



Harnessing Human-AI Collaboration to tackle Spatial Crowdsourcing challenges in Citizen Science

PhD dissertation by MAITE PUERTA BELDARRAIN

within the doctoral program INGENIERÍA PARA LA SOCIEDAD DE LA
INFORMACIÓN Y DESARROLLO SOSTENIBLE

Candidate
MAITE PUERTA
BELDARRAIN

Advisor
DR. OIHANE GÓMEZ
CARMONA

Co-advisor
DR. DIEGO LÓPEZ
DE IPIÑA GONZÁLEZ
DE ARTAZA

MAITE
PUERTA
BELDARRAIN

Digitally signed
by MAITE PUERTA
BELDARRAIN
Date: 2025.06.24
18:10:04 +02'00'

LOPEZ DE
IPIÑA
GONZALEZ DE
ARTAZA DIEGO
- 44670524Q

Firmado digitalmente
por LOPEZ DE IPIÑA
GONZALEZ DE
ARTAZA DIEGO -
44670524Q
Fecha: 2025.06.24
19:26:51 +02'00'

OIHANE
GOMEZ
CARMONA

Firmado digitalmente por
OIHANE GOMEZ CARMONA
Nombre de reconocimiento (DN):
c=ES, ou=Herritar ziurtagiria -
Certificado de ciudadano,
cn=OIHANE GOMEZ CARMONA,
givenName=OIHANE, sn=GOMEZ
CARMONA,
serialNumber=78948965Q
Fecha: 2025.06.24 18:14:37
+02'00'

A esa niña que nunca imaginó llegar tan lejos.
A mi compañero de vida, por su apoyo incondicional.
A mi familia, por su amor y apoyo constante.
Gracias.

Abstract

Citizen Science (CS) has become an essential approach for addressing real-world challenges by actively involving non-professionals in scientific research. Many CS initiatives rely on voluntary contributions rather than financial incentives, making participant engagement and task allocation critical challenges. One of the most complex aspects within CS is the field of Spatial Crowdsourcing (SC), where tasks must be performed in specific geographic locations within a limited timeframe. Traditional SC solutions typically depend on monetary rewards to ensure participation, which is not generally suitable for CS, where motivation is often driven by personal interest, learning, or altruism. For this reason, this dissertation explores Human-AI Collaboration as a mechanism to enhance task allocation in CS, proposing that collaboration between humans and AI-based systems can balance efficient task completion with a positive user perception, fostering greater acceptance and adoption of Spatial Crowdsourcing solutions without relying on financial rewards.

For validating this hypothesis, this research develops a collaborative task allocation strategy, where a spatial task allocation recommender dynamically suggests spatial execution of tasks in the most needed locations while taking into account user preferences, their behavioural patterns, and availability. For that this dissertation introduce three task allocation strategies - Volunteer Task Allocation Strategy (VTAS), Bio-inspired VTAS (Bio-VTAS), and Hybrid-VTAS - which integrate AI-driven recommendations with

an active user involvement in the decision-making process. That is, rather than simply automating task distribution, these systems engage volunteers as collaborative agents, enabling them to influence and adapt AI-driven recommendations based on their own motivations and constraints. By shifting from a purely automated system to a cooperative framework, this approach envisions to increase engagement, optimize task execution, and maintain equitable workload distribution across all locations and participants. The core of this dissertation has been to design and evaluate three alternative spatial crowdsourcing strategies which incrementally satisfy three design requirements, thus leading towards an optimal approach for an effective and user-aware spatial task allocation recommender.

To assess the effectiveness of these spatial crowdsourcing strategies, the research combines simulated experiments with real-world case studies conducted at the University of Deusto. The evaluation examines different conditions, including variations in crowd size, task availability probabilities, and task completion probability. Results, against baseline algorithms which suggest nearest task locations to users, demonstrate that the proposed AI-assisted spatial task allocation recommender not only improves task completion rates and optimizes spatial distribution but also enhances user trust, motivation, and long-term engagement. Feedback from participants was collected from the real-world case studies, both in the form of structured questionnaires and semi-structured interviews. The obtained insight highlights the importance of transparency and control in AI-driven recommendations, reinforcing the idea that collaboration-rather than mere automation-plays a key role in sustaining volunteer participation in CS.

In summary, this research contributes to the advancement of AI-driven crowdsourcing, human-centred computing, and Citizen Science paradigm by offering a scalable and incentive-free approach

to task allocation. The findings suggest that collaboration between humans and AI can create more sustainable, engaging, and effective Citizen Science initiatives, bridging the gap between efficiency and human motivation.

Resumen

La Ciencia Ciudadana (CC) se ha convertido en un enfoque esencial para abordar desafíos del mundo real al involucrar activamente a no profesionales en la investigación científica. Muchas iniciativas de CC dependen de contribuciones voluntarias en lugar de incentivos financieros, lo que hace que la participación de los voluntarios y la asignación de tareas sean desafíos críticos. Uno de los aspectos más complejos dentro de la CC es el campo denominado Spatial Crowdsourcing (SC), donde las tareas deben realizarse en ubicaciones geográficas específicas dentro de un período de tiempo limitado. Las soluciones tradicionales de SC suelen depender de recompensas monetarias para garantizar la participación, lo que no es generalmente adecuado para la CC, donde la motivación suele estar impulsada por el interés personal, el aprendizaje o el altruismo. Por esta razón, esta tesis explora la Colaboración Humano-IA como un mecanismo para mejorar la asignación de tareas en CC, proponiendo que la colaboración entre humanos y sistemas basados en IA para equilibrar la finalización eficiente de tareas con una percepción positiva del usuario, fomentando una mayor aceptación y adopción de soluciones de SC sin depender de recompensas financieras.

Para validar esta hipótesis, esta investigación desarrolla una estrategia colaborativa de asignación de tareas, en la que un recomendador de asignación de tareas sugiere dinámicamente la ejecución de tareas en las ubicaciones donde más se necesitan, teniendo en cuenta las preferencias del usuario, sus patrones de

comportamiento y su disponibilidad. Para ello, esta tesis introduce tres estrategias de asignación de tareas: la Estrategia de Asignación de Tareas Voluntarias (VTAS), la VTAS Bio-inspirada (Bio-VTAS) y la VTAS Híbrida (Hybrid-VTAS), que integran recomendaciones impulsadas por IA con la participación activa del usuario en el proceso de toma de decisiones. Es decir, en lugar de simplemente automatizar la distribución de tareas, estos sistemas involucran a los voluntarios como agentes colaborativos, permitiéndoles influir y adaptar las recomendaciones generadas por IA en función de sus propias motivaciones y restricciones. Al pasar de un sistema puramente automatizado a un marco cooperativo, este enfoque busca aumentar la participación, optimizar la ejecución de tareas y mantener una distribución equitativa de la carga de trabajo en todas las ubicaciones y participantes. El núcleo de esta tesis ha sido diseñar y evaluar tres estrategias alternativas de SC que satisfacen de manera incremental los tres requisitos definidos por nuestra hipótesis, conduciendo así a un enfoque óptimo para un recomendador eficaz y centrado en el usuario de asignación de tareas espaciales.

Para evaluar la efectividad de estas estrategias de SC, esta investigación combina experimentos simulados con estudios de uso en el mundo real realizados en la Universidad de Deusto. La evaluación examina diferentes condiciones, incluidas variaciones en el tamaño de la multitud, probabilidades de disponibilidad de tareas y probabilidades de finalizar las tareas aceptadas. Los resultados, comparados con algoritmos de referencia que sugieren las ubicaciones de tarea más cercanas a los usuarios, demuestran que el recomendador de asignación de tareas espaciales asistido por IA no solo mejora las tasas de finalización de tareas y optimiza la distribución espacial, sino que también aumenta la confianza del usuario, su motivación y su compromiso a largo plazo. Se recopiló la retroalimentación de los participantes a partir de los

estudios de caso en el mundo real, tanto en forma de cuestionarios estructurados como de entrevistas semiestructuradas. La información obtenida resalta la importancia de la transparencia y el control en las recomendaciones generadas por IA, reforzando la idea de que la colaboración, más que la mera automatización, juega un papel clave en el sostenimiento de la participación voluntaria en la CC.

En resumen, esta investigación contribuye un avance del crowdsourcing basada en IA, la computación centrada en el ser humano y el paradigma de la Ciencia Ciudadana al ofrecer un enfoque escalable y libre de incentivos para la asignación de tareas espaciales. Los hallazgos sugieren que la colaboración entre humanos e IA puede crear iniciativas de Ciencia Ciudadana más sostenibles, atractivas y efectivas, cerrando la brecha entre la finalización eficiente de tareas y la motivación humana.

Acknowledgements

La creación de este documento ha sido un largo recorrido de crecimiento profesional y también personal. Por ello, me gustaría agradecer a todas las personas que me han ayudado a llegar hasta aquí. Especialmente:

A mi pareja, porque este viaje no ha sido nada fácil, especialmente en su etapa final, y sin su apoyo y paciencia no habría sido posible. Gracias por estar ahí y por ser mi sostén en los momentos más difíciles.

A Diego López de Ipiña, por confiar en mí desde el inicio y darme la oportunidad no solo de realizar mi tesis con tu apoyo, sino también de haber podido trabajar contigo en el grupo de investigación. Gracias por tu respaldo y paciencia. He aprendido mucho de ti y espero seguir aprendiendo en el futuro. Gracias por cada viaje compartido y por la humanidad con la que trabajas en todo lo que haces.

A Oihane Gómez-Carmona, por su apoyo y paciencia. Gracias por tu ayuda. Sé que aceptar dirigir una tesis no estaba en tus planes a corto plazo, y por ello, y porque sin ti esta tesis sería muy diferente y posiblemente no tan buena, te doy mi más sincero agradecimiento.

A Casado, porque delegar en esta tesis no debió ser una decisión fácil, y sin embargo, lo hiciste. Gracias por tomar esa difícil decisión en un momento tan complicado.

A mi familia, por su apoyo incondicional y su paciencia. Gracias por estar ahí siempre.

To Limming Luke Chen, for giving me the opportunity to do my internship with him. It was one of the best experiences of my life. Thank you very much for the support and affection during the three months I was part of your group.

Al proyecto SOCIO-BEE y a toda su maravillosa gente, en especial a Daniel y Carlos. Un abrazo muy fuerte.

A mis amigos, por su apoyo, su paciencia y por estar siempre ahí. Un agradecimiento especial a Isabela, Maialen, Ainhoa, Irene y Erroxe, quienes en las últimas semanas fueron mi apoyo y una gran motivación en mi vida, siempre brindándome su respaldo.

A mi equipo de investigación en la Universidad de Deusto, MORElab, por todos los consejos y las lecciones aprendidas. Gracias por todo lo aprendido.

A todos mis compañeros de doctorado, por la compañía y los buenos momentos compartidos. Especialmente a Ignacio, Ana y Manu, que han sido una compañía muy cercana en estos tres, casi cuatro años.

También me gustaría agradecer a todas las personas que han participado en los pilotos de Deusto, sobre todo en el segundo, en el que las temperaturas fueron horribles y, sin embargo, estuvieron allí para realizar ese collage conmigo. Gracias por vuestra paciencia y colaboración.

Por último, aunque suene egoísta, quiero dedicarme esta tesis a mí misma. Por todo el esfuerzo, las incontables horas de dedicación y la disciplina. Los últimos meses han sido muy duros, pero por fin, ya está terminado.

Gracias a todos.
Maite Puerta Beldarrain

Table of Contents

List of Figures	vii
List of Tables	xvii
Acronyms	xxviii
Key Concepts	xxxiii
Mathematical Notation	xxxvi
1 Introduction	1
1.1 Context and Motivation	3
1.2 Hypothesis, Objectives and Scope	7
1.3 Research Methodology	10
1.4 Main Contributions	13
1.5 Thesis Outline	15
2 Related work	17
2.1 Citizen Science	18
2.2 Spatial Crowdsourcing	21
2.2.1 Tasks allocation algorithms	24
2.3 Human-AI Collaboration	27
2.3.1 Collaboration cluster	29
2.3.1.1 Humans as an external tool in the collaboration	31
2.3.1.2 Humans-AI Consensus	34
2.3.1.3 Human-AI asynchronous collaboration	37

2.3.1.4	Artificial Intelligent Generated Content	41
2.3.1.5	Human-AI Collaboration making changes at the same time in the environment	44
2.3.1.6	Humans replacing the artificial reasoner of the system	50
2.3.2	Design Principles	52
2.3.2.1	Adaptation & Customization	53
2.3.2.2	Trust & Explicability	56
2.3.2.3	Engagement	59
2.3.2.4	Communication & Feedback	61
2.4	Summary and Conclusions	63
3	Spatial Crowdsourcing Model	65
3.1	Mathematical Model	67
3.1.1	Structure of the Upcoming Chapters and Their Re- lation to the Iterative Process presenting incremental Spatial Crowdsourcing strategies	73
3.2	Summary and Conclusions	75
4	Volunteer Task Allocation Strategy: VTAS	79
4.1	Algorithm Overview	81
4.2	Baseline algorithms, Evaluation metrics and Use Cases details .	90
4.2.1	Baseline Algorithms	90
4.2.2	Evaluation Metrics	93
4.2.3	Use Cases	96
4.3	Simulation Use Case	96
4.3.1	Experiment 1: Crowd Size Variation	102
4.3.2	Experiment 2: User Availability probability	109
4.3.3	Experiment 3: User Task Completion Probability	115
4.3.4	Conclusions of the evaluation of the VTAS Algorithm in the Simulation Use Case	119
4.4	Deusto Use Case	121
4.4.1	Simulation vs real-world Evaluation	124
4.4.2	Survey Evaluation	129

4.4.3	Conclusions of the evaluation of the VTAS in the Deusto Use Case	136
4.5	Summary and Conclusions	138
5	Bio-inspired Volunteer Task Allocation Strategy: Bio-VTAS	141
5.1	Algorithm Overview	144
5.2	Simulation Use Case	152
5.2.1	Experiment 1: Crowd Size Variation	154
5.2.2	Experiment 2: User Availability Probability	161
5.2.3	Experiment 3: User Task Completion Probability	166
5.2.4	Conclusions of the evaluation of the Bio-VTAS Algorithm in the Simulation Use Case	171
5.3	Deusto Use Case	172
5.3.1	Simulation vs real-world Evaluation	175
5.3.2	Survey Evaluation	180
5.3.2.1	Comparison of the Deusto Experiment with the First Evaluation	187
5.3.3	Conclusions of the evaluation of the Bio-VTAS algorithm in the Deusto Use Case	188
5.4	Summary and Conclusions	190
6	Hybrid Volunteer Task Allocation Strategy: Hybrid-VTAS	193
6.1	Face-to-face interview Evaluation	197
6.1.1	Conclusions of the face-to-face interviews	207
6.2	Algorithm Overview	208
6.3	Simulated Use Case	213
6.3.1	Experiment 1: Crowd Size Variation	215
6.3.2	Experiment 2: User Availability probability	222
6.3.3	Experiment 3: User Task Completion Probability	227
6.3.4	Conclusions of the evaluation of the Hybrid-VTAS Algorithm in the Simulation Use Case	231
6.4	Summary and Conclusions	233

7	Conclusions and Future Work	237
7.1	Summary of work and conclusions	238
7.2	Contributions	242
7.2.1	Scientific Contributions	242
7.2.2	Technical Contributions	245
7.3	Hypothesis and objective validation	245
7.4	Relevant publications	251
7.4.1	International JCR Journals and Book Chapters	251
7.4.2	International Conferences	252
7.4.3	Software	253
7.5	Future Work	253
7.6	Final remarks	256
	Bibliography	257
A	Examples of Campaign Shapes	289
B	Deusto Use Case Questionnaire	291
B.1	Demographic information	292
B.2	Satisfactory and activity information	294
B.3	Adoption information	294
B.4	Perceived personalization information	296
B.5	Recommendations and feedback information	297
B.6	Feedback about the Telegram Bot	298
B.7	Comparison of Conducted Pilots	299
B.8	Thank you for your participation	300
C	Telegram Bot	301
C.1	Personal Information Collection	302
C.2	Bot Interaction Process for Image Capture	303
C.3	End of the Experiments	306

D	Deusto personal face-to-face interview	309
D.1	Introduction to the interview	310
D.2	Block 1 - Way of use	311
D.3	Block 2 - Frequency of use	311
D.4	Block 3 - Feedback	312
D.5	Block 4 - Motivation	312
D.6	Block 5 - Conclusions of the interview	313
E	Visual representation of the LFI metric though the experi- ments	315
E.1	Experiment 1	315
E.1.1	100 Users	316
E.1.2	75 Users	319
E.1.3	50 Users	322
E.1.4	25 Users	325
E.2	Experiment 2 - User Availability Probability	328
E.2.1	75% User Availability Probability	328
E.2.2	50% User Availability Probability	331
E.2.3	25% User Availability Probability	334
E.3	Experiment 3 - User Tasks Completion Probability	337
E.3.1	75% User Task Completion Probability	337
E.3.2	50% User Task Completion Probability	340
E.3.3	25% User Task Completion Probability	343

List of Figures

1.1	Representation of the Research Methodology.	13
2.1	A diagram that encompasses all the major research fields in this dissertation.	18
2.2	Types of Citizen Science categorised by the stages of the scientific process in which citizens participate.	20
2.3	Illustration of the workflow in which users serve as an optional element to improve and customize the AI output.	33
2.4	Three different human positions for this integration level. In the first workflow, the human position precedes the AI. In the second, humans and AI work in parallel. In the last, the user follows the AI.	40
2.5	A visual representation of the integration level between humans and AI in collaborative efforts. Both humans and AI have the autonomy to modify the environment and must respond to each other's contributions to achieve effective collaboration.	45
3.1	An example of a campaign created in Bilbao. In this picture, the distribution of the cells are shown.	68
3.2	Several campaigns with different shapes.	69
3.3	Completion Colour scale used to represent the progress of the campaign towards its task-completion goal.	72
3.4	A campaign representation using the completion colour scale defined.	72

3.5	Outline of the Sections and Structure of the Following Three Chapters.	75
4.1	This figure illustrates the interactive workflow of the VTAS with a single user during a campaign. The process encompasses key stages, including campaign configuration, cell-recommendation, and the dissemination of results.	85
4.2	An example of a cell-recommendation using the VTAS strategy. In this campaign, the parameter M_s is set to 8, and the completion colour scale is created based on this value. The user, represented by the person icon, is recommended three nearby cells, indicated by the green and cross icons. The green icon is the accepted cell-recommendation.	89
4.3	An example of a cell-recommendation for the NN algorithm. The user (represented by the person icon) is recommended to the nearest cell (indicated by the green icon) to perform a task within that cell. In the case of this campaign, the M_s value is set to 8.	91
4.4	An example of a cell recommendation for the LNN algorithm: As shown in this picture, the user (represented by the person icon) is recommended the nearest cell that has fewer than M_s completed tasks. The recommended cell is depicted by the green icon in the picture. In the case of this campaign, the M_s value is set to 8.	92
4.5	An example of a cell recommendation using the VTAS algorithm in a simulation. The user's position is depicted by the black person icon, and three nearby recommended cells are shown, marked with green and cross icons, representing the cells with the highest task priority. Among these, the cell with the green icon is the one selected by the simulated user as the recommended cell. This cell is where the user completes a task. The blue person icon signifies the user's final destination. . . .	100

4.6	A visualization of the campaign used to evaluate the VTAS in the simulation use case. This figure was obtained during the initial phase of the campaign, before any tasks were carried out.	101
4.7	The LFI (Eq. 4.4 left-hand side pictures) and realised task (Eq. 4.3 right hand-side figures) over the first slot across different users count (100, 75, 50, and 25 Users). The task allocation algorithms used are the VTAS, LNN, and NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results presented are the mean of five executions. Since the campaign consists of two slots and the tasks-completion objectives were reset at the beginning of each slot, only the first slot is presented.	104
4.8	Visualization of the completed tasks at the end of the first slot, using the VTAS algorithm with 25 users. The completion colour scale indicates the number of completed tasks, based on the M_s variable of the campaign.	107
4.9	Visualization of the completed tasks at the end of the first slot, calculated using the NN algorithm with 25 users. The completion colour scale indicates the number of completed tasks, based on the M_s variable of the campaign.	108
4.10	Visualization of the completed tasks at the end of the first slot, using the LNN algorithm with 25 users. The completion colour scale indicates the number of completed tasks for each cell and it is created based on the M_s variable of the campaign.	109
4.11	The LFI (Eq. 4.4 left figures) and realised task (Eq. 4.3 right figures) over the first slot across different User Availability Probabilities. The algorithms used are the VTAS, the LNN, and the NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used.	112

4.12	The LFI (Eq. 4.4 - left figures) and realised task (Eq. 4.3 - right figures) over the first slot across different User Task Completion Probabilities. The task allocation strategies used are the VTAS, the LNN, and the NN algorithms. To illustrate the evolution of the task realised, the moving average of 20 values was used. The results show the mean of five executions.	117
4.13	The final collage of photos obtained in the Deusto use case using the VTAS algorithm.	123
4.14	The LFI metric (Eq. 4.4- left picture) and Realised tasks(Eq. 4.3 - right picture) in the Deusto real-world campaign (referred to Real-world VTAS) and the results obtained in its simulation using the VTAS, LNN, and NN. To illustrate the realised task over time, the moving average (MA) of 20 values was used in the right hand side pictures.	126
5.1	An example of the neighbouring cells (purple ones) of the orange cell. The white cells are not neighbours of the orange cell but they are still part of the campaign. In the case of this Figure, M is set at 3 to facilitate the visualization.	147
5.2	The threshold colour scale that represents the user's threshold value for each cell. Based on this colour scale, red indicates a minimum threshold value, signifying high specialization, while blue indicates the maximum threshold value, signifying no specialization.	147
5.3	Visualization of the specialization of a user in the cells of a campaign. The threshold colour scale represents the user's threshold value for each cell. In this case, the user has an associated routine and a 48.42% probability of repeating this trajectory. In addition, the θ_{max} is set at 10 and the θ_{min} is set at 1.	148

5.4	An example of a cell-recommendation using the Bio-VTAS algorithm with the completion colour scale. The user, represented by the person icon, is recommended three nearby cells, indicated by the green and cross icons, that have the highest task priority. Each cell has associated with two number in brackets that represent the number of tasks realised and the number of tasks accepted in the cell.	150
5.5	An example of a cell-recommendation using the Bio-VTAS algorithm for a user with the threshold colour scale (see Figure 5.2). The user, represented by the person icon, is recommended three cells, indicated by the green and cross icons, that have the highest task priority.	151
5.6	Visualization of the campaign configuration used to evaluate the effectiveness of the Bio-VTAS solution in this use case. . . .	153
5.7	The LFI metric (Eq. 4.4 left-hand side pictures) and the realised task per minute (Eq. 4.3 right hand-side figures) over the first slot of the campaign across different users count (100, 75, 50, and 25 Users). The task allocation algorithms used are the VTAS, Bio-VTAS, LNN, and NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results presented are the mean of five executions. . .	156
5.8	The task completion campaign involved 25 users using the Bio-VTAS algorithm. This image illustrates the distribution of tasks among the cells at the end of the first slot of the campaign.	160
5.9	The LFI (Eq. 4.4 left figures) and Realised Task (Eq. 4.3 right figures) over the first slot of the campaign across different user availability probabilities. The task allocation strategies used are the VTAS, Bio-VTAS, LNN, and NN algorithms. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results show the mean of five executions.	163

5.10	The LFI metric (Eq. 4.4 - left-hand figures) and the Realised Tasks (Eq. 4.3 - right-hand figures) over the first slot of the campaign across different User Task Completion Probabilities. The task allocation solutions used are the VTAS, Bio-VTAS, the LNN, and the NN algorithms. To illustrate the evolution of the Realised Task, the moving average of 20 values was used. The results show the mean of five executions.	168
5.11	The final collage of photos obtained in the Deusto campaign.	174
5.12	The LFI metric (Eq. 4.4- left picture) and Realised tasks (Eq. 4.3 - right picture) over time in the Deusto experiment and the result obtained in its simulation using the VTAS, LNN, NN and Bio-VTAS. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used.	176
5.13	Figure of a user expertise level at the end of the Deusto use case. In this case, the user has a expertise in the red zone of the map. The three number of each cell represent the number of task realised $rea_{j,i}^t$, number of task accepted $accep_{j,i}^t$ and the threshold of this use for this cell $\theta_{i,j}^t$. The threshold colour scale is based on the threshold value of the user for each cell.	178
5.14	Figure of a user expertise level at the end of the Deusto use case. In this case, the user does not have any expertise in any part of the map. The three number of each cell represent the number of task realised $rea_{j,i}^t$, number of task accepted $accep_{j,i}^t$ and the threshold of this use for this cell $\theta_{i,j}^t$. The threshold colour scale is based on the threshold value of the user for each cell.	179
6.1	An example of N cell-recommendations using the Hybrid-VTAS algorithm. The user, represented by the person icon, is recommended three nearby cells, indicated by the green (the accepted one) and cross icons. In this example, the behaviour of VTAS is observed through the Hybrid-VTAS.	212

6.2	An example of N cell-recommendations using the Hybrid-VTAS algorithm. The user, represented by the person icon, is recommended three nearby cells, indicated by the green (the accepted one) and cross icons. In this example, the behaviour of Bio-VTAS is observed through the Hybrid-VTAS.	213
6.3	Visualization of the campaign configuration used to evaluate the effectiveness of the Hybrid-VTAS solution in this use case. This campaign is the same as the one used in Chapter 4, Section 4.3.	215
6.4	The LFI (Eq. 4.4 left-hand side pictures) and the Realised Task (Eq. 4.3 right hand-side figures) over the first slot across different users count (100, 75, 50, and 25 users). The task allocation algorithms used are the Hybrid-VTAS, VTAS, Bio-VTAS, LNN, and NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results presented are the mean of five executions.	217
6.5	The task completion campaign involved 25 users using the Hybrid-VTAS algorithm. This image illustrates the distribution of tasks among the cells at the end of the first slot of the campaign.	221
6.6	The LFI (Eq. 4.4 left figures) and the realised task (Eq. 4.3 right figures) over the first slot of the campaign across different User Availability Probabilities. The task allocation strategies used are the VTAS, Bio-VTAS, Hybrid-VTAS, LNN, and NN algorithms. To illustrate the evolution of the realised tasks, the moving average (MA) of 20 values was used. The results show the mean of five executions.	224

6.7	The LFI (Eq. 4.4 - left figures) and the Realised Task (Eq. 4.3 - right figures) over the first slot of the campaign across different task completion probabilities. The task allocation strategies used are the VTAS, Bio-VTAS, Hybrid-VTAS, the LNN, and the NN algorithms. To illustrate the evolution of the Realised Task, the moving average of 20 values was used. The results show the mean of five executions.	229
A.1	Several campaigns with different shapes.	290
B.1	The logo, created by a participant for this experiment, symbolizes both the concept of the experiment and the potato omelet that was offered to the participants at the end of the experiment.	292
C.1	Telegram bot requesting and verifying personal information. . .	302
C.2	The bot provides the user with three location-based recommendations.	304
C.3	The user selects "Recommendation 2" and the bot sends the corresponding location.	304
C.4	The bot verifies that the user's location matches the selected recommendation before requesting an image.	305
C.5	The user submits the captured image, and the bot confirms task completion.	305
C.6	The bot informs the user that no further recommendations are available, as the campaign's task-completion goal has been achieved.	307
D.1	A picture of the collage created by the users during the Deusto use case to evaluate the Bio-VTAS, which was presented to them at the beginning of the interview.	310
E.1	Hybrid-VTAS - 100 Users	316
E.2	VTAS - 100 Users	317
E.3	Bio-VTAS - 100 Users	317
E.4	LNN - 100 Users	318

E.5	NN - 100 Users	318
E.6	Hybrid-VTAS - 75 Users	319
E.7	VTAS - 75 Users	320
E.8	Bio-VTAS - 75 Users	320
E.9	LNN - 75 Users	321
E.10	NN - 75 Users	321
E.11	Hybrid-VTAS - 50 Users	322
E.12	VTAS - 50 Users	323
E.13	Bio-VTAS - 50 Users	323
E.14	LNN - 50 Users	324
E.15	NN - 50 Users	324
E.16	Hybrid-VTAS - 25 Users	325
E.17	VTAS - 25 Users	326
E.18	Bio-VTAS - 25 Users	326
E.19	LNN - 25 Users	327
E.20	NN - 25 Users	327
E.21	Hybrid-VTAS - 75% User Availability Probability	328
E.22	VTAS - 75% User Availability Probability	329
E.23	Bio-VTAS - 75% User Availability Probability	329
E.24	LNN - 75% User Availability Probability	330
E.25	NN - 75% User Availability Probability	330
E.26	Hybrid-VTAS - 50% User Availability Probability	331
E.27	VTAS - 50% User Availability Probability	332
E.28	Bio-VTAS - 50% User Availability Probability	332
E.29	LNN - 50% User Availability Probability	333
E.30	NN - 50% User Availability Probability	333
E.31	Hybrid-VTAS - 25% User Availability Probability	334
E.32	VTAS - 25% User Availability Probability	335
E.33	Bio-VTAS - 25% User Availability Probability	335
E.34	LNN - 25% User Availability Probability	336
E.35	NN - 25% User Availability Probability	336
E.36	Hybrid-VTAS - 75% User Task Completion Probability	337
E.37	VTAS - 75% User Task Completion Probability	338

E.38 Bio-VTAS - 75% User Task Completion Probability	338
E.39 LNN - 75% User Task Completion Probability	339
E.40 NN - 75% User Task Completion Probability	339
E.41 Hybrid-VTAS - 50% User Task Completion Probability	340
E.42 VTAS - 50% User Task Completion Probability	341
E.43 Bio-VTAS - 50% User Task Completion Probability	341
E.44 LNN - 50% User Task Completion Probability	342
E.45 NN - 50% User Task Completion Probability	342
E.46 Hybrid-VTAS - 25% User Task Completion Probability	343
E.47 VTAS - 25% User Task Completion Probability	344
E.48 Bio-VTAS - 25% User Task Completion Probability	344
E.49 LNN - 25% User Task Completion Probability	345
E.50 NN - 25% User Task Completion Probability	345

List of Tables

2.1	Some examples of Human-AI Collaboration works in which the user is an optional workflow element used to improve or customize the AI output. This table highlighting the main common characteristics of these group. ($\square$: No; $\boxtimes$: Yes)	32
2.2	Collaboration features of human-AI collaboration papers where the goal of the collaboration is to create a consensus decision. This table highlighting the main common characteristics of these group. ($\square$: No; $\boxtimes$: Yes)	35
2.3	Some examples og Human-AI Collaboration works and their main collaboration characteristic where users are an essential part of the system and where the interaction is thought an interface. The goal of the collaboration is to work together to achieve better results rather than for one agent to assist the other. ($\square$: No; $\boxtimes$: Yes)	38
2.4	Some examples og Human-AI Collaboration works where user and AI agents interact in a physical environment. This table highlighting the main common characteristics of these group. ($\square$: No; $\boxtimes$: Yes)	46
2.5	Some examples of collaborative works where human replacing the AI reasoner component. This table highlighting the main common characteristics of these group. ($\square$: No; $\boxtimes$: Yes)	51

3.1	Glossary of the variables utilized in the Spatial Crowdsourcing model presented in this thesis. The mathematical notation utilized in this chapter is also explained at glossary titled Mathematical Notation	67
4.1	Glossary of mathematical variables used in the VTAS description. These variables are also explained in the Glossary called Mathematical Notation	86
4.2	TCR evaluation over different user crowds (100, 75, 50, and 25) using the VTAS, LNN, and NN algorithms. The values in bold indicate the high TCR in each slot for each user count. .	102
4.3	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count.	105
4.4	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each user count.	107
4.5	TCR evaluation over the different User Availability Probabilities using the VTAS, LNN, and NN algorithms. The highlighted results correspond to the algorithms with the highest TCR in each slot for each tested User Availability Probability.	111
4.6	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the VTAS, LNN, and NN algorithms across different User Availability Probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability.	113

4.7	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the VTAS, LNN, and NN algorithms across different user availability probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each tested User Availability Probability.	114
4.8	TCR evaluation over the different User Task Completion Probability using the VTAS, LNN, and NN. The highlighted results correspond to the algorithms with the highest TCR in each slot for each tested User Availability Probability.	116
4.9	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each Task Completion Probability.	118
4.10	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted values correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability.	118
4.11	TCR metric obtained in the Deusto experiment and in the simulations of this campaign using the VTAS, LNN, and NN algorithms.	125
4.12	MEAN $\pm$ SD of the LFI and RT metrics obtained in the Deusto experiment (real-world VTAS) and in the simulations of this campaign using the VTAS, LNN, and NN algorithms.	128
4.13	Demographic information of the Deusto use case.	129

4.14	Inquiries about the participant's interest level in the Deusto experiment. This table shows the users' questions, possible options, and answers. Question where answers is a scale of the 1 to 5, where 1 is the user strongly disagrees (low satisfied) and 5 is the user strongly agrees (highly satisfied).	130
4.15	Inquiries concerning the system's adoption by participants and their perception of it. This table shows the users' questions, possible options, and answers. Question where answers is a scale of the 1 to 5, where 1 is the user strongly disagrees and 5 is the user strongly agrees.	132
4.16	Questions related to the users' perceived customization of the system. This table shows the users' questions, possible options, and answers. Question where answers is a scale of the 1 to 5, where 1 is the user strongly disagrees and 5 is the user strongly agrees.	133
4.17	Feedback about the Telegram Bot.	134
4.18	Recommendations and feedback information.	135
4.19	The MEAN $\pm$ SD of each quantitative section of the questionnaire.	138
5.1	Glossary of Mathematical Notation used in the Bio-VTAS algorithm description. This table lists variables used in the Bio-VTAS formulation and only the variables that will be used in this chapter are included; for the other variables, see the glossary Mathematical Notation	146
5.2	TCR evaluation over the different number of users (100, 75, 50, and 25) using the VTAS, Bio-VTAS, LNN, and NN task allocation algorithms. The results are the mean of five executions. The bold values are the hight results obtained in each slot for each user count.	154

5.3	The MEAN $\pm$ SD of the Realised tasks (RT - Eq. 4.3) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count.	157
5.4	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each user count.	158
5.5	The TCR performed across different User Availability Probabilities using the VTAS, Bio-VTAS, LNN, and NN task allocation algorithms. The reported results represent the mean of five executions. The highest values indicate the best outcomes achieved in each category for the tested User Availability Probability.	162
5.6	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user availability probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability.	164
5.7	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user availability probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each tested User Availability Probability.	165
5.8	The TCR metric conducted across different task completion probabilities using the VTAS, Bio-VTAS, LNN, and NN algorithms. The highest values represent the best results achieved in each category for the tested User Task Completion Probability.	167

5.9	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each Task Completion Probability.	169
5.10	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted values correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability.	170
5.11	TCR metric obtained in the Deusto experiment and in the simulations of this campaign using the Bio-VTAS, VTAS, LNN, and NN algorithms.	176
5.12	MEAN $\pm$ SD of the LFI and RT metrics obtained in the Deusto experiment (real-world Bio-VTAS) and in the simulations of this campaign using the Bio-VTAS, VTAS, LNN, and NN algorithms.	177
5.13	Demographic information of the Deusto use Case.	180
5.14	Inquiries about the participant's interest level in the Deusto experiment. This table shows the users' questions, options, and answers.	181
5.15	Inquiries concerning the system's adoption by participants and their perception of it. This table shows the users' questions, options, and answers.	183
5.16	Questions related to the users' perceived customization of the system. This table shows the users' questions, options, and answers.	184
5.17	Feedback about the Telegram Bot	185
5.18	Recommendations and feedback information	186

5.19	The mean $\pm$ standard deviation (SD) of each quantifiable section of the questionnaire realised after the campaign conducted at the University of Deusto (Bilbao' Campus) using the Bio-VTAS approach.	187
5.20	Recommendations and feedback information.	187
5.21	Comparison of the mean $\pm$ standard deviation (SD) for each quantitative section of the questionnaire realised after campaigns conducted at the University of Deusto (Bilbao Campus) using the VTAS and Bio-VTAS strategies.	188
6.1	Table of questions related to how users want to interact with for spatial crowdsourcing technologies used in Citizen Science domain.	198
6.2	Table of questions related to the frequency of use of the users.	199
6.3	Table of questions related to the feedback of the users.	201
6.4	Table of questions related to the motivation of the users.	203
6.5	Table of questions related to the conclusion of the interview	205
6.6	TCR evaluation over the different number of users (100, 75, 50, and 25) using the VTAS, Bio-VTAS, Hybrid-VTAS, LNN, and NN task allocation algorithms. The results are the mean of five executions.	216
6.7	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different users counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count.	219
6.8	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each user count.	220

6.9	TCR evaluation over the different user availability probabilities using the VTAS, Bio-VTAS, Hybrid-VTAS, LNN, and NN task allocation algorithms. The highlighted results correspond to the strategies with the highest TCR in each slot for each tested User Availability Probability.	222
6.10	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Availability Probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability.	225
6.11	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Availability Probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each tested User Availability Probability.	226
6.12	TCR evaluation over the different User Task Completion Probability using the VTAS, Bio-VTAS, Hybrid-VTAS LNN, and NN.	228
6.13	The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each Task Completion Probability.	230
6.14	The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted values correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability.	230

7.1	Correlation between gaps, objectives and contributions of this thesis.	250
-----	---	-----

Acronyms

aML	Automatic Machine Learning
AI	Artificial Intelligence
AIGC	Artificial Intelligence Generative Content
AL	Active learning
API	Application Programming Interface
Bio-VTAS	Bio-Inspired Volunteers Task Allocation Strategy
CS	Citizen Science
HiTL	Human-in-the-loop
Hybrid-VTAS	Hybrid Volunteers Task Allocation Strategy
IML	Interactive Machine Learning
IoT	Internet of Things
LFI	Load Fairness Index
LNN	Limited Nearest Neighbours
MA	Moving Average
ML	Machine Learning
MVE	Micro Volunteering Engine

NN	Nearest Neighbours
REST	Representational State Transfer
RL	Reinforcement learning
RT	Realised Tasks per Minute
SC	Spatial Crowdsourcing
SD	Standard Deviation
TCR	Task Completion Rate
VTAS	Volunteer Task Allocation Strategy

Key Concepts

Accepted Tasks or Accepted Cell-Recommendations: When a user accepts a cell-recommendation, it is referred to as an accepted task.

Bio-inspired Volunteer Task Allocation Strategy (Bio-VTAS): The algorithm proposed in Chapter 5, described in Section 5.1.

Campaign: The tasks-completion process for accomplishing a set of tasks is referred to as a *campaign*. The campaign is composed by several cells but always have almost one cells in a campaigns. As the cell are the spatial component of the campaign, the slot are the division of the temporal component of the campaign.

Cell: Static areas in polygonal shape into which the spatial component of a campaign is divided. Tasks in a campaign must be completed at the center of a cell. All cells in the same campaign have the same area. Every campaign contains at least one cell.

Cell-Recommendation: When a user requests a recommendation from the algorithm, it returns N or fewer recommended cells, known as cell-recommendations.

Cell's Priority or Priority: This metric, described in Eq. 4.1, defines the priority of a cell, that is how big is the need of a cell for having another realised task. The priority considers various campaign features, such as the time remaining until the slot ends, the number of completed tasks, and the number of accepted tasks in the cell. Comparing this metric across cells helps

identify which cell requires more tasks to maintain equitable distribution. This metric is referred to as **stimulus** in the Bio-VTAS chapter 5.

Citizen Science (CS): Citizen Science involves the active participation of non-professionals in one or more stages of the scientific process to address real-world challenges (Haklay et al., 2021; Finke, 2014).

Completion Colour Scale: This term refers to the completion colour scale used in the figures to represent the progress of each cell toward its goal of achieving M_s tasks per slot. This representation is introduced for the first time in this thesis in Chapter 3, Section 3.1, in Figure 3.1. Furthermore, this completion colour scale provides a visualization of both the campaign’s progress toward the tasks-completion goal and the LFI metric (Equation 4.4). The selected completion colour scale can be found at Figure 3.3.

Equitable Tasks Distribution: This is evaluated in Goal 3 of the designed algorithms, which is based on ensuring that the number of tasks completed in each cell remains as balanced as possible throughout the campaign. Meeting this condition ensures that, even if participation decreases or abruptly ceases during the campaign, the final dataset will still be representative and retain its statistical validity.

Hybrid Volunteer Task Allocation Strategy: (Hybrid-VTAS): The algorithm proposed in Chapter 6.2, explained in Section 6.2.

Limited Nearest Neighbours (LNN) Algorithm: Another baseline algorithm used in the simulated use case, detailed in Chapter 4, Subsection 4.2.1.

Load Fairness Index (LFI): Measure the equitable distribution of the tasks over the cells at a concrete moment. That is the variance of the set of data composed by the number of complete tasks of each cell in a concrete slot at the time t of the campaign. It is defined in Eq. 4.4 and ensures Goal 2.

Micro Volunteering Engine (MVE): A RESTful API developed to model the Spatial Crowdsourcing problem, perform simulations with baselines, and

host the Telegram bot. It also implements the VTAS, Bio-VTAS, and Hybrid-VTAS algorithms. The API can be found at Github¹.

Moving average (MA): A Moving Average (MA) is a simple way to analyse data trends over time by smoothing out short-term fluctuations. It works by calculating the average of a set of values over a specific period and then continuously updating this average as new data points are added. For example, in a 7-day moving average, you take the average of the last 7 days of data. The next day, you drop the oldest value and add the newest one, recalculating the average. This helps reveal overall trends by reducing the impact of daily ups and downs.

Nearest Neighbours (NN) Algorithm: A baseline algorithm used in the simulated use case. Details can be found in Chapter 4, Subsection 4.2.1.

Overloading Users: This term refers to situations where users are assigned more tasks than they can or are willing to complete. Avoiding overloading is crucial to prevent user demotivation. This also means that the performance of tasks by a user is not an exorbitant and disproportionate effort.

Realised Tasks per Minute (RT): A metric that evaluates the rate at which tasks are completed by users, as defined in Eq. 4.3 to ensure Goal 2.

Representative Values or Cell's Representative Value: At the end of the campaign, a representative value is calculated for each cell during every slot. This value is determined using a statistical metric applied to the tasks completed in one cell during each slot. The dataset composed with the final representative values is the final dataset of the campaign.

Slot: A campaign's duration is divided into sub-periods called slots. All slots within a campaign have the same duration, except for the last one, which may be shorter.

Spatial Crowdsourcing: Spatial Crowdsourcing is a model of crowdsourcing where tasks or data collection activities are assigned to individuals (the crowd)

¹<https://github.com/mpuerta004/RecommenderSystem>

located in specific geographic areas. These tasks often require the participants to be physically present at certain locations or interact with their surroundings in some way (Gummidi et al., 2019).

Tasks-Completion Goal or Tasks-Completion Requirements: This is the target number of tasks to be completed during each slot of the campaign. Achieving M_s tasks in every cell during each slot constitutes the campaign's goal.

Tasks-Completion Procedure: This refers to the process in which a campaign has commenced, and users are actively completing tasks within the campaign's cells.

Task Completion Rate (TCR): A metric used to measure the number of tasks completed by users. It is defined in Eq. 4.3 and ensures Goal 2.

Threshold: Represented by $\theta_{u,j}^t$, this is the threshold of user u for cell j at time t . It defines the user's expertise in a cell and is used in the Bio-VTAS and Hybrid-VTAS algorithms.

Threshold colour scale: A different colour scale was created to represent a user's specialization across all cells of the campaign. Figure 5.2 illustrates the Threshold colour scale used to indicate a user's threshold for all campaign cells. An example of this colour scale is shown in Figure 5.3.

User Availability Probability: A variable in the simulation use case defining the likelihood of a user requesting a task. It ranges from 0 to 100% and is explained in Section 4.3, Chapter 4.

User Task Completion Probability: A variable in the simulation use case defining a user's likelihood of completing tasks. It ranges from 0 to 100% and is explained in Section 4.3, Chapter 4.

Volunteer Task Allocation Strategy (VTAS): The algorithm proposed in Chapter 4, explained in Section 4.1.

Workload Limitation: This defines the maximum number of tasks each cell in the campaign can handle during a slot, as determined by the variable M_s . This limit is directly tied to the task-completion goal, ensuring that once the goal is achieved, the workload constraint prevents any additional tasks from being carried out, thereby ensuring the goal is not exceeded.

Mathematical Notation

$\mathbf{m}$	Total number of cells of a campaign.
$\mathbf{n}$	Total number of slots of a campaign.
$\mathbf{L}$	Set of all cells of a campaign.
j	A cell of the Campaign.
$\mathbf{S}$	Set of all slots of a campaign.
S_i	The i-th slot of a campaign.
d_s	Duration of a slot (excluding the last one).
M_s	Maximum number of tasks that can be completed per cell within each slot.
$St(S_i)$	Start time of slot S_i .
$Ft(S_i)$	End time of slot S_i .
$t_{j,i}^k$	The k-th task assigned to cell j during slot S_i .
C_d	Distance between the centers of two adjacent cells.
P_j^t	Priority of cell j at time t , calculated according to Eq. 4.1.
$accept_{j,i}^t$	Number of tasks accepted by cell j in the slot S_i until time t .
$rea_{j,i}^t$	Number of tasks completed by cell j in the slot S_i until time t .

$\{rea_{j,i}^t\}_{j \in L}$	Set of integers representing the number of tasks completed by each cell in the slot S_i at time t .
$MEAN(rea^t)$	The mean of the set $\{rea_{j,i}^t\}_{j \in L}$, representing the average number of tasks completed in the S_i slot of the campaign at time t .
N	The number of cell recommendations provided by the task allocation solution when a user requests a recommendation. VTAS, Bio-VTAS, and Hybrid-VTAS each recommend this specified number of cells for request. If fewer than N cells are available, the task allocation solution will recommend all available cells.
M	Fixed number defining the radius of a cell's neighbourhood in the Bio-VTAS algorithm.
Pos_i^t	Position of user i at time t .
$n(j)$	Set of cells located in the vicinity of cell j .
$\theta_{i,j}^t$	Threshold of user i for performing the task of cell j at time t .
θ_{max}	Maximum allowable value of the threshold.
θ_{min}	Minimum allowable value of the threshold.
$d(a,b)$	Distance between location a and location b .
ξ_0	Learning coefficient associated with a task completed by a user.
ξ_1	Learning coefficient for tasks related to the completed task from the surrounding area.
φ	Forgetting coefficient.
α, β	Positive constants used in the Bio-VTAS algorithm, as explained in Section 5.1, Chapter 5.

Never doubt that a small group of thoughtful, committed citizens can change the world. Indeed, it is the only thing that ever has.

Margaret Mead

CHAPTER

1

Introduction

Citizen Science (CS) is defined as the active involvement of non-professionals in one or more stages of the scientific process to tackle real-world challenges (Haklay et al., 2021; Finke, 2014). Many Citizen Science initiatives emerge from community efforts aimed at resolving issues that directly impact local communities. Indeed, Citizen Science initiatives are fundamentally dependent on the voluntary and altruistic engagement of individuals, which is motivated by personal development, enjoyment, or altruistic intentions-essentially, non-financial incentives (Levontin et al., 2022). Consequently, effectively addressing Citizen Science initiatives necessitates the development of tailored approaches that take into account user preferences, behaviours, and their willingness to contribute (Vohland et al., 2021)-factors that cannot be fully captured by economic-based models (Meitei and Marchang, 2024b). Furthermore, many applications-such as traffic, noise, and air quality monitoring-often do not provide financial incentives to the users involved (Mousa et al., 2015).

Citizen Science overlaps with the Spatial Crowdsourcing (SC) paradigm, which involves the assignment of participants to tasks that must be performed in designated locations and within specified time (Gummidi et al., 2019). Spatial Crowdsourcing experiments involve gathering and utilizing a wide range of

spatio-temporal data (Tong et al., 2020; Guo et al., 2018), offering significant potential across various applications. These applications span multiple domains, including environmental data collection (Stevens and D’Hondt, 2010), transportation exemplified by Uber¹ and Didi², journalism (Lehmann et al., 2013a,b), and business intelligence platforms such as Gigwalk³ and TaskRabbit⁴. In the literature, addressing Spatial Crowdsourcing problems involves task allocation, where tasks are assigned dynamically among users for execution at different locations (Gummidi et al., 2019). The objective is to ensure that tasks are executed efficiently to achieve the desired outcomes, which is referred to in this dissertation as the tasks-completion goal. Task allocation is widely recognized within the literature, with a multitude of algorithms proposed to address its challenges (Guo et al., 2018; Hettiachchi et al., 2022). Nevertheless, these algorithms primarily emphasize financial incentives as the key strategy for motivating and enhancing user participation.

Typical solutions for Spatial Crowdsourcing do not align with the principles of Citizen Science. Spatial Crowdsourcing strategies mainly rely on economic incentives to encourage user participation. In contrast, Citizen Science requires a more personalised approach that considers users’ preferences, behavioural patterns, and willingness to engage. These aspects cannot be properly addressed by models that depend solely on financial rewards. Therefore, there is a significant potential to develop task allocation strategies that focus on enhancing user experience and promoting sustained, effective participation without relying on monetary incentives. Therefore, there is an opportunity to address this limitation through a Human-AI Collaboration system that assigns tasks based on user preferences and behaviour patterns, operating without financial incentives.

For these reasons, this thesis aims to develop a task allocation strategy for a Spatial Crowdsourcing model applicable to contexts where participants are volunteers and do not receive monetary compensation, that is, in a Citizen

¹<https://www.uber.com/>

²<https://web.didiglobal.com/>

³<https://www.gigwalk.com/>

⁴<https://www.taskrabbit.com/>

Science context. This solution must be capable of adapting to the preferences and behaviours of users in order to foster their acceptance and adoption. Additionally, it is imperative that tasks are carried out efficiently and effectively to meet the objectives of the Spatial Crowdsourcing requirements. Thus, this solution should achieve an equilibrium between the requirements of task completion and the provision of a personalised, positive user experience, all without reliance on financial incentives. In essence, this thesis proposes a Human-AI Collaboration system aimed at fostering a mutually beneficial relationship between the system and its participants, achieving this balance without depending on financial incentives.

The remainder of this chapter is organized as follows: Section 1.1 provides an explanation of the context and motivation behind this research. Section 1.2 outlines the hypothesis, objectives, and scope of the study. Section 1.3 details the methodology employed to achieve these objectives. Section 1.4 enumerates the contributions of this dissertation, and Section 1.5 presents the structure of this dissertation.

In addition, the key definitions, created for this thesis, are provided in the glossary entitled **Key Concepts**.

1.1 Context and Motivation

This thesis aims to develop a task allocation strategy based on a Spatial Crowdsourcing model applicable to the context of Citizen Science through a Human-AI Collaborative approach. To fully appreciate the motivation behind this research, it is essential to examine the misalignments that arise when traditional Spatial Crowdsourcing solutions are applied to altruistic participation, that is, in the Citizen Science context. Recognising these gaps underscores how collaborative approaches can realign task completion processes to provide a more positive user experience. Thus, fostering greater user acceptance and adoption while efficiently meeting the objectives outlined in the Spatial Crowdsourcing requirements. In this thesis, the task completion procedure is defined as the process to of realising the tasks that are part of the

Spatial Crowdsourcing experiment, and the user experience is defined as the perception and feelings of the users when interacting with the system.

As has been introduced Citizen Science represents a collaborative effort between society and the scientific community, where citizens contribute valuable data, local insights, and unique perspectives to the scientific process (Haklay, 2012). This collaboration expands the scope and reach of scientific research, fosters community engagement, enhances scientific literacy, and promotes pro-social behaviour (Bonney et al., 2009b). Citizen Science offers academic benefits through accelerated discoveries and large-scale data collection while delivering social benefits such as informed policy-making and stronger community ties around shared goals.

However, Citizen Science initiatives face several challenges that can hinder their effectiveness. Data quality and reliability are critical concerns, as varying participant expertise may lead to inconsistencies in data collection (Taneja et al., 2024). Additionally, participant retention and motivation are significant challenges; keeping citizens engaged over time requires tailored strategies that resolve issues while motivating users to continue participating without overloading them (Gómez-Carmona et al., 2024; Vohland et al., 2021). Lack of training and accessibility can also limit participation, as many initiatives necessitate specific skills or tools. Effective communication and feedback are essential for maintaining participant engagement and ensuring data quality, yet managing these aspects can be complex, particularly in large volunteer groups. To ensure sustained engagement and effective collaboration, Citizen Science requires consideration of participant preferences and needs, especially since participation in these initiatives is unpaid (Mousa et al., 2015; Levontin et al., 2022).

In this context, it seems imperative to incorporate a more user-oriented perspective into the Spatial Crowdsourcing domain. Indeed, since Spatial Crowdsourcing models rely on maintaining user participation through economic rewards, introducing the Citizen Science context necessitates rethinking both the task allocation strategy and the way the problem is approached. Therefore, it is imperative to develop task allocation strategies that adapt to users' preferences and behaviours in order to encourage their acceptance

and adoption without relying on monetary incentives (Meitei and Marchang, 2024b).

Moreover, the misalignment between Spatial Crowdsourcing and Citizen Science extends beyond economic incentives; in an altruistic context, users are essential to the solution and should be viewed as active participants (Skarlatidou et al., 2019). Their motivations and limitations must be considered, as these factors can significantly impact their willingness to engage (Puerta-Beldarrain et al., 2023). Users should feel motivated to maintain their engagement and participation, while the system must avoid overwhelming them with tasks to prevent a decline in involvement. However, many Spatial Crowdsourcing strategies fail to consider these aspects, leading to a negative user experience. Consequently, problem-solving strategies for the Spatial Crowdsourcing model within the context of Citizen Science need to be flexible and responsive to users' desires and needs, ensuring that solutions are efficient and consider users' preferences and behaviour. As a result, the task allocation solution in the Citizen Science context should be different from typical proposed solutions in other contexts.

Therefore, a user-personalised solution is needed to address the Spatial Crowdsourcing framework within the context of Citizen Science. The proposed solution must balance effective tasks-completion procedures with a positive user perception that fosters sustained and meaningful participation without relying on financial rewards. Given that a positive user experience is reflected in an increase in user acceptance and adoption (Marangunić and Granić, 2015; Ala'a and Ramayah, 2023), these will be the factors that the proposed solutions should promote to determine that a satisfactory user perception has been provided. Consequently, this thesis proposes Human-AI Collaboration as a fundamental tool for achieving a balance between developing optimal solutions that meet tasks-completion requirements and ensuring a positive user perception (that is, increasing the acceptance and adoption of users) without overburdening participants.

Collaboration is understood as two or more entities working together, sharing responsibility for solving problems, and actively interacting to achieve a

common goal (Petri, 2010). While traditional AI systems rely on rigid, one-dimensional strategies, collaboration between humans and AI enables solutions that evolve and adapt to the changing needs of participants. AI has demonstrated extraordinary computational capabilities, excelling in tasks such as pattern recognition, adapting to unforeseen scenarios, and analysing large datasets using advanced techniques like zero-shot learning and pre-trained models (Cao et al., 2023; Wu et al., 2023a; Han et al., 2021). These strengths allow AI to process information at speeds far beyond human capacity, unaffected by stress, motivation, or fatigue (Bosch et al., 2019). However, its performance heavily depends on the quality of the data it is trained on. Biases or inaccuracies in training data can lead to problematic outcomes, particularly in sensitive areas such as ethics, politics, race, and gender (Hofeditz et al., 2023). Additionally, AI often lacks a deep understanding of context and language, limiting its ability to interpret complex situations and make accurate decisions (Wu et al., 2023a). In contrast, humans possess inherent contextual awareness and ethical reasoning, enabling them to navigate ambiguity and take more nuanced decisions (Vicente and Matute, 2023). Human-AI Collaboration combines AI’s computational power with human intuition, ethical judgment, and adaptability. This approach not only enhances AI accuracy but also builds trust and optimises decision-making processes.

Therefore, Human–AI collaboration enables the weaknesses of each team member to be offset by the strengths of the other, as many of these limitations are complementary to the strengths of the other agent. So, Human-AI collaboration allows users to engage in the decision-making process, ensuring that solutions are adaptive to the users’ needs as required by Citizen Science initiatives (Nikolaidis et al., 2017c). Collaborative solutions create adaptive systems that align with users’ preferences and needs, while respecting their participation and limitations. Therefore, Human-AI Collaboration can serve as an effective strategy to involve users in the decision-making process of task allocation, achieving the balance necessary within the context of Citizen Science. Hence, Human-AI Collaboration offers a promising framework to achieve the desired balance by combining the computational power of AI with

the contextual understanding, ethical reasoning, and adaptability of humans. Hence, collaborative systems can better align with participants' needs.

In conclusion, Human-AI Collaboration could serve as the link that ensures solutions address technical requirements, such as creating an effective task completion procedure and considering and promoting a positive user perspective. **By fostering sustained and effective participation without relying on financial rewards, we shift the central design focus from optimal task completion to a balance between technical goals and user perspective.** This shift represents the main objective of this thesis and the primary motivation behind the research.

1.2 Hypothesis, Objectives and Scope

Based on the reviewed fields and misalignments that arise when traditional Spatial Crowdsourcing solutions are applied to altruistic participation, the hypothesis of the present dissertation is:

Hypothesis: Human-AI Collaboration can improve user acceptance and adoption of Spatial Crowdsourcing task allocation recommendations within Citizen Science by balancing efficient task completion procedures with a positive user perception without relying on economic incentives.

Hence, this dissertation sets the following aim to validate the aforementioned hypothesis:

Aim: The aim of this thesis is to develop a task allocation strategy based on a Spatial Crowdsourcing model, tailored to the context of Citizen Science, through a collaborative approach between humans and artificial intelligence. The solution aims to ensure effective task completion whilst enhancing user acceptance and adoption, without relying on economic incentives.

This general aim can be achieved by addressing the following objectives:

- (O1) **Analysis of the Current State of Spatial Crowdsourcing in Citizen Science contexts:** Investigate the existing Spatial Crowdsourcing approaches, focusing on the task allocation problem and the strategies used in Citizen Science contexts.
- (O2) **Analysis of the Current State of Human-AI Collaboration Workflow:** Investigate the current state of the art in Human-AI Collaboration, focusing on the workflow of the collaboration and the integration level of the users in these approaches.
- (O3) **Analysis of Human-AI Collaboration Design Principles:** Understand how to create an effective Human-AI collaborative system, focusing on the design principles necessary to engage users in the decision-making process.
- (O4) **Design of a Spatial Crowdsourcing Model:** Define a Spatial Crowdsourcing model that establishes the spatial and temporal constraints, task completion goals, and workload limitations for Citizen Science experiments. This model should be flexible and adaptable to various Citizen Science contexts, accommodating different data collection requirements and user motivations.
- (O5) **Design and Implementation of a Task Allocation Recommender System:** Create a Human-AI Collaboration task allocation system that balances, on one hand, user engagement and acceptance and, on the other hand, an effective task completion procedure.
- (O6) **Establish an Evaluation Methodology:** Identify appropriate evaluation methodology to validate the proposed system effectively. The proposed solutions are evaluated through a series of experiments, demonstrating the effectiveness of the Human-AI collaborative approach. Results indicate that these solutions outperform naive task allocation solutions (baselines) in terms of both user acceptance and task completion efficiency. In addition, the proposed solution has to be evaluated based on several questionnaires that verify the adoption and acceptance of

the proposed solutions, that is, that the proposed solution improves the perspective of users with respect to the traditional solutions.

The resulting optimised and interactive system should also fulfil the following requirements:

1. **Use Case Independence:** The proposed approaches are designed to be use case-independent, meaning they can be applied to many Citizen Science initiatives that require the concrete task allocation model designed in this thesis. Different use cases are tested to validate the proposed approaches. The proposed approaches exhibit scalability, allowing for adaptation to diverse user and task requirements. The spatial crowdsourcing model or campaign can be customised to address the specific needs of various use cases.
2. **User Engagement:** The proposed approaches aim to engage users in the decision-making process without overloading them with tasks. The goal is to enhance user acceptance and foster the adoption of Spatial Crowdsourcing technologies.
3. **Applicability:** The various use cases analysed in this dissertation are selected to provide a comprehensive evaluation framework. The proposed approaches are designed to be independent of the specific activities chosen as tasks to realize.
4. **Real-world Applicability:** The proposed approaches are designed to be applicable to online scenarios.

Beyond that, the work presented in this dissertation does not deal with the following conditions:

1. **User Interaction:** This thesis does not focus on specific motivation and engagement techniques, but rather on integrating users by fostering their commitment through interaction and active participation.

2. **Data privacy and federated systems:** The proposed solutions are not designed with the aim of creating a federated architecture or a privacy-preserving system. While user privacy is not the main focus of this work, appropriate measures were taken to ensure responsible handling of user data. All questionnaires were conducted in a private manner, and participants' responses were treated with strict confidentiality.
3. **Citizen Science with economic rewards:** The proposed solutions are not designed for Citizen Science initiatives that offer economic rewards to participants. The focus is on altruistic participation and the challenges associated with engaging users in the decision-making process.

This thesis emerged from the requirements of the SOCIO-BEE¹ project. Consequently, several design decisions-such as treating the SC problem as static-were influenced by the project's needs. However, the evaluation of the proposed solutions in this thesis does not include SOCIO-BEE, even though two of the three proposed solutions were successfully tested within its framework. The decision to exclude SOCIO-BEE from the evaluation is based on two main reasons. First, the evaluation criteria for SOCIO-BEE focused on aspects different from those assessed in this thesis. Second, the effectiveness of the tools developed in this thesis, as applied within SOCIO-BEE, is influenced by other components of the project. This interplay makes it impossible to perform a direct comparison or isolate the evaluation of the thesis's contributions. Nevertheless, the successful application of the contribution of this thesis in SOCIO-BEE supports their validity and applicability in real-world scenarios.

1.3 Research Methodology

The research methodology to validate the hypothesis and achieve the outlined research objectives presented in the previous Section 1.2 is structured into several phases. These phases are interconnected and iterative, allowing for the

¹SOCIO-BEE web page

refinement and validation of the research outcomes. The methodology is designed to guide the research process from the initial exploration of the research field to the final dissemination of results. The following phases are included in the research methodology:

1. **Exploratory Phase:** This phase involves examining the literature related to the research field to establish a strong theoretical foundation for the rest of the research process. Although this literature review is presented as an initial step, it is recognised as an ongoing and continuous task that should be conducted throughout the entire research process.
2. **Scope Definition and Validation Scenario:** Following the initial phase and building on existing knowledge, the scope, context, and motivation of this work are defined. The project scenario will then be established in alignment with the set objectives, considering the limitations and advantages of the research proposal. Establishing a robust initial framework is crucial for guiding future research toward the foundational analysis.
3. **Research Design and Methodology:** In this phase, an appropriate research approach is selected, such as qualitative, quantitative, or mixed methods, based on the study's nature. Data collection methods are chosen, including surveys, interviews, experiments, or observations, and relevant instruments or protocols are designed. A sampling strategy is determined to ensure data representativeness and reliability. Ethical considerations are addressed by obtaining necessary approvals, ensuring informed consent, and maintaining data privacy throughout the research process.
4. **Solution Design and Development:** This phase involves designing and developing algorithms, models, or frameworks to address the research problem. The development process is documented, including design, implementation, and testing stages. The appropriate tools and technologies, such as programming languages and development environments, are selected. Implementation involves coding, testing, and

debugging to ensure that the solution meets the specified requirements. Iterative testing and refinement are carried out to enhance the solution's effectiveness and reliability.

5. **Solution Evaluation:** After the design and development phase, the implemented solution undergoes evaluation. This step involves determining whether the system meets the specified requirements and verifying its validity and applicability in real-world environments and contexts.
6. **Data Analysis:** In this phase, the collected data is prepared and pre-processed to ensure quality and readiness for analysis. This includes handling missing values and outliers. Data analysis techniques appropriate to the research approach are applied, such as statistical analysis for quantitative data or thematic analysis for qualitative data. Software tools are utilised for analysis to derive meaningful insights and patterns. The results are compared with existing literature to validate findings and draw conclusions. Additionally, an iterative approach is employed, where the data analysis informs and refines the design and development of alternative approaches, enhancing the alignment with the thesis objectives.
7. **Final Conclusions, Dissemination, and Dissertation Writing:** The concluding task of the research process focuses on analysing the obtained results, drawing conclusions, and highlighting the contributions of the research. These conclusions and contributions are expected to be innovative and significant, impacting both academic and social fields when disseminated. Finally, the research work will be completed and refined through the development, submission, and defence of the PhD dissertation.

Figure 1.1 depicts the research methodology, outlining the various phases and tasks it includes, as well as the iterative process followed to refine the initial conceptualisation of the implemented solutions.

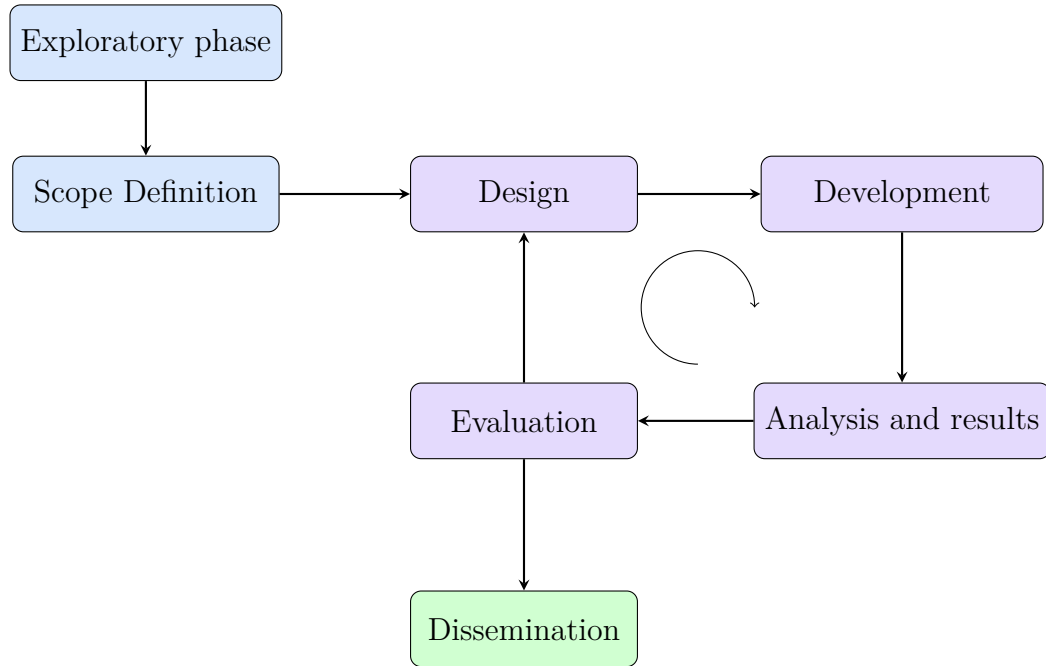


Figure 1.1: Representation of the Research Methodology.

1.4 Main Contributions

The following scientific contributions are reported in this dissertation to achieve the goals outlined in Section 1.2:

- **Analysis the Current State of Spatial Crowdsourcing in the context of Citizen Science:** This contribution examines existing challenges and advancements in Spatial Crowdsourcing within the context of Citizen Science. The analysis highlights the limitations of current approaches for the context of Citizen Science.
- **Analysis of the Human-AI Collaboration Workflows in the literature:** The workflows of existing Human-AI Collaboration systems are examined to identify key components and their interactions. This analysis facilitates the categorisation of different collaborative system architectures and provides a framework for understanding the essential elements and dynamics within these systems.

- **Analysis of the Human-AI Collaboration Design Principles:** An in-depth research of the design principles required to develop an effective Human-AI collaborative system is conducted. This analysis emphasises how these principles can be employed to engage users meaningfully in the decision-making process.
- **Definition of a Spatial Crowdsourcing model:** A Spatial Crowdsourcing model is defined to establish the spatial and temporal constraints, task completion goals, and workload limitations of Citizen Science experiments. This model is designed to be flexible and adaptable to several Citizen Science contexts, accommodating different data collection requirements and user motivations.
- **Design and Implementation of a collaborative Tasks allocation Recommender System:** A human-AI collaborative tasks allocation system is developed to balance user engagement and acceptance while optimising task completion goal efficiency. This system adapts task allocation according to user preferences and crowdsourcing behavioural patterns, fostering user acceptance and enhancing task completion efficiency. Three main solutions are proposed to address the task allocation problem in Citizen Science:
 - **Volunteers Task Allocation Strategy (VTAS) Algorithm:** An initial solution that balances task completion requirements with a positive user experience by distributing tasks evenly across spatial areas. The VTAS algorithm adopts a user-centric approach, allowing participants to manage incomplete tasks, customise task assignments, and reject tasks.
 - **Bio-Inspired Volunteers Task Allocation Strategy (Bio-VTAS) Algorithm:** An advanced version of the VTAS algorithm, incorporating an analysis of user behaviour trajectories to create better adaptive task recommendations. This version aligns task assignments more closely with user crowdsourcing patterns, enhancing user engagement.

- **Hybrid Volunteers Task Allocation Strategy (Hybrid-VTAS)**

Algorithm: The last proposed solution, called Hybrid-VTAS, integrates insights from previous versions to further optimise tasks-completion goals and users' perspectives.

- **Simulation Task Allocation Engine:** A simulation-based engine was developed to evaluate the proposed solutions under various scenarios. This engine models different user behaviours and preferences, offering insights into the effectiveness of the Human-AI collaborative solution proposed in this dissertation.
- **Development of a Telegram bot for Citizen Science:** A Telegram bot is developed to facilitate user interaction with the proposed solutions. This bot allows users to receive recommendations, accept tasks, and manage task assignments, enhancing the overall user experience.
- **Evaluation of Proposed Solutions:** The proposed solutions are evaluated through a series of experiments, demonstrating the effectiveness of the Human-AI collaborative approach. Results indicate that these solutions outperform naive task allocation solutions (baselines) in terms of both user acceptance and task completion efficiency. In addition, the proposed solution has to be evaluated based on several questionnaires that verify the adoption and acceptance of the proposed solutions, that is, that the proposed solution improves the perspective of users with respect to the traditional solutions.

1.5 Thesis Outline

The thesis is structured in seven chapters:

The current one, Chapter 1, introduces the dissertation by presenting the context and motivation for the research alongside the hypothesis, objectives, and key contributions. It also outlines the research methodology used to achieve these objectives and highlights the main contributions of this dissertation.

Chapter 2 reviews the state of the art in Human-AI Collaboration, Citizen Science, and Spatial Crowdsourcing, with a focus on the challenges and limitations of integrating users into the Spatial Crowdsourcing framework in the context of Citizen Science. This chapter analyses the current state of Spatial Crowdsourcing, Human-AI Collaboration design principles, and Human-AI Collaboration workflows in the literature, providing a comprehensive literature review to understand the problems addressed and the state of the art.

Chapter 3 defines the Spatial Crowdsourcing model and presents the nomenclature used in the thesis. The constraints of the model and the requirements to solve the problem are discussed, including the limitation of the workload to avoid overloading users.

Chapter 4 introduces the first solution, which creates a balance between tasks-completion requirements and user perception. This solution maintains an equitable distribution of tasks across spatial areas. The VTAS algorithm is a user-centric approach that can manage incomplete tasks, customize task assignments, and reject tasks, providing a personalised task allocation for each user.

Chapter 5 presents a new version of the VTAS solution based on a bio-inspired division of labour algorithm called Bio-VTAS. This version considers user behaviour trajectories to create task recommendations, aligning task assignments more closely with user patterns and enhancing user engagement.

Chapter 6 introduces an enhanced version of the VTAS, designed to harmonize user preferences with task completion requirements called Hybrid-VTAS. This version incorporates user behaviour trajectories, similar to the Bio-VTAS, to generate task recommendations while maintaining equal task distribution across locations, similar to the VTAS. By aligning tasks assignments more closely with user activity patterns, this approach aims to improve user engagement and participation while maintaining the effectiveness of the task completion process.

Finally, Chapter 7 summarizes this dissertation's main findings and conclusions and outlines future research directions related to this work.

*The key is not to try to beat machines, but to use them
to improve our own skills and knowledge.*

Garry Kasparov

CHAPTER

2

Related work

This thesis addresses the Spatial Crowdsourcing (SC) problem with a focus on its application within the context of Citizen Science (CS) through a Human-AI Collaboration approach. To understand this problem better, a review of the literature is presented in this chapter. The chapter is divided into three main sections: Citizen Science, Spatial crowdsourcing, and Human-AI Collaboration, which are represented in Figure 2.1. The Citizen Science section (Section 2.1) provides an overview of the CS field, highlighting its importance, challenges, and opportunities. The Spatial Crowdsourcing section (Section 2.2) discusses the different task allocation strategies and the strategies that apply to Citizen Science systems. To understand how to create an effective Human-AI Collaboration approach, an analysis of the existing literature on collaborative human-machine workflows is conducted in the collaboration section (Section 2.3). This review encompasses a variety of collaboration techniques and categorizes different collaborative workflow methods, explaining how to integrate users into a collaborative system effectively. This categorization is used to identify the most suitable collaboration techniques and the appropriate level of user integration for addressing the Spatial Crowdsourcing problem. Additionally, several collaboration design principles are examined to guide

the development of the proposed Human-AI Collaboration approach from the user perspective. These design principles help to understand how to design a collaborative system that is both effective and user-friendly.

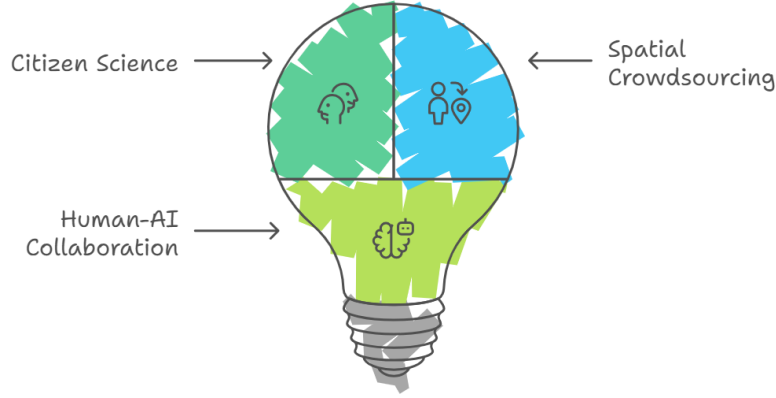


Figure 2.1: A diagram that encompasses all the major research fields in this dissertation.

The chapter concludes with a summary of the main findings and a discussion of the gaps in the existing literature that this thesis aims to address (Section 2.4).

2.1 Citizen Science

Citizen science engages non-professionals in various stages of the scientific process to address real-world challenges (Vohland et al., 2021; Haklay et al., 2021). By fostering collaboration between professional scientists and citizens, CS enhances data availability and contributes to scientific advancement. From a social science perspective, CS aims to democratise science and policy by promoting public engagement, transparency, equity, and justice (Bonney et al., 2009b). It bridges the gap between academic and non-academic communities, offering benefits such as diverse partnerships, integrated methodologies, and innovative solutions (Haklay, 2012). The success of CS initiatives often stems from the collective intelligence, creativity, and dedication of participants, who generate valuable data and insights beyond traditional research methods.

One major factor driving the recent expansion of this field has been the extraordinary advancements in technology, along with the widespread availability of digital systems that facilitate collaboration and empower volunteers to participate actively (Vohland et al., 2021). Another key factor has been the mutual benefits provided by CS projects to both scientists and volunteers. Volunteers can address local concerns, gain practical experience, and make meaningful contributions to scientific research. Simultaneously, scientists benefit from access to local expertise, the ability to collect extensive datasets, and an acceleration of research efforts (Bonney et al., 2009b).

An illustrative example of CS's impact is a project addressing the threat of the invasive mosquito species *Anopheles Stephensi* in Africa, which transmits the disease of Malaria (Carney et al., 2023). In this project, communities collected mosquito samples, and scientists analysed them to identify the presence of *Anopheles Stephensi* and issue public alerts if the species was detected. This example highlights how CS effectively tackles public health issues. CS has also been applied in various fields, including environmental science (Cunha et al., 2017; Froeling et al., 2024), marine research (Earp and Liconti, 2020), astronomy (Masters et al., 2019), archaeology (Wessman et al., 2019) and bird observation (Teng et al., 2024).

CS can be categorised into three primary types based on the stages of the scientific process in which citizens participate (Bonney et al., 2009a):

- **Contributory:** Citizens participate by collecting data according to protocols set by scientists. In some cases, the public may also analyse data using online tools.
- **Collaborative:** Volunteers engage in additional stages of research, such as data analysis, interpretation, and presentation. They may also help design and refine data collection protocols, thereby having a deeper impact on the research process.
- **Co-creative:** This most integrative approach involves citizens in all stages, from project design and data collection to analysis and dissemination. Volunteers often identify research questions related to community

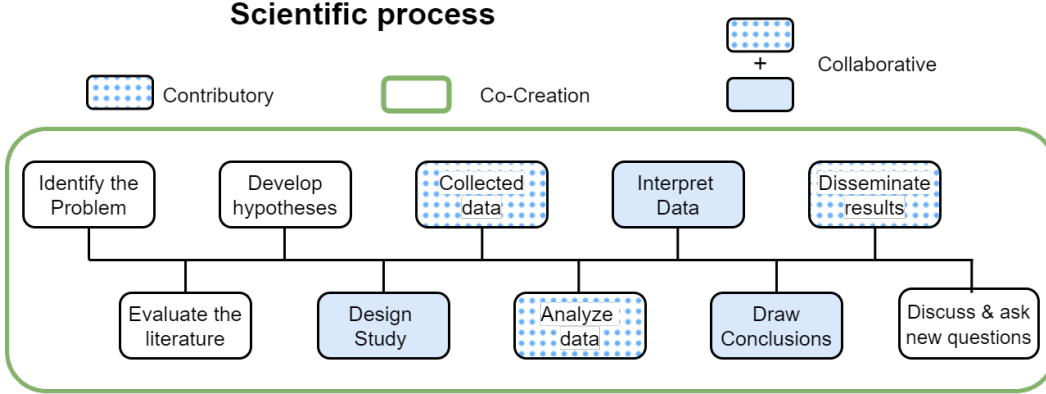


Figure 2.2: Types of Citizen Science categorised by the stages of the scientific process in which citizens participate.

concerns and work with scientists to address these questions and propose solutions (Bonney et al., 2016).

The types of CS are illustrated in Figure 2.2, showing the three types varying the stages of the scientific process in which users participate. While most projects involve volunteers in data collection and partial analysis, achieving full integration across all stages-co-creation-poses challenges both technologically and in terms of engagement (Haklay, 2015; Carney et al., 2023; Pandya, 2012).

CS initiatives rely on volunteers motivated by factors beyond financial rewards, such as personal growth, enjoyment, and altruism (Levontin et al., 2022). The core of CS lies in the community’s intrinsic desire to make a meaningful impact rather than economic incentives. Levontin et al. identify several key motivations for volunteers, grouping them into four categories: intrinsic goals (e.g., personal interest or scientific learning), extrinsic goals (e.g., contributing to research or environmental issues), social goals (e.g., connecting with like-minded individuals), and self-enhancement goals (e.g., skill development or recognition) (Levontin et al., 2022). Understanding these motivations is crucial for managing CS projects effectively, as aligning strategies with users’ motivations ensures sustained participation and valuable contributions. Additionally, CS projects require flexibility to adapt to volunteers’ changing motivations. Rotman et al. emphasise that CS projects must remain

dynamic to address evolving drivers and maintain engagement throughout the project lifecycle (Rotman et al., 2012).

One of the primary challenges in CS is balancing technical efficiency with user experience (Pocock et al., 2024), as both directly impact participation and data quality (Gómez-Carmona et al., 2024; Taneja et al., 2024). New technologies should enhance engagement, build trust, and adapt to users' needs while maintaining technical effectiveness (Puerta-Beldarrain et al., 2023); however, they must also be user-friendly. A focus solely on technical solutions without considering user experience can hinder participation and trust (Skarlatidou et al., 2019). Technologies should be intuitive and accessible, fostering both engagement and effective communication with the system (Sheard et al., 2024). Poor user involvement and experience can lead to reduced engagement, diminished motivation, and compromised data quality.

For CS to succeed, users must view themselves as active contributors to the scientific process, not just data collectors. Engaging participants from the start builds their commitment and ownership of the project (Shirk et al., 2012; Ostrom, 1996). When participants see their input reflected in project outcomes, they feel valued, which strengthens their ongoing engagement (Froeling et al., 2024). Collaborating as equal partners also boosts the community's sense of agency, motivating them to contribute actively to the project's success and future developments (Satcher, 2005).

In summary, the success of CS projects relies on effectively integrating users into the scientific process. This integration requires balancing technical solutions with user experience to ensure that participants feel involved, valued, and empowered throughout the project lifecycle. By fostering engagement and collaboration, CS projects can achieve their goals and produce meaningful outcomes that benefit both the scientific community and society at large.

2.2 Spatial Crowdsourcing

First introduced by Howe in 2006 (Howe, 2006a,b), crowdsourcing is a technique designed to engage non-professionals in the data-gathering phase of the scientific process (English et al., 2018). In recent years, crowdsourcing

has gained popularity as an effective approach for harnessing the collective efforts of large groups to gather information and solve complex problems. Participants typically respond to open calls on the Internet, performing tasks such as data labeling that are better suited to humans than machines. These tasks may have requirements such as the time window in which they must be performed, the skills required, or the cost associated with completing the task. In many cases, users receive financial compensation for their contributions, as seen with platforms like Amazon’s Mechanical Turk¹ (Difallah et al., 2015). By leveraging collective intelligence, collaboration, and diverse contributions, crowdsourcing can address complex problems that are difficult to solve through traditional methods. This approach allows groups of individuals with varying knowledge and skills to engage in tasks through flexible open calls that often include incentives. However, crowdsourcing may be less effective for tasks requiring specific physical locations.

The rise of advanced mobile devices has given birth to a specialized form of crowdsourcing known as Spatial Crowdsourcing (SC), which addresses the physical location limitations of traditional crowdsourcing methods. Unlike conventional techniques, SC requires participants to be physically distributed across different locations and time periods to complete tasks (Gummidi et al., 2019). In many real-world applications, geographic information is crucial, such as in providing real-time traffic updates or monitoring air quality at various locations. SC leverages the spatiotemporal data collected through mobile devices like smartphones, tablets, and wearable sensors (Tong et al., 2020; Guo et al., 2018), unlocking vast potential across diverse applications. A prominent example of SC is Uber², a ride-hailing and peer-to-peer ride-sharing platform. Uber connects users seeking transportation with drivers based on their spatiotemporal context. The system optimizes service delivery by considering factors such as the driver’s maximum daily trip requests, passenger pickup times, and vehicle preferences. Additional constraints, like aligning pickup locations with the driver’s intended route, further enhance efficiency. The

¹<https://www.mturk.com/>

²<https://www.uber.com/>

primary goal is to maximize service delivery while increasing drivers' earnings. SC applications span various domains, ranging from environmental data collection (Stevens and D'Hondt, 2010) and transportation (e.g., Didi¹) to journalism (Lehmann et al., 2013a,b) and business intelligence platforms like Gigwalk² and TaskRabbit³. The tasks that users can perform are categorized into three types (Alt et al., 2010):

- **Sensing Tasks:** Involve gathering data from the environment (e.g., taking pictures, measuring noise levels).
- **Service Tasks:** Involve providing services at a specific location (e.g., delivering items, conducting surveys).
- **Social Interaction Tasks:** Involve interacting with people at a location (e.g., engaging in conversations, gathering opinions).

In the literature, solving SC problems typically involves a tasks allocation, where tasks are allocated to users for execution (Gummidi et al., 2019). Task allocation approaches are categorized into two types: platform-centric and user-centric. In platform-centric tasks allocation, the platform (or server) assigns tasks directly to users. Existing tasks allocation models predominantly focus on platform-centric approaches (Li et al., 2024b; Gao and Zhao, 2022; Huang et al., 2022; Li et al., 2022; Zhan et al., 2018; Zhou et al., 2017). They are primarily concerned with maximizing profit for the platform by minimizing user rewards. On the other hand, in user-centric tasks allocation, users choose from a list of available tasks, and the primary objective remains profit-driven, with an emphasis on increasing user earnings (Yucel and Bulut, 2020; Cai et al., 2020; Chen and Wang, 2020; Cen et al., 2015; Peng et al., 2024; Chen et al., 2024b; Meng et al., 2023; Ding et al., 2022; Zhang et al., 2022; Jin et al., 2021; Wan et al., 2021). Anyhow, all tasks allocation solutions consider the financial reward of the user as the primary strategy to motivate and promote participation.

¹<https://web.didiglobal.com/>

²<https://www.gigwalk.com/>

³<https://www.taskrabbit.com/>

Tasks allocation solutions can be broadly classified into two approaches: online and offline (Wang et al.).

- **Online:** The server cannot predict participant and task information in advance. Tasks with attached attribute requirements are posted periodically. Workers dynamically participate in performing tasks and upload sensing data to the platform for rewards.
- **Offline:** The server knows all worker and task information in advance, such as their locations and resources. In ideal situations, offline algorithms can achieve relatively satisfactory results through optimization.

In conclusion, the tasks allocation problem is complex and challenging due to the dynamic nature of tasks, spatial and temporal constraints, and the need to balance workers' workloads. The SC problem is NP-hard, and finding an optimal solution is computationally expensive. Therefore, heuristic algorithms, such as Greedy algorithms, Genetic algorithms, and Particle Swarm Optimization (Gummidi et al., 2019), are commonly used to find near-optimal solutions within a reasonable timeframe. These algorithms aim to maximize the total number of tasks assigned or the overall payoff and/or minimize the cumulative travel cost for the assigned workers. The choice of optimization objective depends on the application and its purpose. In the next subsection several tasks allocation strategies are analysed.

2.2.1 Tasks allocation algorithms

SC involves outsourcing location-dependent tasks to a large group of mobile workers who perform spatial tasks that require traveling to specified locations (Tong et al., 2020), such as data gathering or taking snapshots of an area. SC employs tasks allocation to distribute tasks to appropriate workers to achieve specific goals while adhering to spatial, temporal, and other constraints. This subsection analyses several tasks allocation strategies to provide a comprehensive overview of the state of the art.

A few studies have handled platform-centric task allocation. First, Li et al. proposed the ContinuousSensing framework, which is a human-robot task allocation model that facilitates cooperation and coordination between different types of robots and humans (Li et al., 2024b). It also features a task migration model that transfers tasks from failed participants to other available participants. This approach improves the quality of crowdsensing task completion while maintaining relatively low task execution costs. Gao et al. developed a personalised task allocation strategy for mobile crowdsensing that minimizes costs by considering both user travel distance and task preferences (Gao and Zhao, 2022). It treats task allocation as a dynamic optimization problem, modelled as a complex version of the travelling salesman problem (TSP). Hang et al. proposed an Optimized Allocation of Time-Dependent Tasks (OPAT), a time-dependent task allocation contribution that considers sensing duration and the user’s sensing capacity (Huang et al., 2022). The goal of this approach is to obtain the most optimal solution, maximizing the sensing capacity of each mobile user. Li et al. developed a fuzzy linguistic method to build the evaluation index matrix to obtain the best allocation strategy (Li et al., 2022). Zhan et al. investigated incentive mechanisms to motivate participation in platform-centric task allocation (Zhan et al., 2018). They employed a one-to-many bargaining framework that considers both the resource limitations and the resource needs of users. Zhou et al. proposed an online auction as a method to incentivize user participation in dynamic systems (Zhou et al., 2017).

On the other hand, other works have created a user-centric task allocation. Yucel et al. created a task allocation to satisfy system stability and user satisfaction (Yucel and Bulut, 2020). They used task matching to minimize unhappy participant pairs while maintaining system stability. Cai et al. proposed user-centric algorithms based on valuation and distance from/to the task (Cai et al., 2020). They calculate the payment according to the task performed. The task owner then selects the participants with the minimum payment. Chen et al. developed an approach called IntelligentCrowd that leverages the power of multi-agent reinforcement learning (MARL) to address

the problem of task allocation (Chen and Wang, 2020). Their work is user-centric in the sense that it aims to maximize the pay-off for the participants. Cen et al. introduced a stochastic task recommendation framework that takes into account workers’ historical trajectories, preferred time budgets, and the inherent probabilistic uncertainty of their future trajectories (Cen et al., 2015). In addition to these studies, many user-centric task allocation approaches also prioritize aligning their objectives with the protection of participant privacy (Peng et al., 2024; Chen et al., 2024b; Meng et al., 2023; Ding et al., 2022; Zhang et al., 2022; Jin et al., 2021; Wan et al., 2021). A key privacy concern with user-centric approaches lies in their reliance on gathering extensive user data, which increases the risk of privacy breaches (Guo et al., 2018).

Some works have tackled incorporating fairness into task allocation solutions. Song et al. developed the Fairness Task allocation Strategy with Distance Constraint (FTAS-D), which leverages Lyapunov optimization to identify a feasible allocation solution and iteratively refines it to find the optimal one (Song et al., 2023). This approach effectively balances the objectives of minimizing costs while ensuring fairness in task distribution among users. In this context, fairness in task allocation involves preventing any single user from being overloaded by distributing tasks evenly to avoid unfair burdens. It also means ensuring that each task is assigned an appropriate frequency so that no task is neglected or favoured disproportionately. Duan et al. developed a task allocation procedure that incorporates a fairness-aware policy (Duan et al., 2020). Their approach focuses on fairness related to the contribution each participant makes to task performance rather than on task distribution or the effort required to complete the tasks. Li et al. introduced incentive mechanisms for task allocation that prioritize both fairness and energy efficiency (Li et al., 2021b). In terms of fairness, the incentives are determined by evaluating the participants’ performance.

The solutions mentioned above are designed to achieve an equilibrium between fulfilling their objectives and being economically advantageous, maximizing the platform’s profit, maximizing the economic reward for users, and/or minimizing transportation costs, among other possibilities. That is, all of these solutions include an economic optimization component. However, many

applications such as traffic, noise, and weather monitoring typically do not offer financial rewards to users involved (Mousa et al., 2015). To the best of our knowledge, only one author has examined task allocation within the context of altruistic and unpaid participation over the past five years. Meitei et al. have developed several approaches in which they eliminated profit-driven motives (Meitei and Marchang, 2020, 2024a,b). In their first work, they proposed two user-oriented task allocation schemes: Nearest User Task Allocation (NUTA) and Nearest User Fair Task Allocation (NUFTA). They focused on achieving an optimized and equitable workload for the users (Meitei and Marchang, 2020). Following this work, they refined their initial two algorithms and introduced two additional user-oriented solutions to accommodate time and distance constraints: Time-Sensitive Task Allocation (TSTAS) and Time-Sensitive Task Allocation with Time Slot Change (TSTA_TSC) (Meitei and Marchang, 2024a). The main goals of these approaches were to 1) minimize the travelling distance between users and their tasks and 2) allocate as many tasks as possible while satisfying time-dependent conditions. Their most recent work introduced a new multi-objective, time-dependent task allocation method called Load-Balanced Task Allocation (LBTA) (Meitei and Marchang, 2024b). The primary objectives of this approach are to 1) allocate as many tasks as possible and 2) create a load-balanced distribution of tasks during the allocation.

However, other challenges related to altruistic participation must be addressed, such as the possible decrease in participation or the overloading of users. Consequently, there is significant potential for developing task allocation strategies that prioritize enhancing the user experience, thereby fostering more sustained and effective participation that is not solely based on financial rewards.

2.3 Human-AI Collaboration

Collaboration has gained significant attention within the AI research community, as collaboration between humans and AI systems has the potential to enhance decision-making processes, improve system performance, and foster

trust between users and AI agents (Wu et al., 2023a; Hofeditz et al., 2023). Although the term Human-AI Collaboration is widely used, its definition remains ambiguous in the literature. Several overlapping concepts, such as Hybrid Intelligence, Collaborative Interactive Learning, and Human-in-the-loop, are often employed to describe related ideas (Sherson et al., 2024; Hanika et al., 2019; Akata et al., 2020; Shi et al., 2022). For the purposes of this work, collaboration is defined as two or more entities working together, sharing responsibility for solving problems, and actively interacting to achieve a common goal (Petri, 2010). This section provides an overview of Human-AI Collaboration systems, focusing on different collaboration clusters and design principles that guide the development of effective Human-AI Collaboration systems. These insights establish the groundwork for addressing challenges in Spatial Crowdsourcing within the broader context of Citizen Science.

AI has demonstrated remarkable computational capabilities, excelling in tasks such as pattern recognition, handling unforeseen scenarios, and adapting to diverse conditions using advanced techniques like zero-shot learning and pre-trained models (Cao et al., 2023; Wu et al., 2023a; Han et al., 2021). These strengths allow AI to identify complex patterns in large datasets and process information at speeds far beyond human capacity, all while being unaffected by stress, motivation, or fatigue (Bosch et al., 2019). However, despite these advantages, there are notable limitations that highlight the critical role of Human-AI Collaboration.

First, AI’s performance depends heavily on the quality of the data it is trained on. Flaws such as inaccuracies or biases in the data can lead to problematic outcomes, especially in sensitive areas like race, gender, ethics, and politics (Wu et al., 2023a; Hofeditz et al., 2023). Additionally, AI systems often lack a deep understanding of language and context, which can lead to misinterpretations and hinder accurate decision-making (Wu et al., 2023a). In contrast, humans possess inherent contextual awareness and ethical reasoning, enabling them to navigate ambiguities and make more nuanced decisions than machines (Vicente and Matute, 2023).

For these reasons, Human-AI Collaboration offers a promising approach that combines AI’s computational power with human reasoning, intuition,

and ethical judgment. Such synergy not only enhances decision-making processes and improves AI accuracy but also fosters trust between humans and AI systems (Wu et al., 2023a; Hofeditz et al., 2023). However, designing effective Human-AI Collaboration systems requires careful consideration of several factors, including the roles of human and AI agents, the nature of their collaboration, and the principles that govern their interaction (Vicente and Matute, 2023).

The following sections provide a classification of Human-AI Collaboration systems and discuss design principles critical to fostering effective interactions between humans and AI. Specifically, the first section explores methods for integrating users into AI systems, highlighting varying levels of interaction and control. The second section outlines key design principles, such as transparency, adaptability, and user-centrality, which are essential for building trust and ensuring effective collaboration. Together, these insights offer a foundation for understanding how collaboration can address the challenges of Spatial Crowdsourcing in the context of Citizen Science.

2.3.1 Collaboration cluster

This section outlines five clusters of Human-AI Collaboration based on various collaboration features and the degree of human integration in AI decision-making. The classification is built upon the following collaboration features:

- *AI Technique (AL, RL, aML, AIGC)*: This feature describes the AI technique used in the collaboration. *AL* refers to Active Learning, *RL* denotes Reinforcement Learning, *aML* stands for automatic Machine Learning algorithms with interactive techniques, and *AIGC* represents Artificial Intelligence Generated Content techniques. We consider automatic machine-learning algorithms that have been modified to include human intervention at specific stages of the workflow, treating them as interactive machine-learning techniques.
- *Expert*: This feature indicates whether a human expert is required for effective collaboration. For instance, certain complex tasks may neces-

sitate expert knowledge, while others can be performed with non-expert humans.

- *Interaction (Interface, Physical)*: This feature describes the nature of the collaboration environment, which can take one of two forms: an *interface* interaction (e.g., software applications) or a *physical* interaction (e.g., collaborative robots working alongside humans).
- *Aim of Collaboration (Cooperation, Improvement, Customization, Consensus, Replace AI)*: This feature categorizes the collaboration’s objectives into five options:
 - *Cooperation*: Agents must continuously cooperate by jointly altering the environment to achieve the final goal of the collaboration. The aim of these works is to enhance communication, physical Cooperation, or mutual understanding between the agents.
 - *Improvement*: The objective is to achieve a better result than that obtained by both users separately.
 - *Customization*: This collaboration aims to obtain improved outputs tailored to the user’s specific needs.
 - *Consensus*: The goal is to reach a unanimous decision or action plan.
 - *Replace AI*: This category applies if the goal is to replace the AI reasoner with human intelligence.

Multiple aims can be selected for one collaborative system, reflecting the complexity of human-AI interactions.

- *Team Composition (1:1 or 1:n)*: This feature describes the composition of the system’s members: *1:1* indicates a system consisting of one AI and one human, while *1:n* denotes a system with one AI and several humans. This distinction is important for understanding dynamics in teamwork and decision-making.

- *Initiative:* This feature describes whether the human agent can initiate communication with the system when desired. A lack of initiative means that human interaction is confined to specific workflow points of the system, limiting the collaboration’s flexibility.

These collaboration features serve as the foundation for the clusters described in the following subsections.

2.3.1.1 Humans as an external tool in the collaboration

Incorporating human input as an optional element to enhance or customize AI outputs is a defining characteristic of the works within this group. The goal of human intervention is to refine future AI outputs or tailor the results, with feedback consisting of a limited set of actions to modify the system’s outcomes. The collaborative workflow for this group is illustrated in Figure 2.3, where users are positioned downstream from the AI output. Table 2.1 summarizes several examples of collaborative works in this group, highlighting their shared characteristics.

Customization is the primary objective of these works. Dawar et al. developed a recommender system that considers music habits, preferences, and facial expressions to suggest suitable songs and playlists to users (Dawar et al., 2023). Users can provide feedback on these recommendations to improve future outputs. Even though face recognition occurs before the AI generates its output, the user’s involvement remains passive, as their cognitive abilities do not actively influence the result. Various recommendation systems are utilized in this group, with AL (Keikhosrokiani and Fye, 2024) and RL (Chen et al., 2024a) being the most common techniques. Another example is the system created by Cai et al., which employs reinforcement learning (RL) for video recommendations (Cai et al., 2023). In this system, users provide feedback, such as likes and follows, to define their preferences and express agreement with the recommendations. A similar approach is taken in the system developed by Murthy et al. (Murthy and Murthy, 2024). These systems are part of our daily lives, embedded in platforms like Spotify, YouTube, and others, offering options to customize recommendations based on user preferences.

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		mInitiative
	AL	RL	aML	AIGC		Interface	Physical	Customization	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
(Allen et al., 2019) - Decision-making Tree	☐	☐	☒	☐	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐
(Cai et al., 2023) - video recommendations	☐	☒	☐	☐	☐	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐
(Dawar et al., 2023) - Music recommendation	☒	☐	☐	☐	☐	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐
(Han et al., 2019) - Topic detection	☐	☐	☒	☐	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☐
(Honeycutt et al., 2020) - Object detection	☐	☐	☒	☐	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐
(Keikhosrokiani and Fye, 2024)- health product recommendation	☒	☐	☐	☐	☐	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐
(Lee et al., 2021) - Health	☐	☐	☒	☐	☒	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐
(Liu et al., 2020) - Classification task	☒	☐	☐	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☒	☐
(Murthy and Murthy, 2024) - video recommendations	☐	☒	☐	☐	☐	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐
(Schroder et al., 2022) - Vegetarian recipe	☒	☐	☐	☐	☐	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐
(Smith et al., 2018) - Topic modeling	☐	☐	☒	☐	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☐
(Yang et al., 2019) - Text classification	☐	☐	☒	☐	☒	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐
(Yeh et al., 2024) - Human-AI Writing	☐	☐	☒	☒	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐

Table 2.1: Some examples of Human-AI Collaboration works in which the user is an optional workflow element used to improve or customize the AI output. This table highlighting the main common characteristics of these group. (☐: No; ☒: Yes)

Han et al. and Smith et al. developed a system that identifies document topics based on word frequency (Han et al., 2019; Smith et al., 2018). These topics can be modified and customized through user input to enhance system performance in subsequent iterations. Another illustrative example is provided by Schroder et al., who analyzed the Plant Jammer system (Schroder et al., 2022). This system recommends vegetarian recipes based on the user’s

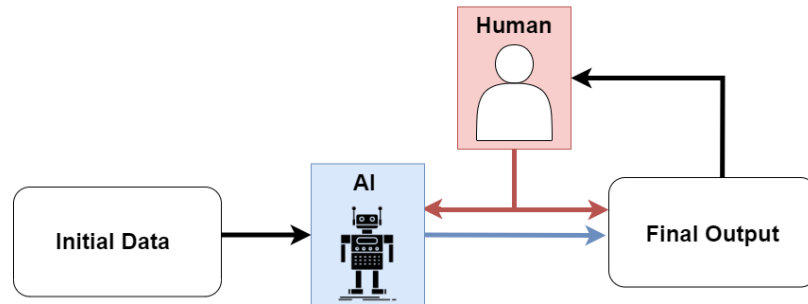


Figure 2.3: Illustration of the workflow in which users serve as an optional element to improve and customize the AI output.

selected ingredients and preferences, with the customization process relying on a rating scale from 'great' to 'okay' for the recipes. A similar approach is utilized by Allen et al., whose system learns users' qualitative preferences to facilitate car selection (Allen et al., 2019). In this case, user preferences are solicited, and the algorithm adapts to their decision-making process. When the model of the user's preferences is presented, users can provide feedback to refine the recommendations. Another example is the work created by Yeh et al. (Yeh et al., 2024), which developed an AI-enhanced writing tool that allows users to exercise greater agency and personalization in the writing process. The system offers suggestions, which users can accept or reject.

Moreover, system customization can also occur through expert human input. For instance, Lee et al. developed a feature selection system that provides quantitative analysis of patients' conditions to enhance rehabilitation assessment practices (Lee et al., 2021). This analysis is presented on a graphical user interface, allowing experts to give feedback on adjusting features to better fit patients' needs and personalize rehabilitation.

Another goal of this level of integration is to leverage human qualities to improve areas where humans excel compared to AIs. For instance, Liu et al. utilized non-experts to label system classification outputs and identify *unknown-unknowns*, learning from them to enhance initial classification (Liu et al., 2020). In this case, a crowd-sourcing platform gathers non-expert labels for large volumes of data without overburdening individuals.

When the system seeks to employ human feedback to enhance outputs, a

learning component is typically incorporated within the AI reasoning framework. However, not all works that integrate humans include this learning element. For example, Yang et al. developed the RulesLearner algorithm, which expresses ML text-analysis rules that experts may modify as needed (Yang et al., 2019). The output of this modification becomes the final result. The importance of learning from human feedback was studied by Honeycutt et al., who developed two systems for detecting human faces in images: one with human feedback and one without (Honeycutt et al., 2020). The human feedback system utilized an iterative workflow in which a human was instructed to correct each AI output. However, the system did not learn from this feedback, even though users were led to believe it did. They aimed to examine how this lack of learning impacts user trust. This work highlighted the necessity of a learning component in iterative user-integrated loops to learn from feedback and maintain user trust.

Therefore, it is evident from the preceding examples that both human experts and non-experts can enhance system performance or customization. The critical aspect of these works is that a previous output is available, making user integration aimed at improving system performance or personalizing the experience optional. Users have become an increasingly integral component of systems in subsequent levels of integration, with expanded freedom of action, communication, and decision-making.

2.3.1.2 Humans-AI Consensus

Works within this integration level aim to establish task allocation agreements and facilitate consensus decision-making among human participants. At this stage, users wield greater decision-making authority compared to previous levels. Typically, the team consists of one artificial intelligence agent and several humans, each of whom may possess unique information regarding the problem at hand. The primary objective is to reach a consensus decision or devise a strategic plan. Table 2.2 summarizes several examples of collaborative projects within this group, emphasizing their key shared characteristics.

One rationale for employing this workflow is that a human coordinator’s decision-making can be swayed by personal motives or biases, whereas an

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
(Lin et al., 2024) - AI assistants	☐	☐	☐	☒	☐	☒	☐	☐	☐	☐	☒	☐	☒	☒	☒
(Pescetelli et al., 2020) - Online game	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☒	☐	☐	☒	☒
(Ramchurn et al., 2016) - Task allocation	☐	☐	☒	☐	☒	☒	☐	☐	☐	☐	☒	☐	☐	☒	☒
(Schurr et al., 2005a,b) - Planning in disaster scenario	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☒	☐	☐	☒	☐
(Zakershahrak, 2021) - Planning in disaster scenario	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☒	☐	☐	☒	☒

Table 2.2: Collaboration features of human-AI collaboration papers where the goal of the collaboration is to create a consensus decision. This table highlighting the main common characteristics of these group. (☐: No; ☒: Yes)

artificial intelligence remains consistently objective. For instance, Pescetelli et al. compared the performance of human and AI coordination in discerning the crowd’s intent in an online game (Pescetelli et al., 2020). They developed an online platform called BeeMe, where users could vote, socialize, and suggest new actions. The coordinator was responsible for interpreting the participants’ intent to determine the next action an actor would take in the game, ensuring that this choice accurately reflected the users’ intentions. A collaborative decision-making algorithm was created that demonstrated a higher level of impartiality compared to a human coordinator.

This workflow is particularly useful in disaster scenarios, where the objectivity of the AI coordinator helps eliminate human bias and its potentially detrimental effects. Additionally, AI excels in efficiently resolving optimization challenges and processing large datasets, which can significantly enhance the effectiveness of interventions in such situations (Pescetelli et al., 2020). The initial efforts to incorporate an AI coordinator into human teams were based on shared control systems, which allowed humans to take control when the system encountered failures (Schurr et al., 2005b,a). However, in dis-

aster contexts, a single lapse in planning can lead to catastrophic outcomes, rendering these approaches less effective.

Consequently, more recent research has focused on developing systems that facilitate communication and promote flexible interactions between humans and AI coordinators, rather than relying solely on shared control techniques. For example, Ramchurn et al. created a flexible human-AI interaction system for generating task allocation plans in emergent scenarios (Ramchurn et al., 2016; Maathuis, 2024). This system empowers users to reject tasks assigned by the AI coordinator, request additional information, and articulate their preferences. Furthermore, the system can inquire about the status of tasks at any time, considering human preferences in the allocation process. As a result, a more adaptable plan emerges from the enhanced communication among team members.

In parallel, other studies advocate for incorporating a third component responsible for facilitating effective communication between humans and artificial agents (Zakershahrak, 2021). Given the fundamental differences in processing between humans and AI, the primary aim of this component is to translate one model into the other and vice versa (Endsley, 2023). Zakershahra developed a collaboration system comprising one AI agent, four warehouse managers, and a facilitator agent to achieve a consensual plan in a disaster scenario (Zakershahrak, 2021). The facilitator utilized the "Wizard of Oz" (WoZ) technique (Musick et al., 2021), a common method for comparing perceptions between human-human and human-agent teams. In this scenario, each participant holds different information about the situation, requiring the facilitator to assist team members in adapting their decision-making to forge a consensus, even when complete information is unavailable or direct communication is limited.

A crucial element for effective communication within teams is the ability of members to share their perspectives. The initiative serves as a communication tool that empowers users to express their views, enhancing their trust and enabling system customization. This initiative involves providing users with tools they can use flexibly, such as customizing the system, declining tasks, or proposing new actions without requiring approval. While the initiative

can take various forms depending on the system, it consistently grants users autonomy and fosters a stronger relationship between the agent and the AI.

For instance, in the BeeMe project developed by Pescetelli et al., users have the option to propose new actions that the system might not consider (Pescetelli et al., 2020). Similarly, in the approach taken by Ramchurn et al., users can request additional information and express their preferences (Ramchurn et al., 2016). Lin et al. designed an AI assistant named Decision-Oriented Dialogue (DoD) to aid one or multiple humans in making complex decisions across three scenarios: assigning reviewers to conference papers, planning a multi-step itinerary in a city, and negotiating travel plans for a group of friends (Lin et al., 2024). In this case, user interaction with the AI occurs through a dialogue environment.

In essence, the most critical aspect of this integration level is the coordination objective of the system. Humans and AI do not necessarily need to work collaboratively on the same task; rather, they must focus on establishing a consensus, which can be viewed as the initial phase of the collaborative process.

2.3.1.3 Human-AI asynchronous collaboration

The integration of users as a fundamental component of AI systems distinguishes the works that comprise this integration level, aimed at enhancing or personalizing performance. Users are essential to the process not only to correct the system but also to actively participate in obtaining the desired output. A defining characteristic of this group is that agents must engage without simultaneously altering the environment. This means human participation is often restricted to specific points in the interaction workflow, or the artificial agent operates in parallel with the user. As such, the collaboration is typically asynchronous. The goal of this collaboration is to work together to achieve improved results, rather than for one agent simply to assist another. However, the user's actions are predetermined and must occur at specific moments within the interaction workflow. Table 2.3 summarizes examples of collaborative works within this group, highlighting their key common characteristics.

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
(Gillotte, 2019) - Generate artworks	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Guimarães et al., 2021) - Similarity detection	☐	☐	☒	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☒	☐
(Hasegawa-Johnson et al., 2017) - Audio descriptions of images	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Hekler et al., 2019) - Medicine, Health	☐	☐	☒	☐	☒	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐
(Holzinger et al., 2019) - Optimization algorithm	☐	☐	☒	☐	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐
(Huang et al., 2024a) - Creation of audio content	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Köbis and Mossink, 2021) - Poems text-generation	☐	☐	☐	☒	☐	☒	☐	☐	☐	☒	☐	☐	☒	☐	☒
(Liu et al., 2023) - Audio storytelling generated	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Liu et al., 2021) - Magnetograms of the Sun	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Liu et al., 2019) - Re-identification tasks	☒	☒	☐	☐	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☐
(Loeschke et al., 2022) - Video generated	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Mozafari et al., 2014) - Classification tasks	☒	☐	☐	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☒	☐
(Pollefeys and Gool, 2002) - 2D images into 3D models	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Schmitt and Buschek, 2021) - Human-ai co-writing	☐	☐	☐	☒	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☒
(Schulman et al., 2022) - ChatGPT	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Sharma et al., 2023) - Text assistant	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Tegen et al., 2019) - Virtual Sensors	☒	☐	☐	☐	☐	☒	☐	☐	☒	☐	☐	☐	☒	☐	☒
(Texler et al., 2020) - Video generated	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Thoppilan et al., 2022; Le et al., 2021) - Question-answering chatbot	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Whittaker et al., 2020) - Create synthetic faces	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Yang et al., 2022) - Human-AI Fiction Co-writing	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒
(Zhang et al., 2024b) - Sepsis Diagnosis	☐	☐	☒	☐	☒	☒	☐	☐	☒	☐	☐	☐	☒	☒	☒
(Zhang et al., 2023d) - Cartographic design processes	☐	☐	☐	☒	☐	☒	☐	☐	☒	☒	☐	☐	☒	☐	☒

Table 2.3: Some examples of Human-AI Collaboration works and their main collaboration characteristic where users are an essential part of the system and where the interaction is thought an interface. The goal of the collaboration is to work together to achieve better results rather than for one agent to assist the other. (☐: No; ☒: Yes)

Consequently, human requirements, such as trust, engagement, and usability, are as crucial as artificial requirements in this integration process (Puerta-Beldarrain et al., 2022). Moreover, a human-centered perspective is employed to incorporate these human requirements. In this framework, humans can assume various roles in the AI system’s learning process, as illustrated in Figure 2.4. For instance, in the study by Hekler et al., both the expert and the AI operate in parallel, independently classifying images of skin cancer, with their classifications combined to arrive at a final diagnosis based on the accuracy of the AI’s predictions (Hekler et al., 2019). In this scenario, both agents exert equal effort to achieve the system’s output by performing the same task independently. Another health-related example is provided by Zhang et al., who developed a method for diagnosing sepsis (Zhang et al., 2024b). Similar to the previous study, both experts and AI work in parallel, with the experts making the final decision.

A comparable methodology was employed by Reverberi et al., who conducted two distinct experiments to explore this collaboration (Reverberi et al., 2022). In the first experiment, endoscopists diagnosed and classified lesions independently, while in the second, they received AI assistance. The study found that endoscopists were influenced by the AI’s recommendations, regardless of its accuracy. However, in cases of disagreement, the endoscopists maintained their original opinions without being swayed by the AI’s input. This exclusive reliance on AI as the sole decision-maker in the healthcare sector can lead to patient aversion (Berger et al., 2021). Conversely, a collaborative approach—where AI assists in the decision-making process—allows us to leverage the capabilities of AI in healthcare, thereby fostering greater acceptance and integration of AI in this domain (Hemmer et al., 2022).

In this system workflow, the human position may precede the AI reasoner, as demonstrated in the study by Guimaraes et al. (Guimarães et al., 2021), which focuses on identifying document pair similarity. In this case, users utilize a crowdsourcing platform to group documents by their authors, after which the AI classifies them based on content similarity. However, users are often positioned after the AI to minimize human effort while maximizing the AI’s

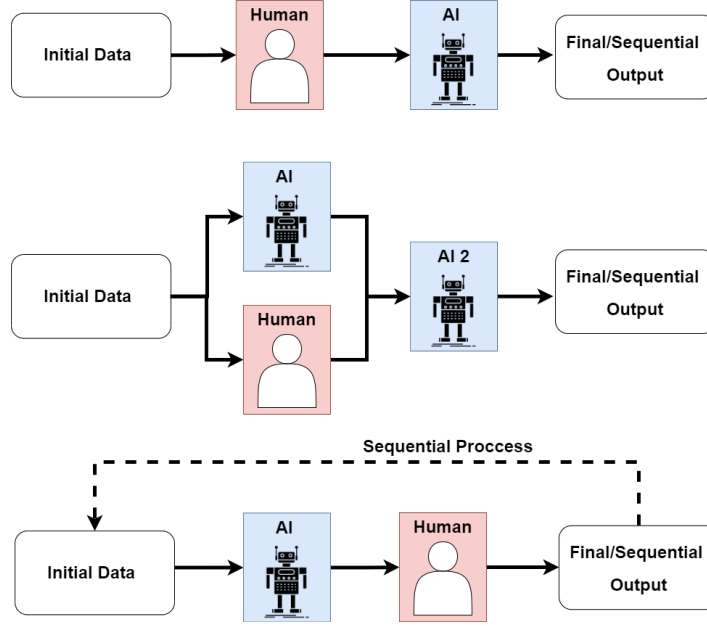


Figure 2.4: Three different human positions for this integration level. In the first workflow, the human position precedes the AI. In the second, humans and AI work in parallel. In the last, the user follows the AI.

data processing capabilities. Figure 2.4 illustrates these different workflows at this integration level.

Common techniques within this integration level include Active Learning and the modification of automatic machine learning algorithms. Another notable technique gaining traction is AI-Generated Content (AIGC), which warrants a separate subsection due to its growing significance and unique method of integrating users.

Crowdsourcing platforms are also widely used since AI algorithms can learn from collective human behaviour, enhancing accuracy over time. This approach helps distribute the human labeling effort among multiple users, as a single user typically cannot label an entire dataset. Some researchers investigate ways to integrate machine learning into crowdsourced databases, aiming to combine the accuracy of human labeling with the speed and cost-effectiveness of machine-learning classifiers. In this vein, Mozafari et al. studied the application of active learning strategies in crowdsourced databases to reduce the number of instances needing labeling by humans, thereby minimizing human

effort and enabling the scalability of this approach (Mozafari et al., 2014).

This challenge is relevant across various contexts. For example, Liu et al. focused on reducing human annotation efforts while maximizing system performance in person re-identification tasks. They developed a deep reinforcement active learning (DRAL) method to guide an agent in selecting training samples in real-time with human annotators.

Another compelling reason to explore ways to reduce human labeling efforts is the susceptibility of humans to mental state influences, such as stress and boredom. In this context, Tegen et al. demonstrated that designing applications where users provide feedback only when the system’s output is incorrect is more efficient than requiring users to provide correct labels consistently (Tegen et al., 2019). Their research indicates that human performance declines when users are forced to exert excessive effort over prolonged periods. In some experiments, they allowed users to actively tag without prompts, empowering users with initiative, which increases freedom in collaboration and significantly contributes to meeting human requirements. However, this flexibility is not present in all studies.

For instance, Holzinger et al. addressed the Traveling Salesman Problem using a snake game to facilitate interaction between the user and the AI (Holzinger et al., 2019). They modified an Ant Colony Optimization (ACO) algorithm, specifically the MMAS (MAX-MIN Ant System), allowing users to select the next target for the snake during interactions. This choice affects the AI’s behaviour, enabling humans to correct deficiencies in the algorithm. However, human interactions are time-restricted, which may not align with the optimal timing for adjustments; addressing issues may be more critical in subsequent iterations or earlier phases.

2.3.1.4 Artificial Intelligent Generated Content

One notable subset of this group is the AIGC (Artificial Intelligent Generated Content) algorithms. These algorithms have gained significant popularity in recent years and have revolutionized various fields and applications, such as medicine (Shao et al., 2024) and tourism (Carvalho and Ivanov, 2024). As mentioned earlier, the performance of these methods is quite straightforward:

the user provides prompt input, and the AI generates content that uses this human input as a guide. This interaction, however, is not inherently interactive. Currently, we are at the forefront of developing this technology, with interactive methods increasingly being integrated into generative AI (GAI) models. For example, ChatGPT has integrated human feedback to enhance performance (Schulman et al., 2022; Wu et al., 2023b). Moreover, some AIGC systems allow for the repetition of this input-output loop, occasionally incorporating memory as a dialogue feature, which enables improvement and customization based on multiple iterations and/or specifications. These algorithms belong to this integration level because the scope of human input is primarily limited to the type of prompt that the AIGC accepts. However, human input is not restricted in other respects. Additionally, human intervention is crucial for collaboration, as users are the most beneficial contributors.

Although these methods have recently gained traction and transformed the way many tasks are performed, particularly in education (Luther et al., 2024; Nguyen et al., 2024; Dhillon et al., 2024), they also hold potential influence over the academic peer review process (Sarker et al., 2024).

The diverse forms of AI-generated content can be classified into three categories: text, image, and audio generation:

Text generation. Creative writing and dialogue writing are the primary subfields of AI-generated text (Li et al., 2024a; Sefeni et al., 2024).

- *Creative Writing:* This involves generating text with greater openness, creativity, and nuance. For instance, in the work of Yang et al., users can co-write a short sci-fi story with a GPT-2-based text generation model (Yang et al., 2022). Another application includes generating poems, as demonstrated by Kobis et al. in their research (Köbis and Mossink, 2021). Additionally, Sharma et al. developed a tool for recommending co-written text messages for mental health support (Sharma et al., 2023).
- *Dialogue Writing:* This category encompasses chatbots that interact with users through text. These bots are designed to answer questions

and provide information. Such AIGC algorithms are considered interactive due to the inherent nature of dialogue (Marri; Ait Baha et al., 2024). For instance, ChatGPT is designed for conversational usage, producing human-like responses by drawing on its extensive knowledge base (Wu et al., 2023b; Li et al., 2024c). ChatGPT can remember previous interactions within the same conversation, thereby facilitating a continuous dialogue. Notably, LaMDA is a question-answering chatbot focused on two domains: Mount Everest education and music content recommendations (Thoppilan et al., 2022; Le et al., 2021). Another example is the software developed by Schmitt et al., which supports the creation of fictional characters through interactive conversation and gradual development (Schmitt and Buschek, 2021). AI-generated text remains the most expansive category of AIGC, partly due to the influence and success of ChatGPT.

Image Generation. By leveraging AIGC, users can modify existing images or generate new ones that meet specific requirements based on given prompts (Stock et al., 2021; Ren et al., 2021), all without needing advanced skills or knowledge. AI-generated images have various applications, ranging from artwork (Gillotte, 2019) to synthetic faces (Whittaker et al., 2020) and complex magnetograms of the Sun (Liu et al., 2021). DALL-E, a widely used AIGC application, generates images from textual descriptions (Ramesh et al., 2021; Adetayo, 2024; Vayadande et al., 2024; Ramesh et al., 2022). Zhang et al. have developed a novel application that integrates cartographic design processes based on DALL-E (Zhang et al., 2023d). Additionally, some applications can translate 2D images into 3D models (Pollefeys and Gool, 2002). Further examples can be found in the survey conducted by Zhang et al. (Zhang et al., 2023a). These AI-generated images can also be used to create videos. AIGCs operate similarly in video generation, processing each frame individually and utilizing AI algorithms to generate video clips, which can be employed to create trailers and promotional videos (Loeschcke et al., 2022; Texler et al., 2020).

Audio Generation. AIGCs can also generate audio, which falls into two primary categories: text-to-speech synthesis and voice cloning (Zhang et al., 2023b; Liu et al., 2024a).

- *Text-to-Speech Synthesis:* This process converts input text into speech that mimics a specific speaker’s voice, commonly used in robotics and voice broadcasting applications (Yuan et al., 2024).
- *Voice Cloning:* This technique takes specific target speech as input and transforms it to match the speech patterns of the designated speaker. Some applications automatically generate audio descriptions of images (Hasegawa-Johnson et al., 2017). For instance, WavJourney, developed by Liu et al., creates structured scripts for audio storytelling from text descriptions of auditory scenes (Liu et al., 2023). Another innovative application, AudioGPT, created by Huang et al., excels in generating diverse audio content, including speech, music, sound effects, and talking head tasks (Huang et al., 2024a,b).

Despite having a restricted role within this workflow, users remain essential. As observed, the human position is not static; instead, it depends on the objectives of collaboration and the balance of design effort within the system. Users’ flexibility and decision-making power tend to increase in subsequent integration levels.

2.3.1.5 Human-AI Collaboration making changes at the same time in the environment

This integration level shares the same objectives as the previous one; however, in this stage, agents have greater freedom to modify the environment through their actions, leading to more equitable collaboration, as illustrated in Figure 2.5. This strategy affords users enhanced flexibility. Furthermore, humans possess the autonomy to initiate actions without requiring explicit instructions. This results in a more balanced distribution of roles between agents, as the activities performed by humans closely resemble those executable by AI. Consequently, there is an increased freedom for humans and a greater

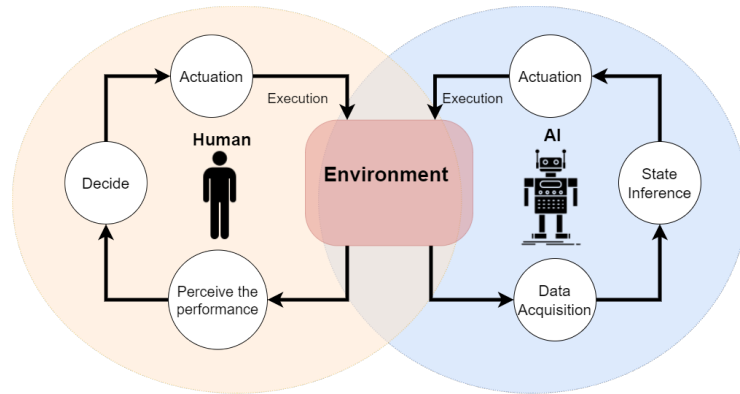


Figure 2.5: A visual representation of the integration level between humans and AI in collaborative efforts. Both humans and AI have the autonomy to modify the environment and must respond to each other’s contributions to achieve effective collaboration.

capacity for decision-making, which enhances the level of human integration. To foster collaboration, all participants must be aware of each other’s actions to inform their subsequent responses, thus increasing overall engagement and awareness. Table 2.4 summarizes examples of collaborative projects in this group, highlighting the key characteristics shared among them.

One notable example of this integration level is the algorithm proposed by Holzinger et al. They introduced a solution for optimizing the Travelling Salesman Problem (Holzinger et al., 2017) using the behaviour of the Ant Colony Optimization algorithm (Deng et al., 2020). In this study, the authors provided users with two innovative tools for modifying the ants’ behaviour: a human interaction matrix that enables control over the ants’ movements between nodes and a human impact factor that allows users to define the probability that an ant will follow a human suggestion. This empowers users to act autonomously within the system’s environment, as they can adjust the matrix and factors at their discretion.

Gómez-Carmona et al. explored how to develop a system that empowers users to take initiative (Gómez-Carmona et al., 2021). They designed an intelligent drink monitoring system linked to a customizable interface. This system records all user interactions with the smart bottle to analyse hydration patterns. Users are asked to label their actions to ensure the system accurately

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
(Cao et al., 2024) - Robot plan assistant	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(Gómez-Carmona et al., 2021) - Smart drink monitoring	☒	□	□	□	□	☒	☒	☒	□	□	□	□	☒	□	☒
(Holzinger et al., 2017) - Optimization algorithm	□	□	☒	□	□	☒	□	☒	□	□	□	□	☒	□	□
(Li et al., 2024d) - Emergency Indoor Patrolling	□	□	☒	□	□	□	☒	☒	□	□	□	□	☒	□	☒
(Lou et al., 2023) - Game	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(McCamish et al., 2017; Termehchy and Touri, 2015a) - Mutual understanding	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(Mehak et al., 2024) - Human-AI assemble Task	□	□	☒	□	☒	☒	□	☒	□	□	□	□	☒	□	☒
(Nikolaidis et al., 2017a) - Collaboration task	□	☒	□	□	□	□	☒	☒	□	☒	□	□	☒	□	☒
(Nikolaidis et al., 2017c) - Collaboration task	□	☒	□	□	□	□	☒	☒	□	☒	□	□	☒	□	☒
(Schelble et al., 2021) - Game; RL	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	□
(Strouse et al., 2021) - Game	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(Tao et al., 2021) - Collaboration task	□	☒	□	□	□	□	☒	☒	□	□	□	□	☒	□	☒
(Wang and Zhao, 2024) - Human-AI coordination	□	□	☒	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(Xing et al., 2024) - Leadership transition	□	□	☒	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(Xu et al., 2012) - Game	□	□	☒	□	□	☒	□	☒	□	☒	□	□	☒	□	☒
(Yu et al., 2023) - Game	□	☒	□	□	□	☒	□	☒	□	□	□	□	☒	□	☒
(Zhang et al., 2024a) - Collaborative tasks	□	□	☒	☒	□	☒	□	☒	□	□	□	□	☒	□	☒
(Zhou et al., 2021) - Tracking a trajectory	□	□	☒	□	□	□	☒	☒	□	□	□	□	☒	□	☒

Table 2.4: Some examples of Human-AI Collaboration works where user and AI agents interact in a physical environment. This table highlighting the main common characteristics of these group. (□: No; ☒: Yes)

tracks drinking gestures. To improve accuracy, the system allows users to record examples of their specific drinking movements at any time, augmenting the training dataset. Following an action, users can also verify whether the gesture was correctly classified. This two-way communication enhances system performance, while the customization options improve user comfort during the learning process and regulate their effort.

Cao et al. developed a strategy for robot actions to assist humans using the Theory of Mind (ToM) and reinforcement learning. In this approach, the robot observes user behaviour and communicates with them to facilitate task completion more effectively (Cao et al., 2024).

The primary goal of these works is to achieve cooperation between agents in real-world environments, rather than simply creating a plan or reaching a consensus. Another example of this objective is the work by Zhou et al., who developed a human-robot Cartesian co-manipulation task that navigates random obstacles, promoting mutual communication and accommodation in unstructured environments (Zhou et al., 2021). Due to the unpredictable barriers, robots face challenges in managing environmental uncertainty, and the task requires tracking specified trajectories while avoiding unexpected obstacles. This system maintains a high level of performance while minimizing cognitive load on users.

A similar study by Xu et al. builds upon previous research on human-human collaboration (Xu et al., 2012, 2009). They developed a Waiter Agent Interactive Training Experimental Restaurant (WAITER) robot that learns to collaborate with human trainers to provide customer service in a virtual restaurant. In this context, both agents (the robot and human) access different data and must communicate effectively to complete the task. Through this dialogue, agents can adapt their behaviours to align with those of their counterparts, employing various machine learning algorithms to facilitate this adaptability, including Bayesian networks and linear prediction.

However, collaboration members sometimes possess disparate information about the environment. For instance, Tao et al. developed a physical collaborative task involving a human and a robot, wherein both participants shared the same information (Tao et al., 2021). In their experiment, a human and a

robot worked together to manipulate a ball on a platform, aiming to place it in the center hole without letting it fall out. While both agents could observe the environment, they had to learn to anticipate each other’s intentions and coordinate their actions without direct communication.

When defining human-robot collaboration, it is crucial for AI to be trained using human data to understand how to respond to various situations. However, this training can impose a significant burden on humans. One method to streamline this process and reduce the human workload involves employing reinforcement learning (RL) strategies. For example, Strouse et al. explored how to train robots for effective collaboration with human partners without relying on human data (Strouse et al., 2021). They utilized a self-play strategy (Heinrich et al., 2015) to replace the human effort during training. Their study employed a cooking simulator game called *Overcooked*, where both agents work together to prepare recipes within a set time. Coordination is critical, as both agents share information and can select their subsequent actions. They utilized a V-MPO (Song et al., 2020), an RL algorithm featuring a ResNet plus LSTM architecture, to facilitate this coordination. Another example of human-AI collaboration within the *Overcooked* framework is the work by Zhang et al. (Zhang et al., 2024a), which developed a chat system enabling users and AI to discuss strategies for resolving the game.

Furthermore, human corrections to robot actions can demand substantial mental and physical resources (Nassar et al., 2024; Liu et al., 2024b). Consequently, some researchers focus on reducing human involvement while enhancing the robot’s capability to adapt its role in collaboration. For instance, Wang et al. developed a dynamic role-shifting system for human-robot collaboration in assembly tasks, aiming to alleviate human fatigue (Wang and Zhao, 2024). Similarly, Xing et al. devised a fuzzy logic-based arbitration rule to regulate the transition of the robot’s role in collaboration (Xing et al., 2024). Another approach by Mehak et al. involves creating a system capable of recognizing human actions and predicting forthcoming actions to foster improved team coordination and communication (Mehak et al., 2024). In addition to addressing human fatigue, maintaining the trust of human agents is

critical in collaboration. For example, Li et al. developed a human-AI collaboration system for indoor patrolling during sudden power outages, focusing on preserving human trust in the agent (Li et al., 2024d).

Moreover, Yu et al. employed self-play strategies to mitigate human biases in decision-making (Yu et al., 2023). They introduced the Hidden-Utility Self-Play (HSP) method, which explicitly models human biases as hidden reward functions. Similarly, Lou et al. utilized self-play strategies in conjunction with a policy ensemble method to enhance partner diversity within the population. Their context-aware approach enables agents to analyse and identify potential policy primitives of their partners, allowing for appropriate action selection (Lou et al., 2023).

The Overcooked game has been used for various research purposes, including studying human collaboration across different task loads in human-AI environments (Rosero et al., 2021; Wu et al., 2021). Despite its popularity, other games have also been employed in similar loops, utilizing RL techniques. The combination of gaming and RL is advantageous due to its inherent qualities (Tahboub, 2019; Qin and Yu, 2024). Therefore, it is crucial to examine how the selection of RL and game theory can foster varying levels of cooperation within hybrid teams. Schelble et al. conducted an empirical study to explore how different modern RL algorithms and game theory scenarios could influence cooperation levels in human-AI teams (Schelble et al., 2021). Their findings highlight the importance of selecting appropriate RL techniques to enhance team effectiveness and composition.

Additionally, Nikolaidis et al. designed a physical collaborative task in which a human and a robot must move a table through a door (Nikolaidis et al., 2017a). Initially, both agents must decide on their actions. This study emphasizes mutual adaptation, where each member must modify their behaviour based on the actions of the other (Nikolaidis et al., 2017b). They developed a Bounded-Memory Adaptation Model (BAM), an RL algorithm, alongside a Mixed Observability Markov Decision Process (MOMDP) policy to manage adaptability functions. In subsequent work, they improved mutual adaptability to preserve human trust, although the most efficient task solution was not achieved (Nikolaidis et al., 2017c).

In addition to mutual adaptability, other researchers focus on fostering mutual understanding. For example, Mccamish et al. aim to establish a mutual comprehension of human intent through queries (McCamish et al., 2017; Termehchy and Touri, 2015b). Using an RL method, they strive to create a shared understanding of human intent (Termehchy and Touri, 2015a). In this context, a query serves as a vital communication element between users and the database system, facilitating mutual understanding. Both humans and AI are rewarded based on their agreement, positioning them as equals with their respective methodologies for understanding and reasoning.

These examples illustrate that, in contrast to earlier groups, the degree of human freedom is significantly greater in this integration level. While all members share the collaboration’s goals, they actively cooperate to achieve them. The next level of integration involves utilizing humans as the reasoning element within the collaboration.

2.3.1.6 Humans replacing the artificial reasoner of the system

In this collaboration group, users act as the primary reasoning element of the system, effectively replacing the AI’s reasoning capabilities through an interface that equips them with the necessary tools to solve problems. The design of this tool varies depending on the specific problem, and the interface is automatically updated when users modify the data. This decision-update process is iterative, enabling users to engage in their own reasoning work. In this group, humans are the sole members with decision-making authority, creating an unbalanced dynamic within the system. Table 2.5 summarizes several examples of collaborative projects within this group, highlighting their key shared characteristics.

For instance, Kartoun et al. developed a classification system for clinical narrative notes to determine patients’ smoking status (Kartoun, 2017). In this project, the AI pre-processed the data by removing all non-smoking narratives and then utilized a random sample to allow humans to evaluate the selected instances.

Another example is provided by Wiethof et al., who created an interface that alerts customer service employees to tasks that may cause delays or be

Paper - Topic	Techniques				Expert	Interaction		Aim of collaboration					AI-human		Initiative
	AL	RL	aML	AIGC		Interface	Physical	Cooperation	Improvement	Customization	Consensus	Replace AI	1:1	1:n	
(Amershi et al., 2015) - Build a ML algorithm	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☐	☒	☒	☐	☐
(Song et al., 1999) - Build Decision Tree	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☐	☒	☒	☐	☐
(Kartoun, 2017) - Task classification	☒	☐	☐	☐	☐	☒	☐	☐	☐	☐	☐	☒	☒	☐	☐
(Talbot et al., 2009) - Build ensemble classifier	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☐	☒	☒	☐	☐
(Wiethof and Bittner, 2022) - Customer service	☐	☐	☒	☐	☐	☒	☐	☐	☐	☐	☐	☒	☒	☐	☐

Table 2.5: Some examples of collaborative works where human replacing the AI reasoner component. This table highlighting the main common characteristics of these group. (☐: No; ☒: Yes)

overlooked, thereby improving work efficiency (Wiethof and Bittner, 2022). This agent learns from users’ behaviours to provide timely alerts, but ultimate decision-making authority remains with the human agents. Both expert and non-expert customer service employees evaluate the agent’s recommendations.

In addition, Talbot et al. engaged users in constructing an ensemble classifier, which combines several classifiers to create one that outperforms its individual components (Talbot et al., 2009). To facilitate this process, they designed a visually interactive interface using the confusion matrix technique, enabling users to explore and combine individual classifiers effectively.

Similarly, Ankerst et al. developed a system that allows non-expert users to create a decision tree (Song et al., 1999). They employed a pixel-oriented technique known as Squarele Segments (Keim, 2000) for intuitive visualization of high-dimensional data, which enhances users’ understanding of the classifications and the data involved. Additionally, they designed a system where users can build a machine-learning algorithm through a visual interface (Amershi et al., 2015). In this case, users have the ability to label data, edit and select features, and view various statistical metrics, such as the confusion matrix, to assess model performance. While this approach enhances users’ comprehension of the output and the underlying data, it is important to note

that human capacity to detect patterns in large datasets is limited compared to artificial intelligence (Jirgl et al., 2018; Berger et al., 2021). This limitation underscores the necessity of achieving the right balance in collaboration, where each participant contributes according to their strengths.

Overall, this integration level is characterized by positioning humans as the reasoning element within the collaboration, granting them the highest level of decision-making authority compared to all other levels discussed in this work. Furthermore, an interface facilitates communication among members at this integration level. This framework emphasizes the importance of human agency in collaborative systems, highlighting that effective collaboration requires not only advanced tools but also a clear understanding of the roles and responsibilities of each participant.

2.3.2 Design Principles

This section defines the essential design principles that must be considered during the design phase of a Human-AI Collaboration system. The design principles that are not directly relevant to promoting such collaborations are excluded. These principles are described based on their impact on collaboration within the system. It is important to note that these principles are part of a broader set of considerations for effective, sustainable, and ethically sound technology design (Manion, 2002; Siau and Wang, 2020). However, the focus of this work is solely on those principles that are vital for establishing mutually beneficial collaboration where all participants feel comfortable, integrated, and actively engaged in the process.

Based on a review of the literature, the following design principles have been identified: Adaptation & Customization, Trust & Explainability, Engagement, and Communication & Feedback. A detailed description of each principle is provided, emphasizing their relevance to the integration of human participants. Moreover, practical examples illustrating the application of each principle in real-world scenarios are included to enhance understanding and implementation.

2.3.2.1 Adaptation & Customization

Adaptability is the capacity to alter behaviour based on an assessment of the environment. In a collaborative context, this involves modifying the initial approach according to the actions of other agents and the capabilities of each team member. Consequently, adaptability encompasses communication, perception, and an interactive process (Zhao et al., 2022). According to Li et al., the success of a collaborative system depends on the members' ability to work together and adjust their strategies to achieve mutually beneficial outcomes while completing tasks (Li et al., 2021a). This flexibility is essential for members to adapt to one another in a complementary manner. To demonstrate this, Li et al. developed an adaptive agent capable of identifying and selecting the most complementary policy based on the actions observed from the human partner. Their findings indicate that performance improved with the adaptive approach compared to non-adaptive agents. This experiment supports the notion that agents must complement each other by adapting their behaviours to enhance collaboration.

Furthermore, adaptability is essential for collaboration as it enhances human performance, reduces fatigue, and increases the effectiveness of outputs (Nassar et al., 2024; Liu et al., 2024b). This assertion was evaluated by Buehler et al., who explored the relationship between human reciprocal cooperation and its correlation with collaborative performance (Buehler and Weisswange, 2020). The results suggest that the cooperative behaviours of an automated agent can significantly influence human cooperativeness. Participants demonstrated greater resource sharing and improved performance when working with a high-cooperation agent compared to a low-cooperation agent. However, it is crucial for the automated agent to calibrate cooperation within the system to avoid inefficient use of resources.

The adaptation of one agent depends on the adaptability of the other. Many authors postulate that adaptation should occur for both human and AI agents (Nikolaidis et al., 2017a). A common approach is the mutual adaptation technique, which asserts that both agents must continuously adjust their behaviour based on each other's actions (Nikolaidis, 2017). Nikolaidis et al. focus on achieving mutual adaptation without compromising user trust,

aiming to balance task execution with computational accuracy and the maintenance of user trust (Nikolaidis et al., 2017a,c). Other authors offer different perspectives on this concept. For instance, Xu et al. examine adaptation from the agent’s viewpoint, suggesting that mutual adaptation could enable the agent to manage flexible mappings between human instructions and agent actions (Xu et al., 2012). They developed formal conditions and demonstrated that effective collaboration requires enabling the mutual adaptation phenomenon. Key elements include the continuous adaptation of each agent to the actions of the other and the necessity for agents to express their internal states. Both experiments, along with their respective developments, have proven superior to one-way adaptation strategies while maintaining user confidence in the progress of collaboration. This aligns with the work of Zhou et al., which establishes mutual adaptation in unstructured environments that promote mutual communication and adaptation (Zhou et al., 2021).

Despite these considerations, effective collaboration between humans and machines requires not only behavioural changes but also the accuracy of those changes. This knowledge is acquired through the co-learning process inherent in their interactions (Sawaragi, 2005). Consequently, an approach based on the co-learning process is essential for fostering effective human-AI collaboration. In this context, Van Zoelen et al. examine co-learning patterns in human-robot collaboration (Van Zoelen et al., 2021). They assert that the initial step of co-learning is co-adaptation, defined as partners adjusting their actions to align with both the task and each other. This research identifies a list of recurring interaction patterns applicable to stable situations as well as those requiring sudden adaptation (changes in behaviour occurring over short periods). In stable situations, patterns include actively coordinating actions or working while waiting for a team member. In contrast, sudden adaptation situations may involve confusion regarding the agent’s behaviour or strategy, attempts to communicate, or feelings of isolation without providing assistance to the agent. Alternatively, Tahboub et al. present a different perspective, advocating that co-adaptation and co-learning occur simultaneously during collaboration (Tahboub, 2019). They propose and demonstrate that reinforcement learning (RL) strategies, particularly those utilizing policy

gradients, effectively fulfil the requirements of co-adaptation due to their capacity for iterative learning from interactions in dynamic environments.

To facilitate mutual adaptation, it is essential to study humans' ability to adapt rather than solely focusing on the benefits of achieving mutual adaptation. Although several studies have concentrated on how robots should adapt during collaboration (Zeng et al., 2020), a new paradigm is emerging that seeks to understand how human adaptation evolves through interactive collaboration (Nikolaidis, 2017). The primary objective is to learn how this process occurs to incorporate this knowledge into the artificial agent, resulting in more adaptive behaviour during collaboration.

For instance, in addition to developing a navigation collaboration system in which each agent possesses partial information, Mohammad et al. investigate the evolution of human adaptation in collaborative settings (Mohammad and Nishida, 2008). Their findings demonstrate that the human adaptation rate is not fixed; it begins with a non-adaptive phase, during which users attempt to discover the robot's capabilities. Following this initial phase, users enter an adaptation phase, which is influenced by their perception of the robot's ability to complete the task. This second stage is particularly evident among users with high expectations. Consequently, the time required for users to adjust their behaviour is contingent upon their perception of the robot's capabilities. Similarly, Nikolaidis et al. develop a robot model that accounts for human perceptions of the robot's capabilities, allowing the robot to choose between actions that reveal its abilities to the user and actions that are optimal based on the available information (Nikolaidis et al., 2017b). This model adapts its actions based on the understanding that user adaptation is influenced by perceptions of the robot's capabilities. Their research indicates that considering the human perspective enhances human-AI collaboration compared to policies that assume the human will fully adapt to the robot. In another collaboration paradigm, Van Zoelen et al. study how humans adapt to continuous leadership shifts in human-robot collaboration and how this influences their trust and understanding (Van Zoelen et al., 2020). They demonstrate that, over time, humans develop a greater appreciation for collaboration when they as-

sume less of a leadership position. This decision is influenced by the human’s perception and confidence in the robot’s capacity to complete the allocation.

Customization is a fundamental aspect of human-AI collaboration, enabling the system to adapt to user preferences and requirements. This personalization can be achieved through various techniques, including user feedback, user modelling, and reinforcement learning, among others (Paripati et al., 2024). For example, ChatGPT utilizes prior user input to generate more customized responses (Hua et al., 2024; Wu et al., 2023b; Schulman et al., 2022; Tabari et al., 2024). The level of customization varies according to the context, goals, and learning techniques employed. Personalization has been extensively explored in the development of recommender systems (Li et al., 2023; Mysore et al., 2023; Radlinski et al., 2022). However, predicting user acceptance of novel recommendations remains a significant challenge. Additionally, customization in recommender systems may raise concerns regarding user privacy (Wang et al., 2024). Balancing diversity, user engagement, and privacy constraints is therefore essential. Techniques proposed by Kelly et al. focus on enhancing exploration without deviating from the system’s core objectives (Kelly and Bridge, 2006). In this context, recommender systems can adjust aspects such as suggestion timing, the number of recommendations, and other factors to maintain user engagement while ensuring a balance between exploring new preferences and adhering to the system’s goals (Rafter and Smyth, 2005). This adaptive strategy helps ensure that recommendations align closely with user preferences and system objectives, ultimately enhancing the overall user experience.

2.3.2.2 Trust & Explicability

Understanding the role of trust in engaging humans in collaboration, as well as how trust is influenced during interactions, is crucial for establishing and promoting effective and productive collaborative relationships between humans and AI (Ostheimer et al., 2021; Mehrotra et al., 2024). This understanding is vital throughout all phases of collaboration, beginning with the design of the system and continuing through its ongoing interactions with the user. For

instance, Yang et al. indicate that the level of human involvement in the machine learning (ML) development process significantly impacts their trust in the models (Yang et al., 2018). Additionally, in the interactive collaboration process, initial interactions play a critical role, as demonstrated by Tolmeijer et al. (Tolmeijer et al., 2021). Their findings suggest that if the initial interactions between humans and these systems fall below expectations, it can negatively affect user adoption. Trust, however, is not a static concept; it evolves over the course of interactions. In this process, it is essential for the system to maintain human trust without diminishing it (Nikolaidis et al., 2017c,b) or exceeding it beyond reasonable limits throughout the collaboration.

For these reasons, human trust is a fundamental factor in effective collaboration, particularly when assessing the appropriate level of trust required by users in the decision-making process (Janhunen et al., 2024; Lane and Klave, 2024). A lack of user trust can hinder their willingness to participate, negatively affecting the adoption of intelligent systems (Eiband et al., 2021). For instance, Yang et al. analyse how users perceive different tracking devices, noting that perceptions of accuracy can vary depending on individual expectations and personal circumstances (Yang et al., 2015). Furthermore, Zerilli et al. highlight that a user's limited understanding of how these devices function, including data collection and measurement calculations, can significantly influence their perception of accuracy (Zerilli et al., 2022). Therefore, human-AI systems should prioritize fostering effective and transparent collaboration and communication.

The concept of explicability and the emphasis on interactive processes stand in contrast to the "black box" nature that AI and machine learning (ML) systems may exhibit when human knowledge is not integrated into their problem-solving processes (Schmitt et al., 2024; Endsley, 2023; Chander et al., 2024). This lack of transparency has significantly influenced AI development by increasing human aversion and diminishing user trust (Scharowski, 2024; Wang and Ding, 2024). Providing explanations enables users to gain a clearer understanding of the machine's behaviour, identify potential limitations and errors in its reasoning patterns, establish a sense of control, and modulate their level of trust in the technology (Teso et al., 2023; Saeed and

Omlin, 2023). Empirical findings from Yang et al. suggest that incorporating interactive processes can lead to greater confidence among non-experts in learning outcomes, particularly when explainability is prioritized (Yang et al., 2018). This indicates that when users engage with ML systems, it is crucial for them to comprehend how the system reaches its conclusions in order to trust its decision-making processes and provide accurate feedback (Teso et al., 2023). Enhancing explicability without compromising computational performance presents a significant challenge for the development of AI-generated content (AIGC) technologies (Saeed and Omlin, 2023).

Achieving an appropriate level of explicability in AI systems is also essential. Overly detailed explanations can distract users and may negatively impact their trust in the system. Conversely, the inability to understand a given explanation can lead users to lose trust in the system prematurely and unjustifiably (Honeycutt et al., 2020). Therefore, finding the right balance between user engagement, the amount of explanation provided, and the necessary human effort is crucial to maintaining user trust within appropriate boundaries (Tutul et al., 2024).

Additionally, it is essential for communication tools within interactive systems to establish a bi-directional communication channel and to consider how these tools are presented to users (Teso et al., 2023). In this context, McCallum et al. investigated the role of user-driven feature representation in systems, revealing that users often have a limited understanding of the impact of their contributions to the model (McCallum and Fiebrink, 2019). Similarly, Amershi et al. examined the potential advantages of collaborative approaches that allow users to assess the quality of the model and decide on further inputs, enhancing both user experience and the learning system itself (Amershi et al., 2014). They concluded that further research into new interaction methods is necessary to determine whether user influence on the model can indeed lead to more effective systems. Meza et al. also emphasized the importance of evaluating different forms of interaction and their impact on user trust, highlighting the need for a deeper understanding of these dynamics at this stage (Meza Martínez et al., 2019).

In summary, the interactive nature of collaborative solutions contributes to building trust in AI systems by increasing user involvement in the decision-making process and enhancing their understanding of how AI functions (Vössing et al., 2022). Overall, factors such as user perception of accuracy, system learning progress, explainability, the level of user integration, and the interactive nature of collaboration significantly influence user trust in these systems. These aspects should be carefully considered in the development of future collaborative technologies.

2.3.2.3 Engagement

Lowering the barriers that hinder effective interaction between humans and technology is a primary challenge in promoting collaboration between them (Abdelghani et al., 2019). Indeed, the collaboration between humans and AI involves active participation to create a joint environment where they work together to achieve a common goal. However, motivating users to collaborate over an extended period can be challenging and requires considerable involvement. In addition**,** there may not always be a perceived value produced. For instance, in an experimental evaluation by Masson et al. on user involvement with AI systems, participants initially took an active role in data capture (Masson et al., 2016). Still, eventually, most of them quit due to the significant time required, poor accuracy, and perceived lack of reward. In this line, Ramos et al. delved into user willingness to interact with a deployed system and determined that it may vary with context, task, and individual characteristics (Ramos et al., 2019).

These concepts of involvement, interest, and attention relate to the idea of engagement, which is crucial in human-agent interaction as it is usually a prerequisite for the objective of the interaction (Glas and Pelachaud, 2015). Hence, engagement with AI systems is vital for comprehending the user's role and, consequently, enriching this relationship and enhancing user experience (Amershi et al., 2014). In this context, user engagement refers to the degree of involvement, interaction, and emotional connection a user has with an AI system or technology (O'Brien and Toms, 2008). From the AI agent's

perspective, engagement can be viewed as the goal of maintaining user interactions. From the human perspective, engagement goes beyond just the functional use of the system; it refers to the user's active participation and willingness to continue using the system (Wang et al., 2020). Oertel et al. reviewed previous efforts addressing the relevance of engagement in human-AI interaction from both perspectives (Oertel et al., 2020). According to them, knowing the level of user engagement can be advantageous for customizing agent behaviour and gauging the quality of interaction and user satisfaction with the system.

At the same time, the perception of relevance is also pivotal for promoting user engagement, as users need to feel that their inputs contribute to the system's value (Dudley and Kristensson, 2018). In this regard, reducing the effort to interpret outputs and express feedback about the potential outcomes of the provided inputs may enhance user engagement. For this reason, it is essential to maintain a good trade-off between the efforts made and the perceived value of such efforts. Therefore, in the context of human-AI collaboration, user engagement also encompasses the level of communication and transparency perceived from the AI system (Rezwana and Maher, 2022). These conclusions highlight the need for increased collaboration between AI/ML and HCI design research fields. In this line, other scholars have evaluated such engagement practices to reduce the effort needed to cooperate with the system. Those efforts aim to support the iterative design, development, and dissemination processes of learning systems oriented toward conversational AI agents (Zheng et al., 2022) or to define a human-centered thinking approach to applying IML methods (Mathewson, 2019).

In essence, user engagement is critical in ensuring that the collaboration between humans and AI is effective, efficient, and beneficial for both parties. An effective system needs to actively engage users in the task being performed to motivate them to achieve the desired outcomes of the human-AI collaboration.

2.3.2.4 Communication & Feedback

The more effectively agents comprehend each other’s capabilities, limitations, inputs, outputs, and context, the more efficiently they can collaborate to achieve their shared objectives (Holder et al., 2021; Zhang et al., 2023e). As Marathe et al. concluded, AI systems must be capable of understanding the state and intent of their human co-workers; simultaneously, humans must clearly articulate their intentions (Marathe et al., 2018). Furthermore, it is crucial for humans to recognize the limitations of AI; otherwise, they may develop unrealistic expectations regarding AI’s capabilities, which can negatively impact their engagement, adaptation, and trust (Riveiro and Thill, 2021; Milani et al., 2024).

Effective collaboration necessitates bi-directional communication to establish a mutual understanding that enables the participants to perceive each other through natural, fluid interactions (Schaefer et al., 2018; Breckner et al., 2024). In a formal sense, communication refers to a reciprocal exchange in which team members share messages to coordinate ideas, norms, and strategies to achieve their collective goals (Salas et al., 2015; Cooke et al., 2013).

Several scholars have explored strategies to foster this bi-directional communication in human-AI systems. For example, Holder et al. provide an overview of best practices and challenges in establishing bi-directional transparency at various stages of the human-AI-robot systems (HARTs) lifecycle (Holder et al., 2021; Duan et al., 2023). Similarly, Koop et al. apply principles of human-to-human communication as the foundation for developing effective human-AI interactions (Kopp and Krämer, 2021). They argue that agents should engage in cooperative and incremental dialogue construction with human users, supported by AI’s coordination mechanisms. They advocate for the next phase in human-AI communication research to focus on developing agents with the capacity to intertwine incremental metacognitive processes with socio-communicative behaviours, enabling them to co-create seamless interactions and rectify misunderstandings, misinterpretations, and dialogue interruptions.

Building on this concept, McCamish et al. have introduced a collaborative communication approach that implements a formal framework for represent-

ing and understanding information needs in database querying, aiming to develop a mutual language for representing these needs (McCamish et al., 2017; Termehchy and Touri, 2015b). This mutual language allows users to articulate their intentions precisely when querying databases and simultaneously provides feedback mechanisms that help users understand whether their intentions align with the AI’s responses. Their findings indicate that users tend not to explore alternative methods once they discover a reliable way to express their intentions. Moreover, Hanna et al. developed a communication model called Speech Act Theory (SAT), which explores how utterances can facilitate actions in collaborative dialogues (Hanna and Richards, 2019). This model also examines the influence of human perceptions of AI speech acts on collaborative performance, showing that a positive human perspective is associated with the AI’s ability to articulate its intentions in a comprehensible manner, thereby promoting mutual communication.

When addressing communication scenarios, the agent’s initiative, or proactive communication capabilities, is also a critical factor. Zoelen et al. developed proactive communication techniques to enable more fluid human-AI collaboration by using context and adaptability to guide communication strategies (Zoelen et al., 2020). They demonstrated that technology trained to communicate proactively outperforms those without such capabilities, especially in dynamic environments like disaster scenarios or military decision-making, where adaptability and re-planning are essential (Van Den Bosch and Bronkhorst, 2018). Ramchurn et al. further exemplified this concept by showing that proactive agents empower users to decline tasks, request information, and communicate preferences, while also allowing the AI to inquire about task statuses as needed (Ramchurn et al., 2016). This dynamic enables both parties to initiate communication when appropriate, fostering a deeper understanding of each other’s perspectives.

Chatbots, as conversational agents, have evolved significantly in recent years, enhancing their ability to simulate human interactions through text, voice, and visual communication (Aggarwal et al., 2023; Saeed and Omlin, 2023). These advancements have led to personalised and engaging conversational experiences that can provide on-demand health interventions, among

other services (Kuhail et al., 2023). As AI technologies continue to progress, we can anticipate even greater improvements in chatbot capabilities, enabling them to handle increasingly complex tasks and offer more valuable user assistance (Sharma et al., 2023).

Feedback is a crucial element in human-AI collaboration, serving as the primary mechanism through which each agent communicates its internal state in response to its partner’s actions. Selecting the appropriate feedback type for the human collaborator is essential for fostering co-adaptation and co-learning within the partnership (Zbyszynski et al., 2019). Knox et al. illustrated the significance of robot feedback through their work on TAMER, a robot that learns behaviours based on human numerical feedback in navigation tasks (Knox et al., 2013). Other studies have examined the role of feedback in dialogue, emphasizing the importance of providing users with clear indications of progress, which helps to meet their expectations during collaborative interactions (Pérez-Quñones and Sibert, 1996). These studies identify several feedback states-ready, processing, reporting, busy-no-response, and busy-delayed-response-that AI systems can simulate to manage communication effectively.

Finally, it is important to consider how feedback influences an AI’s behaviour relative to its current contextual knowledge versus the information provided by the human collaborator. Misinterpretations in feedback can arise due to the user’s personal circumstances rather than the AI’s actual performance. Therefore, effective communication strategies are vital for resolving these misunderstandings and correctly interpreting the feedback to enhance the collaborative process.

2.4 Summary and Conclusions

This thesis explores the integration of Human-AI Collaboration in the context of Citizen Science and Spatial Crowdsourcing. This chapter presents a comprehensive review of the literature on Spatial Crowdsourcing, Citizen Science, and Human-AI Collaboration workflows and design principles. The review was divided into three main sections: Citizen Science, Spatial Crowdsourcing, and Human-AI Collaboration.

The Citizen Science section provided an overview of the Citizen Science field, highlighting its importance, challenges, and opportunities. The Spatial Crowdsourcing section discussed various task allocation strategies, focusing on those designed for altruistic participation contexts. Many of these solutions include an economic optimisation component. However, the altruistic nature (unpaid) of participants is often not considered in the task allocation strategies. Therefore, there is a significant need to develop task allocation strategies that prioritise enhancing the user experience, thereby fostering more sustained and effective participation that is not solely based on financial rewards. This thesis addresses this challenge and others related to altruistic participation in the Spatial Crowdsourcing model, such as potential decreases in participation or user overload.

The Human-AI Collaboration section examines the literature in this field to illustrate how collaboration can create new and better solutions that consider users as active participants in the system. Since Citizen Science requires tailored solutions that involve users without overloading them (Vohland et al., 2021), Human-AI Collaboration is used in this thesis to create a better system that considers users as a key part of the system. Hence, the Human-AI Collaboration section examined a classification of collaboration workflows based on several characteristics to understand the diverse ways to integrate users into the solution. Additionally, it analysed the design principles of Human-AI Collaboration to understand how to create a better system that considers the user as a key part of the system.

Together, these sections provided a comprehensive understanding of the challenges and opportunities in the Spatial Crowdsourcing field within the Citizen Science context. They highlighted how integrating Human-AI Collaboration can address these challenges by engaging users and creating adaptive solutions. This approach ensures that the system is better aligned with the needs and preferences of participants, ultimately enhancing the effectiveness and sustainability of Citizen Science projects.

*Science has not yet taught us if madness is or is not the
sublimity of the intelligence.*

Edgar Allan Poe

CHAPTER

3

Spatial Crowdsourcing Model

The objective of this chapter is to describe a model for the Spatial Crowdsourcing (SC) experiments conducted in this thesis. This model is designed to generate a spatio-temporal data map that meets the requirements of Citizen Science (CS) initiatives. This model has to define the spatial and temporal constraints associated with each task and specify the number of tasks needed to achieve the initiative's goals, that is, the tasks-completion requirements.

Given the diverse nature of Citizen Science experiments, it is imperative that this model possesses sufficient flexibility to accommodate various contexts of Citizen Science. For instance, in a Citizen Science experiment focused on measuring air quality in an urban setting, tasks will entail assessing air quality at multiple locations throughout the city. Conversely, if the experiment is focused on counting birds in a designated area, the tasks will involve tallying the bird population within that specified region. Consequently, the model developed in this chapter must be adaptable and agnostic to the experiments' domain, thereby ensuring that the types of tasks remain independent from the core formulation of the model.

Similarly, the model must support varying tasks-completion requirements. For instance, measuring air quality requires repeated measurements over time at the same location to calculate averages and assess trends, whereas identifying non-accessible areas for people with mobility disabilities may need only one or two tasks to label each location. The model should be capable of adjusting to the different tasks-completion requirements.

While more data generally improves accuracy and data quality, in Citizen Science initiatives, we must consider the human effort involved. Unlimited demands on users can lead to overloading, decreased participation, or a decline in users' motivation. Additionally, as financial incentives may not be feasible for all Citizen Science initiatives, the model must avoid imposing an unsustainable workload as part of the goal of creating a positive user perception. Hence, to prevent user fatigue and promote sustained participation, it is essential to regulate task completion requirements with participants' capacities. Thus, the model must be flexible regarding the number of users and their motivation levels, including limits on workload when needed.

Finally, the model must accommodate diverse spatial and temporal constraints specific to each Citizen Science experiment. Some experiments may require data collection at specific times of the day or week, while others may benefit from random task intervals. The model must be able to handle these constraints and adapt to each experiment's unique requirements.

In conclusion, the model must generate a spatio-temporal data map, define spatial and temporal constraints, and specify tasks-completion requirements. It must accommodate various Citizen Science contexts, support different tasks-completion needs, and prevent user overload to maintain participation and motivation.

The chapter has the following structure: Section 3.1 details the model formulation, including the spatial and temporal constraints, the tasks-completion goals, and the workload limitations. Finally, Section 3.2 summarizes the key findings of this chapter.

The fundamental definitions pertinent to this thesis are compiled at the glossary titled **Key Concepts**. Likewise, the mathematical notation employed throughout this chapter is detailed at the glossary titled **Mathem-**

atical Notation. Readers are encouraged to consult these glossaries as necessary to ensure a comprehensive understanding of the material presented in this chapter. The most critical definitions in this thesis are highlighted in italics to signify their importance. Detailed explanations of these terms can be found both at glossaries and within the text of this chapter.

3.1 Mathematical Model

In this section, model for Spatial Crowdsourcing experiments is formulated. This solution can be applied to multiple Citizen Science initiatives. As part of our goal, this formulation must integrate a workload limitation to improve the user perception and prevent overloading. Table 3.1 lists variables used in the model formulation.

Variable	Definition
C_d	The distance between the centres of two adjacent cells.
m	Number of cells of a campaign.
L_j	The j -th cell of a campaign.
L	Set of cells of a campaign.
d_s	Duration of the time slots (except the last).
n	Number of time slots of a campaign.
S_i	The i -th slot of a campaign.
S	Set of slots of a campaign.
$St(S_i)$	Starting time of the slot S_i .
$Ft(S_i)$	Ending time of the slot S_i .
M_s	Maximum number of completed tasks allowed for each cell within each slot.
$t_{j,i}^k$	The k -th task of the cell (L_j) at slot S_i .

Table 3.1: Glossary of the variables utilized in the Spatial Crowdsourcing model presented in this thesis. The mathematical notation utilized in this chapter is also explained at glossary titled **Mathematical Notation**.

In this thesis, the process of accomplishing tasks, that is the *tasks-completion procedure*, is designated as a *campaign*. The campaign includes a spatial component, defined as an area, which is circular by default in this thesis. However, the model is flexible enough to accommodate diverse area geometries.

This circular area is centred on a specific point and has a specified radius. This area is further subdivided into a grid of squares, called *cells*, for task distribution. This subdivision is based on the input parameter C_d , which defines the distance between the centres of two adjacent cells. By utilizing this parameter and the defined circular area, a grid is created. Each campaign contains at least one cell within the entire campaign area. Figure 3.1 illustrates the spatial component of a campaign and how the cells are distributed within this area. Tasks must be performed at the centre of these cells. Based on this initial shape, to create new shapes for other Citizen Science initiatives, some cells can be removed to achieve the desired form. Examples of other shapes are shown in Figure 3.2 and additional examples of campaign shapes can be found in Appendix A. The cells remain consistent throughout the campaign duration since it is a **static** task allocation model. The total number of cells in a campaign is denoted by m , the j -th cell is represented as L_j , and L denotes the set of all cells within the campaign (Eq. 3.1). In the rest of this dissertation, the L_j cell is referred as j to simplify the notation.

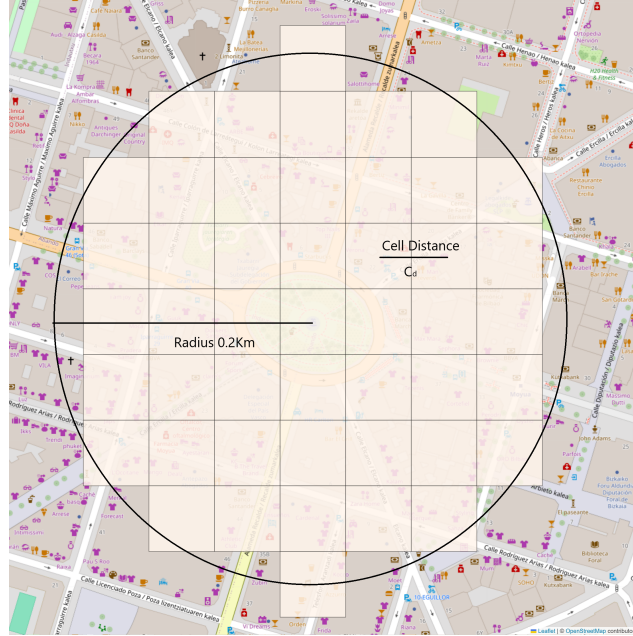
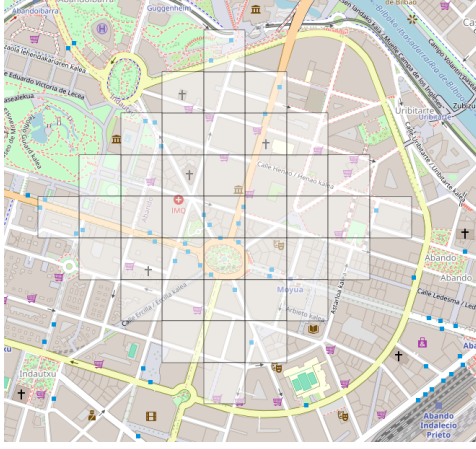
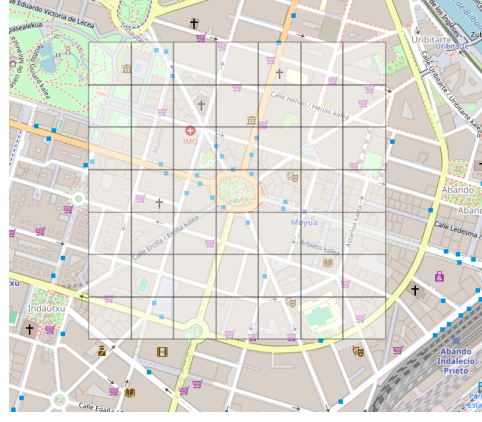


Figure 3.1: An example of a campaign created in Bilbao. In this picture, the distribution of the cells are shown.



(a) An example of a campaign with a diamond shape.



(b) An example of a campaign with a rectangular shape.

Figure 3.2: Several campaigns with different shapes.

The campaign's duration is divided into sub-periods called *slots*. All campaigns always have at least one slot. The number of slots is calculated based on the input d_s , which defines the slot duration. Hence, the duration of one campaign is divided into sub-periods of equal length, except for the last slot, which may span a shorter duration. Based on this input, the number of slots is defined and denoted as n . The i -th slot is described as S_i , and S is the set composed of all the campaign slots (Eq. 3.2). The following equations show the formulation of the sets L and S , where L is composed of the campaign cells, and S is composed of the campaign's time slots.

$$L = \{L_0, \dots, L_{m-1}\}; \quad m \geq 1 \quad (3.1)$$

$$S = \{S_0, \dots, S_{n-1}\}; \quad n \geq 1 \quad (3.2)$$

As this model is a time-dependent problem, each slot, S_i , is associated with two times: the time it starts, $St(S_i)$, and the time it ends, $Ft(S_i)$. Based on this notation, the start of the campaign can be expressed as $St(S_0)$, and the final time as $Ft(S_{n-1})$. The problem must satisfy three temporal constraints: (Eq. 3.3) the start time of a slot must occur before its end time; (Eq. 3.4) the duration of all slots is the same, d_s , except for the last slot, which may

be shorter; and (Eq. 3.5) the ending time of one slot must coincide with the beginning time of the next slot.

$$St(S_i) < Ft(S_i); \quad \forall i \in \{0, \dots, n-1\} \quad (3.3)$$

$$Ft(S_i) - St(S_i) = Ft(S_0) - St(S_0) = d_s; \quad \forall i \in \{0, \dots, n-2\} \quad (3.4)$$

$$Ft(S_i) = St(S_{i+1}); \quad \forall i \in \{0, \dots, n-2\} \quad (3.5)$$

Limiting the workload is paramount to enhancing user perception. Since participation is altruistic, avoiding overloading users is essential to sustaining engagement. Hence, in this work, each campaign is associated with a fixed number that defines the maximum number of allowed completed tasks in each cell at each slot, denoted as M_s . This variable defines the *workload limitation* of the campaign. In the case that a campaign consists of only 15 cells and the variable M_s is set to 8, the maximum number of tasks that can be performed per slot is 120. Specifically, the model requires that 8 tasks be completed in each cell per slot. The Spatial Crowdsourcing model defined in this context is not designed to allow the execution of a greater number of tasks in one cell within the same slot. Once a slot is completed, the task count is reset and the process begins again in the next slot. In this thesis, although the tasks are defined by the Citizen Science problem, all tasks are essentially the same; they are merely carried out at different times or locations.

The variable M_s depends on the campaign and its specific requirements. When the task involves capturing information by a sensor, multiple sensor measurements (tasks) are needed to ensure that the average value is representative, as the sensor readings can vary based on measurement conditions. If the tasks are unrelated to sensor measurements, it may be sufficient to perform only one task per cell. The idea is to use the M_s tasks of each cell to obtain a representative value for the cell on this slot. These *representative values* compose the final data time series map. Consequently, only M_s tasks can be realised at each location of L for each slot of S . Thus, a task can be represented as $t_{j,i}^k$, which is the k -th task of cell j at the slot S_i . The complete set of tasks for the campaign is formally defined in Eq. 3.6. The total

set of tasks defined by this equation outlines the *tasks-completion goals* of the campaign, representing the tasks-completion effort users must undertake to complete the dataset. It should be noted that, for this model, all tasks are of the same type.

$$\{t_{j,S_i}^0, \dots, t_{j,S_i}^{M_s-1}\}_{j \in L}; \quad M_s \geq 1 \quad (3.6)$$

A *completion colour scale*, presented in Figure 3.3, was designed to visually represent the progress of the campaign toward its task-completion goal. This scale consists of five colours, each associated with a range of completed tasks based on the campaign's M_s variable. To represent the progress of tasks within a campaign, each cell is coloured according to the number of tasks completed within it. For example, if the M_s variable is 8, cells with one or two completed tasks will be assigned the lighter colours of the scale. Only when a cell reaches 8 completed tasks will it be coloured with the darkest brown of the scale. A visualisation of the campaign (as Figure 3.4) using this completion colour scale can be generated at any point during the campaign, providing a clear overview of the progress toward the goal. The aim is to reach the M_s value in each cell, and when all cells display the darkest brown colour, it indicates that the goal of M_s tasks has been achieved across all cells. Figure 3.4 illustrates the **campaign representation** using this completion colour scale. In this example, each cell is coloured based on the number of tasks completed and labelled with the corresponding task count.

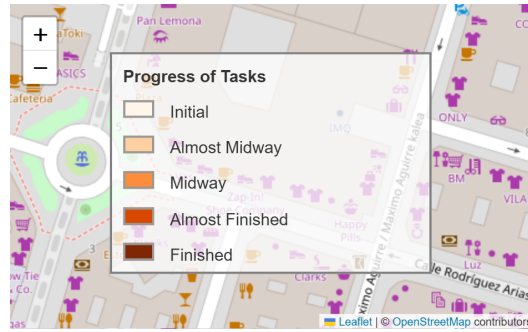


Figure 3.3: Completion Colour scale used to represent the progress of the campaign towards its task-completion goal.

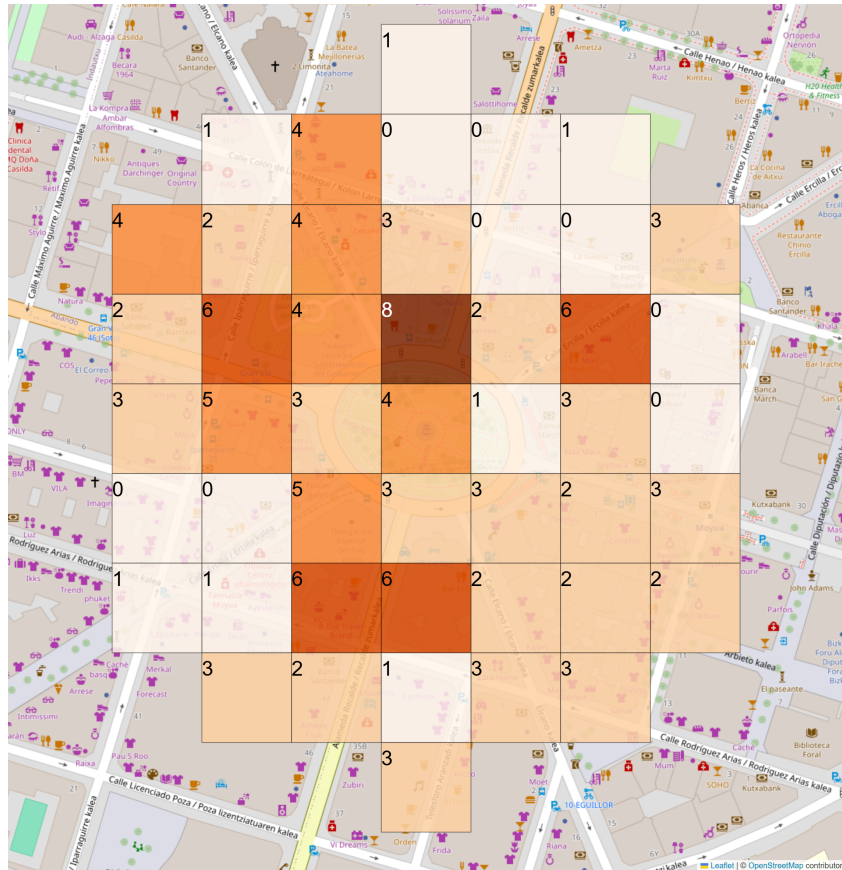


Figure 3.4: A campaign representation using the completion colour scale defined.

3.1.1 Structure of the Upcoming Chapters and Their Relation to the Iterative Process presenting incremental Spatial Crowdsourcing strategies

The following chapters of this thesis will introduce three spatial crowdsourcing strategies, each designed to align with the model defined in the previous section. These strategies are developed iteratively, aiming to progressively increase user participation at each stage while ensuring that the proposed hypothesis is met. This iterative approach means that the results obtained from each strategy will serve as the foundation for improving the next one. In other words, the second strategy will be developed by incorporating feedback and insights gained from the first, while the third will integrate the improvements derived from the second.

To facilitate the reading and understanding of this thesis, Figure 3.5 provides an outline of the structure of the next three chapters, which correspond to the proposed strategies. Chapter 4 introduces the VTAS strategy, Chapter 5 presents Bio-VTAS, and Chapter 6 is dedicated to Hybrid-VTAS. Each strategy is explained in detail in Sections 4.1, 5.1, and 6.2, respectively. To avoid redundancy and enhance clarity, some key concepts are presented only once and then assumed in subsequent chapters. The general objectives that these strategies must fulfil are explained in Section 4.1 and will not be repeated in later chapters. The initial baselines are detailed in Subsection 4.2.1, and as new strategies are developed, they will also include previously explained approaches as reference points. Similarly, the evaluation metrics are introduced in Subsection 4.2.2 and will not be explained in detail again in later chapters. The use cases employed to evaluate these strategies are described in Subsection 4.2.3. These encompass two main types: the Simulation Use Case and the Deusto Use Case. The methodology for the first use case is explained in Section 4.3, while its evaluation can be found in Sections 4.3, 5.2, and 6.3 for VTAS, Bio-VTAS, and Hybrid-VTAS, respectively. The second use case, the Deusto Use Case, is applied only to VTAS and Bio-VTAS. Its initial description, along with the Telegram bot used for its implementation, is provided in Section 4.4 and Appendix C. The evaluation of this use case

for VTAS and Bio-VTAS of this use case is presented in Sections 4.4 and 5.3, respectively. Finally, the conclusions for each strategy are found in Sections 4.5, 5.4, and 6.4 for VTAS, Bio-VTAS, and Hybrid-VTAS, respectively.

All this information is summarised in Figure 3.5, which serves as a visual guide to understanding the structure of the following chapters and the logical flow of the thesis. With this image and explanation, it is expected that the reader will be able to follow the progression of the research more easily and understand how the various elements of the proposed strategies are interconnected.

		Chapter 4	Chapter 5	Chapter 6
		VTAS	Bio-VTAS	Hybrid-VTAS
Algorithm Overview		Section 4.1	Section 5.1	Sección 6.2
GOALS		Section 4.1		
Baselines		Subsection 4.2.1: LLN & NN		
			VTAS	Bio-VTAS
Evaluation metrics		Subsection 4.2.2		
Use case Details		Subsection 4.2.3		
Simulation Use Case	Simulation explanation	Section 4.3		
	Simulation evaluation	Section 4.3	Section 5.2	Section 6.3
Deusto Use Case	Telegram bot	Section 4.4, Appendix C		
	Deusto Evaluation	Section 4.4	Section 5.3	
Summary and Conclusions		Section 4.5	Section 5.4	Section 6.4

Figure 3.5: Outline of the Sections and Structure of the Following Three Chapters.

3.2 Summary and Conclusions

In this chapter, the Spatial Crowdsourcing model, designed to generate a spatio-temporal data map, is explained. The model offers several customizable features, including the maximum number of tasks that can be completed per time slot and cell, the number and duration of available time slots, the experiment area, the total number of cells, and their respective sizes. This

flexibility allows that the tasks-completion requirements can be adjusted based on the number of participants, their motivation levels, and the unique characteristics of each Citizen Science experiment. The type of tasks also depends on the specific Citizen Science initiative, but for this model all tasks are the same type of task. This model is adaptable across various Citizen Science contexts, allowing it to be tailored to the specific needs of each initiative. The proposed model is flexible, accommodating different tasks-completion requirements while limiting workload to prevent participant overload, thereby supporting sustained long-term engagement. This adaptability enables the creation of campaigns that can be adjusted to satisfy both user and campaign needs.

This model has been deployed as a microservice offering a RESTful API, called MVE, which is publicly available as open source in a Github¹ repository. RESTful APIs (Representational State Transfer APIs) are a standardized approach to building web services that allow seamless communication between clients and servers by adhering to stateless architectural principles. This design enables the API to efficiently handle requests and responses using standard HTTP methods such as GET, POST, PUT, and DELETE. In the context of MVE, the API facilitates the creation and management of campaigns by providing endpoints for defining their parameters and structure.

The API acts as the backbone for interacting with the Spatial Crowdsourcing model, serving as a bridge between the problem domain and the spatial task allocation strategies that are explained at the following chapters. It provides the necessary operations for initializing campaigns and retrieving their configurations, ensuring the system is flexible and scalable. Alternative strategies for tackling the task allocation problem will be explained in the following chapters. Such strategies are designed to work with these campaigns generated via the API. It ensures that tasks are efficiently assigned to users while leveraging the Spatial Crowdsourcing model. The API defines a set of endpoints to configure, run and retrieve results of Spatial Crowdsourcing campaigns. Based on the core Spatial Crowdsourcing model, diverse strategies,

¹<https://github.com/mpuerta004/RecommenderSystem>

representing the soul of the micro-volunteering engine (MVE), are activated, depending on the context and needs of the sought Citizen Science campaign.

In each of the following three chapters, alternative Spatial Crowdsourcing strategies are proposed based on the model defined in this chapter. Each strategy builds upon the previous one and seeks to achieve significant improvements to meet the research objectives outlined in this thesis and to validate the proposed hypothesis.

*The solution to a problem often reveals the simplicity
that was always there, waiting to be understood.*

Robert A. Heinlein

CHAPTER

4

Volunteer Task Allocation Strategy: VTAS

Spatial Crowdsourcing techniques are not inherently well-suited for contexts such as Citizen Science due to a fundamental misalignment between the two fields. This misalignment stems primarily from their contrasting incentive structures and user engagement strategies. While Spatial Crowdsourcing typically relies on financial incentives to drive participation, Citizen Science thrives on volunteers motivated by altruism, personal interest, or community engagement. Traditionally, Spatial Crowdsourcing strategies have prioritized achieving measurement objectives, often overlooking the most critical component of Citizen Science: the volunteers—the human participants who power these initiatives. **Addressing this fundamental difference requires a paradigm shift in task allocation strategies, moving away from economic models toward user-centred approaches that account for individual preferences and behaviours, ensuring sustained engagement without monetary rewards.** As previously discussed, these strategies must actively involve users

in decision-making, balancing task completion efficiency with fostering a positive user experience to encourage greater acceptance and adoption of Spatial Crowdsourcing approaches. Additionally, they should prevent user overload by respecting individual limits and avoiding reliance on financial incentives. This thesis proposes three collaborative strategies that not only enhances the acceptance and adoption of these technologies but also ensures an efficient approach to the Spatial Crowdsourcing problem within the context of Citizen Science. Thus, these Spatial Crowdsourcing strategies must achieve a balance between user experience and overall efficiency. The primary objective is to develop a solution that not only improves efficiency but also enhances user acceptance and adoption—an aspect traditionally overlooked in Spatial Crowdsourcing strategies. **By addressing this gap, the proposed approach seeks to bridge the disconnect between Citizen Science and Spatial Crowdsourcing, ensuring both usability and effectiveness in task allocation. In doing so, it aims to validate our hypothesis that Human-AI collaboration can enhance acceptance and adoption by balancing efficient task completion with a positive user experience, without relying on economic incentives.**

To this end, it is important to recall that the Spatial Crowdsourcing model underpinning these strategies is detailed in Chapter 3, which outlines the spatial and temporal constraints, task completion objectives, and workload limitations inherent to Spatial Crowdsourcing campaigns.

The first strategy of this dissertation is proposed in this chapter, marks an initial step towards a more user-centric approach while building on a traditional model. To address this, the chapter introduces the Volunteer Task Allocation Strategy (VTAS), a strategy designed for the Spatial Crowdsourcing model that prioritises altruistic participation over financial incentives. VTAS begins by engaging users at a basic level, progressively incorporating users' feedback to refine the approach in subsequent iterations (chapters). These refinements, detailed in the following chapters, aim to achieve a better balance between efficiency and users' perspective while deepening user involvement in decision-making. Through this iterative process, the solution evolves to

better address users' needs, empowering them to play a more active role and ensuring continuous adaptation based on ongoing evaluations.

Hence, this chapter offers a detailed examination of the VTAS and its evaluation, with the insights gained forming the basis for developing more user-centric strategies in subsequent chapters. These new strategies will be informed by the evaluation of VTAS, which is based on both simulation and real-world assessment. The chapter is organized as follows: Section 4.1 presents the proposed collaborative solution: the VTAS. Section 4.2 outlines the baseline algorithms, metrics and use cases employed to evaluate VTAS's performance. Section 4.3 details the simulation use case, which assesses VTAS's performance in a controlled environment, while Section 4.4 presents a real-world use case evaluation. Finally, Section 4.5 provides a summary of the key findings of this study. To better understand the structure of this chapter and its relationship with the following ones, it is recommended to refer to Figure 3.5. This image provides an overview of the sections and key concepts that will be explained here and assumed in the following chapters.

Be reminded that the **Key Concepts** glossary contains the most important definitions of this thesis, and the **Mathematical Notation** glossary provides explanations of the mathematical notation used in this thesis. The reader is encouraged to consult these glossaries as needed for a better understanding of the content presented in this chapter. The most critical definitions in this thesis are highlighted in italics to signify their importance.

4.1 Algorithm Overview

This section explains the first task allocation strategy devised in this dissertation, the Volunteer Task Allocation Strategy (VTAS). This solution aims to match users with tasks while achieving the following goals efficiently. These goals have been defined based on the hypothesis to be tested in this thesis and will serve as the reference for all proposed strategy. Thus, any algorithm that fails to meet these goals will not be considered a valid strategy for verifying the hypothesis of this thesis. The goals are as follows:

Goal 1: Enhance User Perception: This solution seeks to cultivate a positive user perception of the system by adhering to two fundamental principles: (1) respecting the workload limitation established in the Spatial Crowdsourcing model (Chapter 3), thereby ensuring that the campaign's task completion objectives are not exceeded. The fulfilment of this first condition confirms that users are not overloaded with tasks during the campaign's execution. (2) Implementing a user-centric task allocation process that actively involves users in decision-making. Thus, this user-centred solution, which prevents task overload and respects workload constraints, aims to enhance user perception, ultimately fostering greater acceptance and adoption of the technology. Consequently, all proposed solutions (strategies) in this thesis must satisfy these two conditions and contribute to promoting a positive user perspective to achieve this goal.

Goal 2: Optimised Task Allocation Solution: The solution aims to maximise task completion without surpassing the campaign's task-completion goal and the workload limitation. To achieve this, tasks should be allocated to volunteers flexibly and dynamically.

Goal 3: Achieving a Functional Balance Between the First Two Goals: The solution must recognise and respect user participation, understanding that, in some cases, meeting the target number of tasks (M_s) per cell at each slot may not be feasible. However, the task completion process must remain efficient, even under such circumstances. To address this, the solution should implement a task distribution strategy that ensures the collected data remains statistically robust and reliable. This involves maintaining comparability between the representative values derived from different cells within the same slot. As described in Chapter 3, these representative values are calculated from data gathered across all tasks completed in a given cell in the same slot, and the final campaign map (the final spatio-temporal data map i.e the final dataset) is constructed from them. Suppose one cell's representative value is based on a small sample (fewer tasks), while

another cell's is based on a larger sample. In that case, the datasets may lack statistical compatibility, undermining the final map's reliability. To prevent such inconsistencies, the solution must ensure an **equitable task distribution** across all cells throughout each slot. This ensures that even in cases of reduced participation, the data will retain statistical validity, and the task-completion process remains efficient. In conclusion, the solution must strike a balance between facilitating user participation and ensuring high-quality data collection, which is achieved by **maintaining an equitable distribution of tasks across all cells**.

The following section explains how the VTAS was designed to achieve these goals.

To address the second constraint of the Goal 1, the VTAS adopts a user-centric approach, working as a recommender system where users must first express their intention to participate. Upon processing each request (i.e. for each cell-recommendation issued to a user), the VTAS generates a set of N cells, referred to as *cell-recommendations*, tailored to each query. From these N cell-recommendations, the user is required to select one cell in which to perform a task, thereby committing to complete it. By recommending cells instead of tasks, this approach ensures that the user receives N distinct locations in which to participate. It prevents the user from being assigned multiple tasks related to the same cell, so the VTAS allows users to choose the cell that best align with their preferences. As a result, the same cell may be recommended to multiple users, but it will not appear more than once in a single request, meaning a recommendation issued to a user. The number of recommended cells is limited to N to minimise cognitive choice overload effect (choice difficulty) by avoiding too many options (Bollen et al., 2010; Beierle et al., 2017).

Since it cannot be assumed that all accepted tasks will be completed, the system must reassign unselected cells (i.e. incomplete tasks) to other users to ensure that the campaign's task completion goal can be met (Goal 2). For this reason, the VTAS incorporates a management system that allows

incomplete tasks to be reassigned to new users. Without this reassignment process, incomplete tasks would stall progress toward achieving the campaign’s tasks-completion goals. To address this challenge, the task associated with each generated cell recommendation can have different statuses that reflect its current state in the process of being completed by the user. The states are: *‘notified’* if the task has been recommended to a user, *‘accepted’* if the user agrees to perform a task in this cell, *‘realised’* if the user completes it, and *‘not realised’* if the user does not complete it within the timeout period while the task remains in the accepted status but is not performed. A task automatically moves to *‘not realised’* status if it remains in *‘notified’* or *‘accepted’* status for a concrete period of time. In this thesis, this timeout was set to seven minutes but can be adjusted according to the campaign’s requirements. This duration should be determined during the campaign design, considering factors such as the participation population (e.g., elderly people, young people, etc.) and the size of the cells. This automatic status change allows the task to be reassigned to other users dynamically, ensuring that no task remains uncompleted for an extended period.

Figure 4.1 illustrates how the VTAS interacts with an individual and outlines the general workflow of the interaction. The process begins when a user expresses interest in participating in a campaign. For this to happen, an active campaign must be in progress. If no campaign is currently active, a new one must first be configured by defining the hypothesis, geographical area, temporal constraints, and task-completion objectives, as detailed in Chapter 3. Once a campaign is active and a user requests a task, the VTAS generates N cell-recommendations. The user selects one of these recommendations and proceeds to complete a task within the chosen cell. Upon acceptance of a cell recommendation, a pending task within that cell is assigned to the user. No additional cell-recommendations are generated for the user until the status of the assigned task changes either due to its completion, in which case the task status is updated to *realised*, or after a specified time without completion, in which case the task status is updated to *non-realised*. If the user has no accepted tasks and requests a new task, the system generates

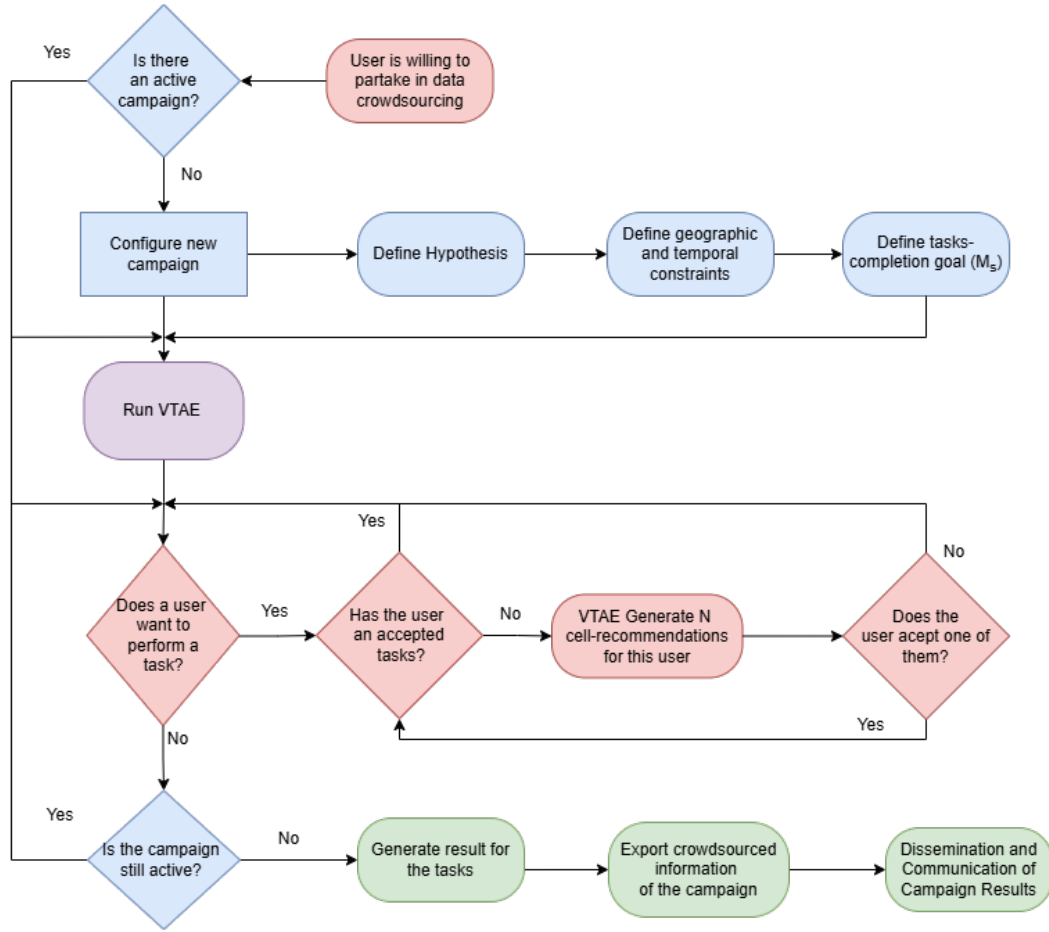


Figure 4.1: This figure illustrates the interactive workflow of the VTAS with a single user during a campaign. The process encompasses key stages, including campaign configuration, cell-recommendation, and the dissemination of results.

a new set of N cell-recommendations tailored to the user. The status associated with each task enables the system to track the interaction's progress and adjust the VTAS's responses to align with the campaign's objectives. At the campaign's conclusion, all completed tasks' results are consolidated, analysed, and interpreted. Notice that the VTAS is envisaged to streamline the effective execution of data crowdsourcing processes which are essential in the organization and execution of Citizen Science campaigns. Indeed, thanks to VTAS's operation, promoters of a Citizen Science campaign will be able to effectively disseminate and communicate the findings to maximise their value and achieve the intended social benefits. While Figure 4.1 illustrates the in-

Variable	Definition
N	Number of cell-recommendations generated for each request.
m	Number of cells of a campaign.
n	Number of slots of a campaign.
L	Set of Cells of a campaign.
j	A cell of a campaign.
S	Set of slots of a campaign.
S_i	The i -th slot of a campaign.
d_s	Duration of the slot (except the last).
M_s	Maximum number of completed tasks allowed for each cell within each slot.
$St(S_i)$	Starting time of the slot S_i .
$Ft(S_i)$	Ending time of the slot S_i .
$t_{j,i}^k$	The k -th task of the cell (L_j) at slot S_i .
C_d	The distance between the centres of two adjacent cells.
P_j^t	Priority of the cell (L_j) at time t .
$accep_{j,i}^t$	Number of tasks accepted by cell j in the slot S_i until time t .
$rea_{j,i}^t$	Number of realised task at cell j in the slot S_i until time t .

Table 4.1: Glossary of mathematical variables used in the VTAS description. These variables are also explained in the Glossary called **Mathematical Notation**.

teraction of a single user with the VTAS solution, the system is designed to concurrently assign tasks to multiple users simultaneously.

A priority metric was created to achieve an equitable distribution of realised tasks among the cells (Goal 3). It describes the *cell's priority*, indicating how big is the need of a cell for having another realised task. This metric considers several campaign features, such as the remaining time until the slot expires and the number of completed and accepted tasks in the cell. This metric quantifies a cell's urgency for additional tasks by comparing its current progress of accepted and realize task in the current slot with an ideal uniform distribution of the M_s desired tasks throughout the slot. To illustrate this concept, consider a scenario where $M_s = 3$, then a uniform distribution would imply that the first task is completed at the beginning of the slot, the second at the midpoint, and the final one at the end. To perform this comparison, the metric consists of two factors: (1) A time-based factor, which measures

the proportion of the slot that has elapsed. (2) A workload-based factor, which represents the percentage of completed and accepted tasks relative to the expected M_s tasks. By subtracting these two values-which ideally range between 0 and 1 if the algorithm adheres to workload limitations-an indicator of the cell's priority is obtained. A positive value of this metric indicates that a cell is lagging behind its expected task completion rate and requires additional tasks to maintain fairness. Conversely, a negative value suggests that the cell has exceeded its expected progress. The mathematical notation of this metric is based on the following variables: Let t be the current time, S_i the slot corresponding to the time t , d_s the span of the slots, $rea_{j,i}^t$ the number of realised tasks of the cell j in the slot S_i at time t (realised tasks from the start of the slot to the current slot time), $accep_{j,i}^t$ the accepted tasks of the cell j in the slot S_i at time t , and M_s the maximum number of completed tasks allowed per cell in each slot. Then, the priority of the cell j at time t can be denoted as P_j^t and is defined in Eq. 4.1. The mathematical notation utilised is defined in Table 4.1.

$$P_j^t = \frac{(t - St(S_i))}{d_s} - \frac{accep_{j,i}^t + rea_{j,i}^t}{M_s} \quad \forall j \in L; \forall t | St(S_0) \leq t \leq Ft(S_{n-1}) \quad (4.1)$$

By comparing this metric across various cells, it can be determined which cell needs a task more than another to maintain an equitable distribution. This metric facilitates efficient task management by dynamically adjusting cell priorities based on real-time demand. An automated scheduled process assesses the priority of cells for a campaign every two minutes. This means that the VTAS always has access to up-to-date priority values for each cell, removing the need to recalculate them every time a cell-recommendation is requested.

Furthermore, as previously explained, the VTAS is a user-centric solution, meaning that it works as a recommendation algorithm in its interaction with the user. Based on the approach outlined regarding the various implemented factors, the only remaining aspect to explain is how the cell-recommendations are generated. The selection of cell-recommendations for a user is determined

by the priority of the cells, the user's location, and the number of accepted and completed tasks of each cell in the current slot. To adapt the recommendations to the users, only cells close to the user are considered for recommendation. In this way, the user is not required to travel long distances to complete a task. The definition of proximity for cells is a campaign parameter that can be customized for each campaign, based on its specific requirements and the participating population. In this thesis, a cell is considered to be in proximity to the user if, and only if, the distance between the cell's center and the user's location is less than five times the C_d distance. As mentioned, this is a predefined criterion, but it can be adjusted according to the specific needs of the campaign. For example, if the C_d is large, more time will be required to change the task's state from '*accepted*' or '*notified*' to '*not realised*'. Therefore, it is crucial to define these variables based on the campaign's objectives and the characteristics of the participating population. Hence, to create the cell-recommendations for a specific user, the initial step is to select the cells closer to the user. In this list, all cells that already have M_s tasks counting the realised plus accepted in the current slot are removed. In this way, the VTAS respects the workload limitation of the Spatial Crowdsourcing model (Goal 1). To create the cell-recommendations, the priority, the number of accepted and completed tasks of each cell, and the distance from the user to each cell on the resulting pruned list are calculated. The list of cells is then ordered by fewer accepted plus completed tasks, higher priority, and closer proximity. The top N cells of this ordered list are the cell-recommendations. If fewer than N cells are present in the ordered list of possible cell-recommendations for a user, the number of available cells in the list will determine the number of cell-recommendations.

In the context of this thesis, the variable N is fixed at 3 to limit the cognitive overload for the user (Bollen et al., 2010; Beierle et al., 2017).

Figure 4.2 shows a cell-recommendation generated for a user using the VTAS algorithm. In this campaign, the M_s parameter is set to 8, and the completion colour scale, described in Chapter 3, is created based on this value. Each cell in the image is labelled with a tuple (x, y), where the first number represents the tasks completed in the cell in the current slot, and the second

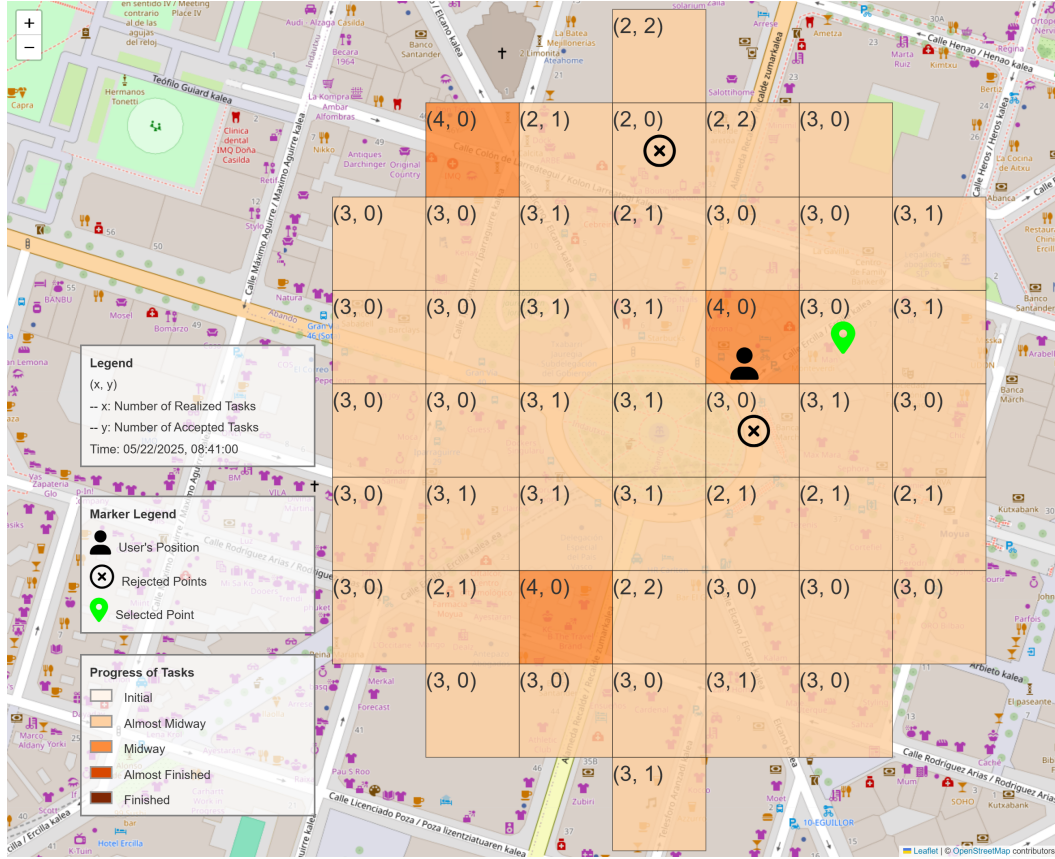


Figure 4.2: An example of a cell-recommendation using the VTAS strategy. In this campaign, the parameter M_s is set to 8, and the completion colour scale is created based on this value. The user, represented by the person icon, is recommended three nearby cells, indicated by the green and cross icons. The green icon is the accepted cell-recommendation.

represents the tasks accepted in the cell in the current slot. The user, represented by the person icon, is recommended three nearby cells (based on the proximity definition used in the algorithm), indicated by the green and cross icons. The green icon is the accepted cell-recommendation.

This collaborative task allocation strategy, VTAS, has been connected to the RESTful API, called MVE, and is available on GitHub¹. This API, referred to as the Micro Volunteering Engine (MVE), supports the creation of campaigns based on the Spatial Crowdsourcing model described in Chapter 3.

¹<https://github.com/mpuerta004/RecommenderSystem>

The MVE can be enhanced through various spatial crowdsourcing strategies. In this chapter, the VTAS is introduced and developed as an initial step toward advancing beyond traditional spatial crowdsourcing approaches in the Citizen Science domain. The MVE handles task allocation by efficiently matching available users with tasks through the VTAS, irrespective of the task type or the specific Citizen Science initiative. This approach ensures broad adaptability and inclusivity in task assignments.

4.2 Baseline algorithms, Evaluation metrics and Use Cases details

This section focuses on three key aspects: (1) the baseline algorithms used as benchmarks to assess the efficiency of VTAS, (2) the evaluation metrics employed to measure the effectiveness of VTAS, and (3) the selected use cases for evaluating VTAS’s performance. Through these use cases, the performance of VTAS will be analysed using the defined metrics and compared with the results of the baseline algorithms.

4.2.1 Baseline Algorithms

The VTAS algorithm was evaluated using multiple metrics and compared against two baseline algorithms. The first baseline, referred to as the Nearest Neighbours (NN) algorithm, is an adaptation of the 1-Nearest Neighbours algorithm tailored to this specific context (Zhang et al., 2023c; Halder et al., 2024). It assigns users to the nearest cell to perform a task. The second baseline, referred to as the Limited Nearest Neighbours (LNN) algorithm, is created based on the NN algorithm and incorporates the same workload limitation as the VTAS. This workload limitation prevents more tasks from being completed than those defined by the task-completion goal, i.e., no more than M_s tasks per cell in each slot. For both algorithms, if two cells are equidistant, the cell with the higher priority (as determined by Eq. 4.1) is recommended. Furthermore, both algorithms recommend only one cell to the user for each request. This limitation stems from the fact that these

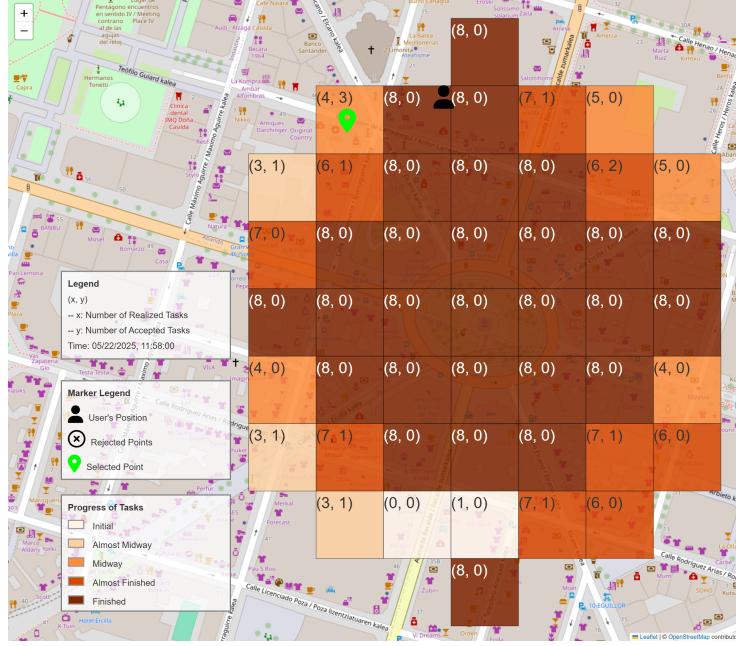


Figure 4.4: An example of a cell recommendation for the LNN algorithm: As shown in this picture, the user (represented by the person icon) is recommended the nearest cell that has fewer than M_s completed tasks. The recommended cell is depicted by the green icon in the picture. In the case of this campaign, the M_s value is set to 8.

tasks. Additionally, the figure highlights that the cell recommendation is based solely on proximity to the user rather than prioritizing the cell with the greatest need for tasks in the user's surroundings.

Unlike the NN algorithm, the LNN algorithm (Figure 4.4) respects the workload limitation defined by the Spatial Crowdsourcing model. As a result, all cells in this figure have task counts less than or equal to M_s , ensuring that no cell exceeds M_s completed tasks. As shown in the picture, among the cell recommendations provided to the user (represented by the green icon), can not be included cell with more than M_s tasks, including both accepted and realised tasks as the algorithm respects the workload limitation. However, similar to the behaviours of the NN algorithm, this figure shows that the cell-recommendation is still based on proximity to the user rather than prioritizing the cell with the greatest need in the user's surroundings.

4.2.2 Evaluation Metrics

To evaluate the achievement of the goals of this research, specific metrics were established. The performance metrics used to evaluate the VTAS are given below:

Questionnaire to evaluate the acceptance and adoption of users to the Spatial Crowdsourcing technologies: To assess the adoption and acceptance of users (Goal 1), a questionnaire inspired by the Technology Acceptance Model (TAM) (Davis et al., 1989) and System Usability Scale (SUS) (Grier et al., 2013) was created. The Technology Acceptance Model (TAM) is a theoretical framework used to evaluate users' acceptance of a technology based on two key factors: *perceived usefulness* (how useful users believe the technology will be in improving their performance) and *perceived ease of use* (how easy they find it to use) (Davis et al., 1989). These factors influence the attitude toward the technology, its intention to use it, and its actual adoption. TAM is widely applied in research to predict and understand user behaviour regarding new tools or technological systems (Marangunić and Granić, 2015). The System Usability Scale (SUS) is a standardized questionnaire designed to measure the *perceived usability* of a system, product, or service, such as an application, website, or software (Grier et al., 2013). It was developed by John Brooke in 1986 and has become one of the most widely used methods for assessing user experience (Vlachogianni and Tselios, 2022). The created questionnaire can be found in Appendix B.

Task Completion Ratio (TCR): This metric measures the ratio of tasks completed to the total number of tasks specified by the task-completion goal of the campaign (Goal 2). It must be calculated at the end of each campaign slot. A higher TCR indicates a greater percentage of tasks completed. Values exceeding 100% suggest that the algorithm has surpassed the workload limitation. In these cases, the campaign has more tasks accomplished than the M_s tasks per cell at the end of the slot, exceeding the campaign's task-completion goal, thus indicating that the workload limitation has been exceeded. Specific algorithms, such as NN, are not constrained by this limitation and can surpass this threshold, leading to a TCR exceeding 100%. The following equation

defines the TCR value at the end of the S_i slot of the campaign:

$$TCR^{S_i} = \frac{\sum_{j \in L} \sum_{S_i \in S} rea_{j,i}^{Ft(S_i)}}{m * M_s} \quad \forall S_i \in S \quad (4.2)$$

Where S_i denotes the slot of the campaign, $Ft(S_i)$ is the time at which slot S_i ends, m is the number of cells in the campaign, M_s is the maximum number of tasks allowed to be completed for each cell in each slot, L is the set of cells in the campaign, $rea_{j,i}^{Ft(S_i)}$ is the number of tasks performed in cell j in the slot S_i at time $Ft(S_i)$ and S the set of slots of the campaign.

Realised Tasks per minute (RT): This metric measures the number of tasks completed within a single minute. In other words, a count will be made of the tasks completed between, for example, 16:15 and 16:16, and this data will be analysed as part of the time series consisting of the values of realised tasks for each minute throughout the duration of the campaign. It provides insights into the efficiency of task allocation algorithms by indicating which algorithm leads to more tasks being completed per minute during the campaign. A higher RT value reflects greater task completion efficiency (Goal 2). The equation for this metric is as follows:

$$\text{Realised Tasks at time } t = RT^t = \sum_{j \in L} (rea_{j,i}^t - rea_{j,i}^{t-1}) \quad (4.3)$$

where $rea_{j,i}^t$ is the number of tasks performed in cell j at time t , L the set of cell of the campaign.

This metric is calculated every minute during the campaign, resulting in a time series that enables the visualization of this metric over time. As a time series, for each slot, the mean and standard deviation, referred to as $MEAN \pm SD$, are calculated. The time series visualization and the $MEAN \pm SD$ of this metric for each slot will be used to evaluate the performance of the VTAS to assess the Goal 2.

Load Fairness Index (LFI): Measure the equitable distribution of the tasks over the cells at a concrete moment. That is the variance of the set of data composed by the number of complete tasks of each cell in a concrete slot at the time t of the campaign. Low LFI denotes an equitable distribution of

tasks across cells, whereas high LFI signifies a non-equitable distribution. So, based on the Goal 3, the VTAS has to achieve a low LFI at every moment of the campaign. This variable can be calculated at any time t of the campaign with the following equation:

$$LFI^t = Var(\{rea_{j,i}^t\}_{j \in L}) = \frac{\sum_{j \in L} (rea_{j,i}^t - MEAN(rea^t))^2}{m - 1} \quad (4.4)$$

Where t represents the current time, L is the set of cells in the campaign, $rea_{j,i}^t$ is the number of tasks completed in cell j in the slot S_i at time t , and $MEAN(rea^t)$ is the average of the set $\{rea_{j,i}^t\}_{j \in L}$, i.e., the average number of tasks completed per cell in the slot S_i until time t . Additionally, m denotes the number of cells in the campaign. This metric is calculated every minute during the campaign, generating a time series that visualizes the distribution of tasks over time. For each slot in the series, the mean and standard deviation (denoted as $MEAN \pm SD$) are computed based on the values recorded in that specific slot. The time series visualization and the $MEAN \pm SD$ will be used to evaluate the performance of the VTAS to assess the Goal 3.

Furthermore, this metric can be visually observed using the completion colour scale defined in the previous chapter (Figure 3.1). The bigger number of colours used in the campaign visualization corresponds to a higher value of this metric, as it signifies a more diverse range of tasks performed in each cell. This representation will be employed in the experiments to illustrate this metric and to compare the task distribution achieved by each algorithm at the conclusion of each slot.

The LFI and RT metrics are represented as time series, calculated every minute throughout the campaign to generate a time serie that is subsequently visualised for assessing the performance of VTAS. Computing these metrics every second would demand excessive computational resources and, in the case of RT, would yield limited meaningful insights. Therefore, the decision to calculate and obtain this time series at one-minute intervals was made to ensure a sufficient volume of data while maintaining an optimal balance between computational efficiency and the accuracy required for a representative analysis of VTAS behaviour.

4.2.3 Use Cases

The assessment of the VTAS solution relies on the following two uses cases:

- **Simulation Use Case:** This simulation of the task allocation strategy aims to assess the behaviour and performance of the VTAS under various conditions. It primarily provides data on how effectively the algorithm distributes tasks (Goal 3) and highlights the differences between the VTAS and the baseline algorithms considered. Additionally, the simulation allows for the observation of which behavioural variables of the simulated users have a significant impact on achieving the task-completion goal of the campaign (Goal 2).
- **Deusto Use Case:** This involves a targeted, small-scale trial (12 participants) conducted at the University of Deusto (Bilbao' campus). This use case is not truly a Citizen Science initiative, as it is not an initiative that has arisen from the social need of users. However, it simulates one in which the aim was to create a photo collage that visually represented the users' perception of the university and its surroundings. This use case provided an opportunity to assess the effectiveness of the VTAS algorithm in fostering altruistic participation within the context of a real-world experiment, as well as to evaluate user adoption and acceptance (Goal 1).

In the subsequent sections, the use cases assessment is articulated in detail. The evaluation of the VTAS was conducted using the previously established metrics and it included a comparative analysis with the selected baseline algorithms and the result obtained with the VTAS.

4.3 Simulation Use Case

In this section, we present the simulation use case. The simulation models the task allocation problem based on the Spatial Crowdsourcing model defined in Chapter 3. It has the goal of assessing the behaviour and performance of the VTAS under various conditions. The simulation allows for the observation of

which behavioural variables of the simulated users have a significant impact on achieving the task-completion goal of the campaign (Goal 2). Additionally, it primarily provides data on how effectively the VTAS distributes tasks (Goal 3). Furthermore, in cases where the task-completion goal cannot be fully achieved, it is essential and necessary for the algorithm to generate an equitable distribution of tasks, thereby preserving the usability and validity of the collected data. This matter is also assessed in the simulation.

This simulation creates fictional users who, while following their routines (if applicable), participate in the defined campaign. These simulated users follow the same interaction workflow (Figure 4.1) that a real user would follow. This simulation is designed to replicate real-world behaviours and includes several customisable variables to simulate it. These variables make it possible to simulate different levels of user commitment and availability, as well as inconsistencies or lack of dedication in executing assigned tasks. The customisable variables in the simulation are as follows:

- **Number of Users:** This parameter defines the number of simulated participants who will take part in the campaign. Once the tasks-completion goal are established, we can evaluate the individual effort expected from each user, specifically the average number of tasks each user is expected to perform. While this metric is theoretical and does not reflect actual participation-since some users may engage minimally while others may participate extensively-it can provide an estimate of the average effort that users should put forth during the campaign. For example, if the variable M_s is 8, the number of cells (denoted as m) is 10, and there are 5 users, the average individual effort per slot would be 16 tasks per slot. That is, in this scenario, it is expected that each user will complete 16 tasks per slot. However, if the number of users increases to 10, the individual effort per slot decreases to 8 tasks per slot. Thus, if the expected individual effort is high, the likelihood of failing to meet tasks-completion objectives increases.
- **User Task Completion Probability:** This probability indicates how likely it is for a user to complete an accepted task. A high User Task

Completion Probability signifies strong user commitment to accepted tasks, whereas a low probability suggests a lack of commitment. In cases where the probability is low, the tasks-completion goal may be at risk. Therefore, the algorithm needs to effectively distribute tasks to ensure the creation of a reliable dataset.

- **User Availability Probability:** This parameter indicates the likelihood of a user requesting a cell-recommendation at any given time. However, it does not account for the probability of the user accepting any of these recommendations. Users are deemed unavailable if they have already accepted tasks. A high User Availability Probability indicates a strong motivation to participate i.e. high level of pro-activeness. When users are more motivated and available, they are likely to engage more, leading to a greater completion of tasks compared to situations where this probability is low.

The experiments conducted in this use case tested different values of these customisable variables to evaluate the performance of the VTAS under various conditions. Before presenting the results of these experiments, an explanation of how the simulation operates is provided.

Based on the values of customisable variables, users participate in the campaign and complete tasks along their simulated trajectories. At the start of the simulation, each user is randomly assigned an initial trajectory representing their daily routine, along with a probability that defines how strictly they will adhere to this routine. As users move along their trajectories, their positions gradually progress toward the final destination (i.e., the endpoint of the trajectory). Every minute, the User Availability Probability determines whether a user is willing to participate. If the probability confirms the user's availability and the user has no accepted tasks, then they submit a request. The algorithm responsible for task allocation (VTAS, LNN, or NN) processes this request by generating corresponding cell-recommendations. The simulated user then selects the cell with the smallest deviation from their current trajectory, though they may also choose not to accept any recommendation. If a recommendation is accepted, the user is assigned a random time to reach the

selected cell. This time had to be shorter than the period required for tasks to transition from *accepted* or *notified* to *non-realised*. Once this time elapses, the user is assumed to have arrived at the cell, where they must complete the assigned task. The User Task Completion Probability determines whether the user successfully completes the task, enabling the simulation of different levels of user commitment. Regardless of task completion, the simulation continues to model the user’s journey along their trajectory until they reach the end-point. Throughout their trajectories, the process of requesting and completing tasks is simulated continuously. Upon reaching the end of their trajectories, the Routine Repetition Probability determines whether they will repeat the same path or be assigned a new, randomly generated trajectory. If a new trajectory is assigned, the routine is simulated again, with users participating in and completing tasks based on the predefined probabilities.

Figure 4.5 illustrates an example of three cell recommendations generated by the VTAS algorithm in a simulation, depicting the user’s trajectory. The trajectory is represented by the user’s current position (indicated by a black person icon) and their final destination (indicated by a blue person icon). In the figure, the three cell recommendations computed by the VTAS algorithm are displayed. The cell accepted by the user is marked with a green icon, while the other two recommendations are indicated with cross icons. The completion colour scale of the cells corresponds to the M_s parameter, which is set to a value of 8 in this simulation.

The simulation was implemented in Python, and the code is publicly available in the GitHub repository¹ where the microservice offering a RESTful API, named MVE, has been published. The simulation supports three task allocation methods: VTAS, LNN, and NN. Each of these strategies can be executed via a dedicated POST request, which runs the simulation for a predefined campaign. Additionally, the simulation automatically generates visual representations of the campaign’s state, using the completion colour scale defined by the variable M_s .

Using this simulation, several experiments were conducted using different values for the customised variables to analyse the performance of the VTAS.

¹<https://github.com/mpuerta004/RecommenderSystem>

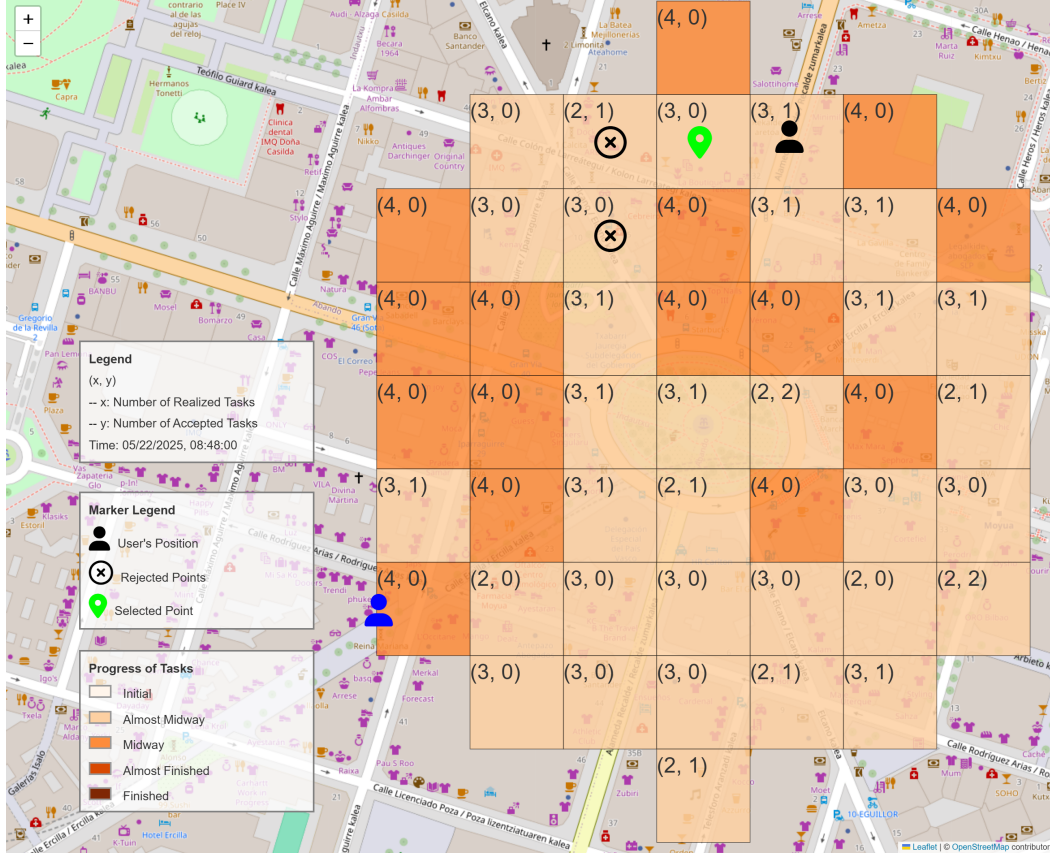


Figure 4.5: An example of a cell recommendation using the VTAS algorithm in a simulation. The user's position is depicted by the black person icon, and three nearby recommended cells are shown, marked with green and cross icons, representing the cells with the highest task priority. Among these, the cell with the green icon is the one selected by the simulated user as the recommended cell. This cell is where the user completes a task. The blue person icon signifies the user's final destination.

These experiments maintained the campaign configuration while varying the customised variables to assess their impact and behaviour. The **campaign configuration** for all the experiments included two time slots, each lasting 2 hours. The campaign was located in the centre of the city of Bilbao, Basque Country, SPAIN at coordinates longitude of -2.9349889768009008 and latitude of 43.26302147532318, with a radius of 0.2 km. The campaign has 47 cells, so the variable m is 47. The campaign started at 08:00 AM and ended on the same day at 12:00 AM. The C_d was 0.05km, and M_s was 8 tasks. This variable

defines the **task-completion goal** of the campaign, which consists of 8 tasks for each of the 47 cells per slot, totalling 375 tasks per slot. Additionally, this variable also sets the **workload limitation**, which is violated if more than 8 tasks are completed in the same cell in a given slot. A visualisation of this campaign is shown in Figure 4.6.

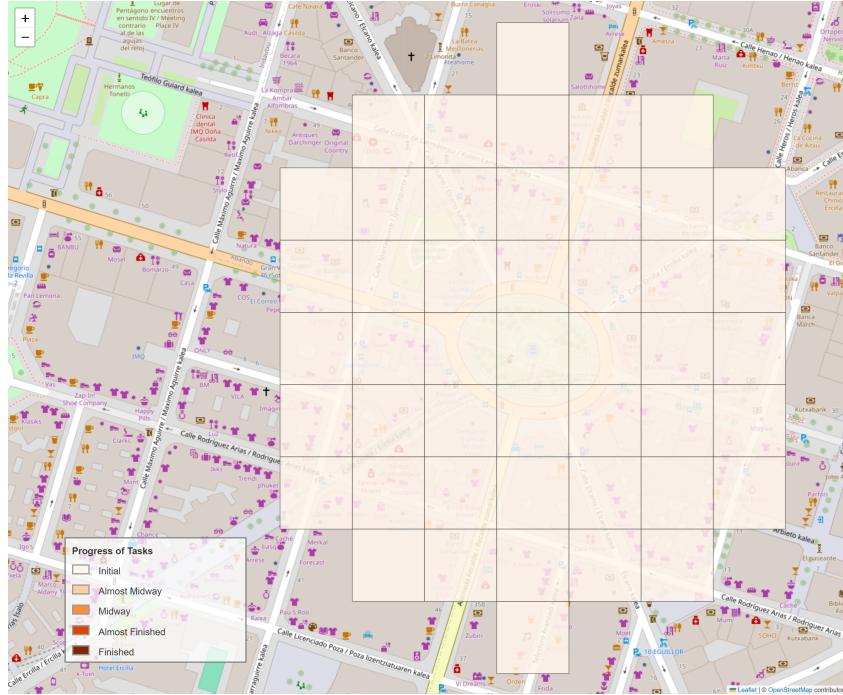


Figure 4.6: A visualization of the campaign used to evaluate the VTAS in the simulation use case. This figure was obtained during the initial phase of the campaign, before any tasks were carried out.

The remainder of this section is organised as follows: Subsection 4.3.1 presents the first experiment, where the VTAS is evaluated with different numbers of users to assess its effectiveness. Subsection 4.3.2 presents the second experiment, in which different User Availability Probabilities are tested to evaluate the VTAS’s performance. Subsection 4.3.3 presents the third experiment, where different User Task Completion Probabilities are tested to further evaluate the VTAS. Finally, subsection 4.3.4 presents the conclusions of the simulation use case.

4.3.1 Experiment 1: Crowd Size Variation

This experiment evaluates the effectiveness of the VTAS across different crowd sizes. As the number of users decreases while maintaining the same task-completion goal, individual effort increases due to fewer users sharing the workload. Consequently, with fewer users, the likelihood of successfully achieving the task-completion goal decreases. This section examines how the VTAS supports task completion under varying levels of individual effort (Goal 2), including cases where achieving the goal may not be feasible. In such scenarios, ensuring a fair distribution of tasks is essential. Therefore, we also assess the VTAS’s effectiveness in promoting equitable task distribution (Goal 3).

In this subsection, we evaluate the VTAS’s performance across different crowd sizes: 100 users (low individual effort), 75 users, 50 users, and 25 users (high individual effort). The campaign configuration and task-completion goal remained consistent with the parameters defined in the previous section. For these campaigns, the User Task Completion Probability was set at 75%, reflecting a high level of commitment, while the User Availability Probability was fixed at 50%. The VTAS’s performance was compared against the LNN and NN baselines. Each scenario was simulated five times using each task allocation strategy, and the results presented here represent the average values from these simulations.

		TCR			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	100%	100%	94.52%	50.48%
	LNN	100%	98.56%	74.95%	40.59%
	NN	160.96%	121.22%	83.40%	39.89%
Slot 2	VTAS	100%	100%	95.43%	49.15%
	LNN	100%	98.67%	73.46%	39.521%
	NN	157.13%	122.02%	82.29%	41.12%

Table 4.2: TCR evaluation over different user crowds (100, 75, 50, and 25) using the VTAS, LNN, and NN algorithms. The values in bold indicate the high TCR in each slot for each user count.

The Task Completion Ratio (TCR) for the four cases tested is shown in

Table 4.2. The table shows that as the number of users increases, the TCR improves across all algorithms. The first observation is that algorithms with a workload limitation (VTAS and LNN) never exceed 100% in the TCR metric, which is expected since the workload limitation prevents surpassing the task-completion goal. In contrast, the NN algorithm exceeds this limit when 100 and 75 users are involved, as it has no such restriction, thereby exceed the VTAS and LNN in terms of TCR. However, in scenarios with 25 and 50 users, where the NN does not exceed the task-completion goal, the VTAS consistently achieves a higher TCR. It suggests that without the workload limitation, the VTAS would outperform the NN in TCR. Additionally, the TCR of the LNN algorithm is always equal to or lower than that of the VTAS.

The right-hand side of Figure 4.7 illustrates the evolution of the realised tasks (Eq. 4.3) per minute for the first slot of the tested campaign using the VTAS, LNN, and NN algorithms. The moving average¹ of 20 values was used to illustrate the evolution of the realised tasks. Since the campaign consists of two slots and the task-completion objectives were reset at the beginning of each slot, only the first slot is presented. The graph shows that the VTAS consistently achieves a higher number of realised tasks than the LNN and NN algorithms. The superiority in the number of tasks completed is consistently maintained, except at the end of the slot in scenarios with 100 and 75 users, where the VTAS graph shows a sharp decline. This drop occurs because the task-completion goal has been fully met or near to, prompting users to stop receiving cell-recommendations as a deliberate mechanism to prevent overload. As a result, the number of realised tasks decreases due to the enforcement of the workload limitation. This confirms that the TCR achieved by the LNN and VTAS is not higher in the case of 100 and 75 users due to the workload limitation. Furthermore, since the realised tasks graph for the VTAS remains superior at all times until the workload limitation

¹A Moving Average (MA) is a simple way to analyse data trends over time by smoothing out short-term fluctuations. It works by calculating the average of a set of values over a specific period and then continuously updating this average as new data points are added. For example, in a 7-day moving average, you take the average of the last 7 days of data. The next day, you drop the oldest value and add the newest one, recalculating the average. This helps reveal overall trends by reducing the impact of daily ups and downs.

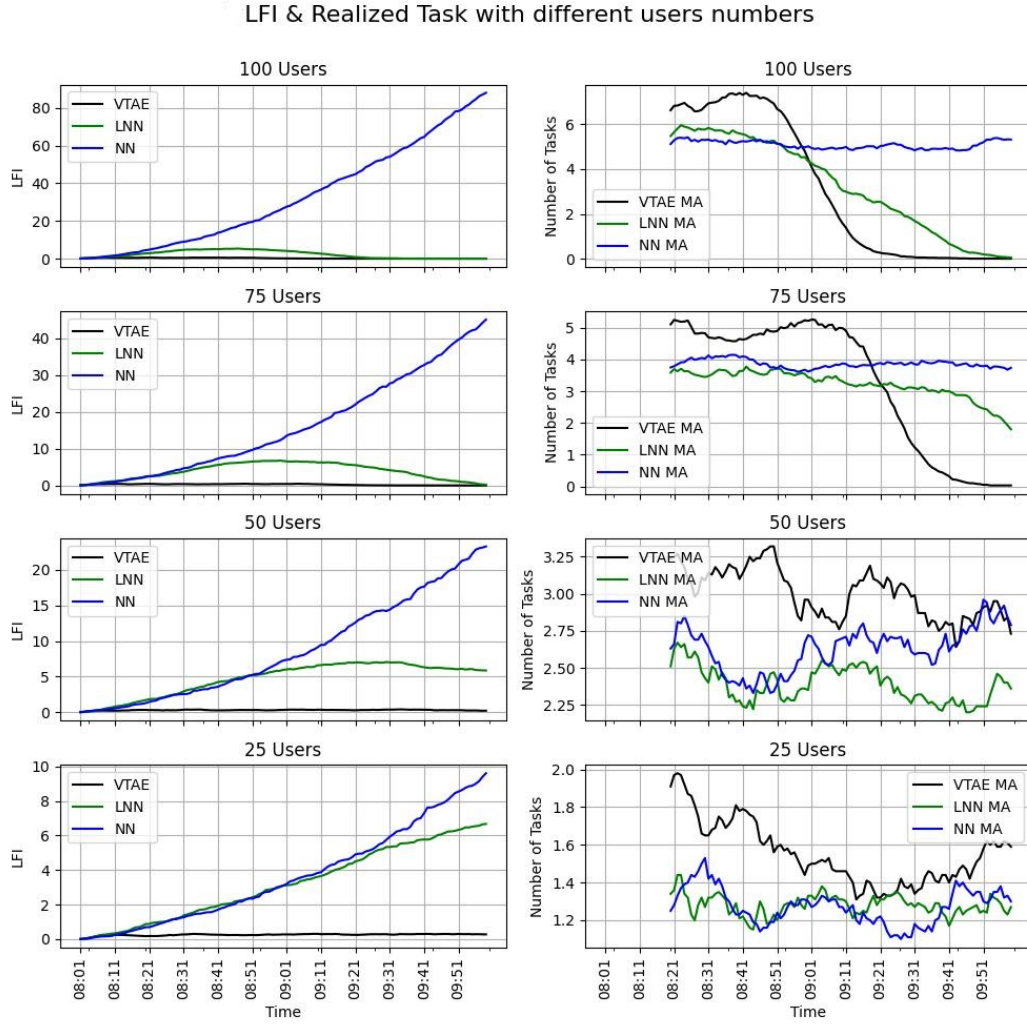


Figure 4.7: The LFI (Eq. 4.4 left-hand side pictures) and realised task (Eq. 4.3 right hand-side figures) over the first slot across different users count (100, 75, 50, and 25 Users). The task allocation algorithms used are the VTAS, LNN, and NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results presented are the mean of five executions. Since the campaign consists of two slots and the tasks-completion objectives were reset at the beginning of each slot, only the first slot is presented.

intervenes, it can be concluded that, without this restriction, the VTAS would outperform the NN both in the number of realised tasks and in the TCR in scenarios with 100 and 75 users. However, this limitation is crucial, as we are in a Citizen Science domain where overloading users could lead to a decline

in their motivation and participation. Therefore, in cases where the workload limitation applies, it is better not to exceed this limit even when we would desire to achieve better results of the TCR or RT metric, as this helps avoid placing excessive demands on users. Consequently, although the NN achieved better values in some cases, the performance of the VTAS is more appropriate in these scenarios.

Table 4.3 show the MEAN $\pm$ SD of the realised tasks metric (RT). The highlighted results correspond to the algorithms with the highest mean in each slot for each user count. These values support what was previously mentioned. The VTAS outperforms the NN algorithm only in the case of 25 and 50 users, that is when the workload limitation is not reached. In the other two tested cases, the NN outperforms the VTAS in this metric.

		MEAN $\pm$ SD (RT)			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	3.16 $\pm$ 3.38	3.16 $\pm$ 2.42	2.99$\pm$0.81	1.59$\pm$0.00
	LNN	3.16 $\pm$ 2.40	3.11 $\pm$ 1.08	2.37 $\pm$ 0.70	1.28 $\pm$ 0.56
	NN	5.06$\pm$0.99	3.83$\pm$0.77	2.64 $\pm$ 0.77	1.26 $\pm$ 0.48
Slot 2	VTAS	3.13 $\pm$ 3.27	3.13 $\pm$ 2.33	2.99$\pm$0.89	1.54$\pm$0.58
	LNN	3.13 $\pm$ 2.44	3.09 $\pm$ 1.11	2.30 $\pm$ 0.73	1.24 $\pm$ 0.47
	NN	4.92$\pm$1.20	3.82$\pm$0.96	2.58 $\pm$ 0.73	1.29 $\pm$ 0.57

Table 4.3: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count.

The left-hand side of Figure 4.14 displays the LFI (The Load Fairness Index, Eq. 4.4) during the first slot of the campaign for the VTAS, LNN, and NN algorithms. It is evident that the LFI graph for the NN algorithm demonstrates exponential behaviour across all tested scenarios. This is expected, as the NN algorithm does not impose any workload limitation. A similar exponential pattern is observed with the LNN algorithm in the case of 25 users. These patterns indicate that, in these scenarios, both algorithms fail to maintain an equitable distribution of tasks across cells. In contrast,

the VTAS algorithm consistently maintains a minimal LFI in all tested scenarios, showcasing its ability to ensure a more equitable distribution of tasks across cells. VTAS ensures an equitable distribution at all times, resulting in consistently low LFI metrics across all scenarios and throughout the entire campaign. This highlights that VTAS not only meets Goal 3 throughout the entire campaign but also achieves it with greater efficiency compared to the selected baselines.

A peculiar behaviour is observed in the results obtained with the LNN for the cases of 75, 50, and 100 users. The LNN's LFI increases until it reaches a maximum, after which it starts to decrease (in the cases of 75 or 100 users) or remains stable (as in the case of 50 users). This behaviour occurs because, when some cells reach M_s tasks, the workload limitation prevents more tasks from being assigned to those cells. As a result, tasks are more equitably distributed among the remaining cells, leading all cells to eventually reach a workload similar to M_s . This, in turn, causes the LFI metric to begin decreasing. Hence, the reduction in LFI observed in the case of LNN is attributable to the workload limitation. This conclusion is also supported by the fact that the only difference between NN and LNN is precisely this workload limitation that the LNN algorithm has and the NN does not.

The MEAN $\pm$ SD of the LFI metric is presented in Table 4.4. The highlighted results represent the algorithms with the lowest mean for each user count in each slot. These values reinforce the earlier findings: due to the exponential behaviour of the NN algorithm, its mean LFI values are considerably high. In contrast, the VTAS algorithm consistently maintains low LFI values across all cases, indicating an equitable task distribution.

As explained above, the completion colour scale defined in Chapter 3 also provides a graphical representation of the LFI metric. This completion colour scale represents the range of completed tasks, based on the M_s variable of the campaign. Figures 4.8, 4.9 and 4.10 illustrate the distribution of completed tasks at the end of the first slot, using the VTAS, NN, and LNN algorithms, respectively, with 25 users. This scenario was chosen because the higher individual effort required makes achieving the task completion goal more challenging, highlighting the importance of equitable task distribution. In the

		MEAN $\pm$ SD (LFI)			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	0.22$\pm$0.22	0.24$\pm$0.17	0.32$\pm$ 0.06	0.26$\pm$0.05
	LNN	2.23 $\pm$ 1.99	3.83 $\pm$ 2.21	4.70 $\pm$ 2.27	3.20 $\pm$ 2.10
	NN	32.55 $\pm$ 26.57	16.23 $\pm$ 13.32	8.75 $\pm$ 7.05	3.64 $\pm$ 2.81
Slot 2	VTAS	0.22$\pm$0.23	0.25$\pm$0.18	0.36$\pm$0.09	0.28$\pm$0.07
	LNN	1.96 $\pm$ 1.87	3.54 $\pm$ 2.22	4.35 $\pm$ 2.22	3.08 $\pm$ 2.04
	NN	28.91 $\pm$ 25.09	15.12 $\pm$ 12.82	8.99 $\pm$ 7.40	3.64 $\pm$ 3.01

Table 4.4: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each user count.

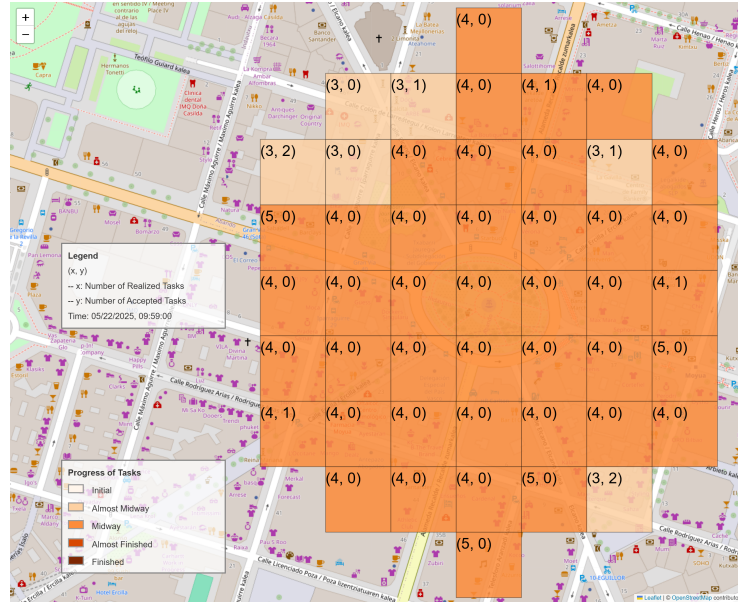


Figure 4.8: Visualization of the completed tasks at the end of the first slot, using the VTAS algorithm with 25 users. The completion colour scale indicates the number of completed tasks, based on the M_s variable of the campaign.

cases of NN and LNN, the representations display four or five distinct colours, indicating an inequitable task distribution, as each colour corresponds to a different range of completed tasks. In contrast, for the VTAS, the cells exhibit more uniform colouring, with only two colour ranges shown, suggesting a more equitable distribution of tasks. This visual representation complements

the LFI results. While this chapter presents the representation for 25 users, the corresponding images for the other scenarios in this experiment can be found in Appendix E, Section E.1.

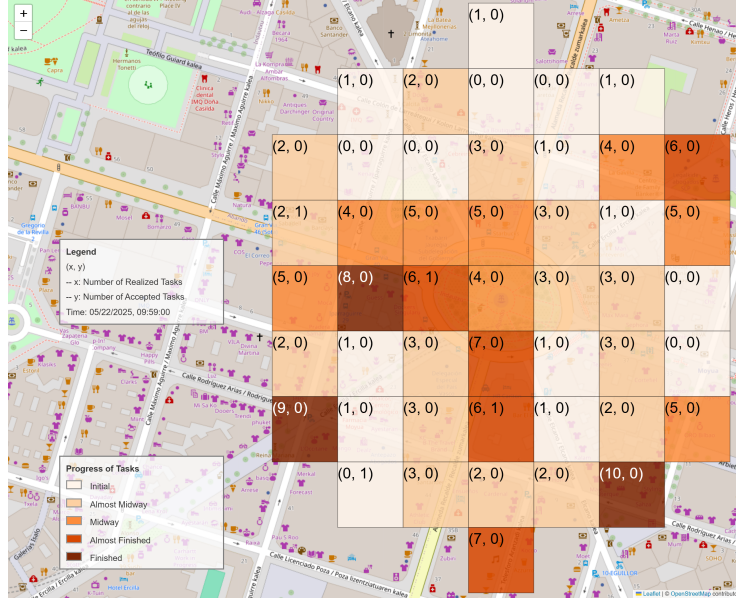


Figure 4.9: Visualization of the completed tasks at the end of the first slot, calculated using the NN algorithm with 25 users. The completion colour scale indicates the number of completed tasks, based on the M_s variable of the campaign.

The results of Experiment 1 highlight the robustness of the VTAS strategy in managing varying crowd sizes. The VTAS consistently achieves a higher TCR compared to the LNN and NN algorithms when workload limitations are not exceeded. Furthermore, the VTAS outperforms the LNN and NN algorithms in the total number of realised tasks, except in scenarios involving 100 and 75 users where the workload limitation is reached. These metrics collectively demonstrate that, if the workload limitation was not achieved, the VTAS would also surpass the NN and LNN in the remaining cases. However, the VTAS strictly adheres to the imposed workload limitation, ensuring it operates within a defined task threshold. Consequently, while it does not exceed the TCR or the total number of tasks performed in all scenarios, its overall performance remains superior to that of the NN. This outcome aligns with

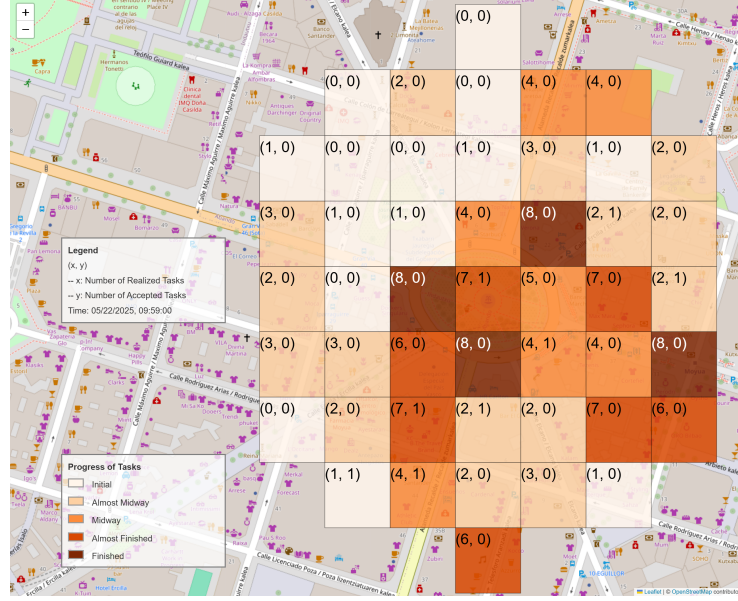


Figure 4.10: Visualization of the completed tasks at the end of the first slot, using the LNN algorithm with 25 users. The completion colour scale indicates the number of completed tasks for each cell and it is created based on the M_s variable of the campaign.

the motivation of this dissertation to enhance and streamline spatio-temporal data crowdsourcing within the Citizen Science domain, where preventing the overburdening of users with excessive tasks is a critical requirement that the strategy must address. In conclusion, the VTAS algorithm proves to be more suitable for this context achieving better results than the baseline algorithms in fulfilling the Goal 2. Furthermore, VTAS maintains a low Load Fairness Index (LFI) across all tested scenarios, indicating an equitable task distribution even under varying levels of user effort, thereby fulfilling the Goal 3. To conclude, the VTAS algorithm is more appropriate for this context, as it yields better results than baseline algorithms in addressing Goal 2 and 3.

4.3.2 Experiment 2: User Availability probability

This experiment evaluates the effectiveness of the VTAS algorithm under varying User Availability Probabilities. Campaigns that have high User Availability Probabilities are characterised by motivated users who make it easier to

achieve task completion goals. In contrast, when User Availability Probability is low, user involvement and motivation decrease, leading to a decreased likelihood of successfully meeting the task completion goal. This section examines how effectively the VTAS supports the achievement of the task-completion goal under different User Availability Probabilities (Goal 2), including scenarios where achieving the goal may not be feasible (i.e., in cases of low motivation). In these situations, ensuring an equitable distribution of tasks becomes essential. Consequently, the ability of the VTAS to maintain an equitable task distribution was also evaluated (Goal 3).

In this subsection, the performance of the VTAS algorithm, in relation to the defined goals, was evaluated under User Availability Probabilities of 75% (high motivation and participation), 50%, and 25% (low motivation and participation). In all cases, the campaign configuration and task-completion goal were defined using the same parameters specified in the previous section. For these campaigns, the User Task Completion Probability was set at 75%, reflecting a high level of commitment from users to their assigned tasks, while the number of users involved was set at 50. The performance of the VTAS strategy was compared against the baseline algorithms: LNN and NN. For each user availability level, five simulations were conducted with each task allocation algorithm, and the results presented in this subsection represent the mean values obtained from these five executions.

The Task Completion Ratio (TCR) metric can be observed in Table 4.5. The results obtained are similar to those observed in the previous experiment. The VTAS demonstrates better performance in most cases, except in the scenario with 75% of User Availability Probability, where the NN outperforms the VTAS. In all other cases, the VTAS significantly outperforms both the NN and the LNN. Furthermore, it is evident that this is due to the workload limitation, as the performance of the VTAS does not exceed 100%, indicating that it has reached the maximum number of tasks allowed in the campaign, preventing users from completing additional unnecessary tasks. In contrast, the NN exceeds this limit in 75% of the cases, achieving TCR values of 101.9% and 102.2%. However, as in the previous experiment, in Citizen Science domain, it is better to maintain the TCR at 100% than obtain a higher TCR as

it prevents user overload. Another observation is that, in this case, the LNN does not fulfil the workload limitation, reaching a maximum of 91.7% in the TCR metric.

		TCR		
		User Available Probabilities		
		75%	50%	25%
Slot 1	VTAS	100%	94.52%	70.27%
	LNN	91.7%	74.95%	48.62%
	NN	101.9%	83.40%	49.04%
Slot 2	VTAS	100%	95.43%	61.38%
	LNN	90.7%	73.46%	48.62%
	NN	102.2%	82.29%	46.33%

Table 4.5: TCR evaluation over the different User Availability Probabilities using the VTAS, LNN, and NN algorithms. The highlighted results correspond to the algorithms with the highest TCR in each slot for each tested User Availability Probability.

The right-hand side of Figure 4.11 illustrates the realised tasks throughout the first slot of the campaign using the VTAS, LNN, and NN algorithms. Since the campaign consists of two slots, and the task completion objectives were reset at the beginning of each slot, only the results from the first slot are presented. To depict the evolution of task completion, a moving average (MA) of 20 values was applied. The moving average metric is defined in the glossary under **Key Concepts**. This figure demonstrates that the VTAS consistently achieves a higher number of realised tasks compared to the NN and LNN algorithms. The sole exception occurs toward the end of the 75% of User Availability Probability scenario, where the realised tasks for the VTAS decline due to the workload limitation imposed by the campaign. This behaviour mirrors what was observed in Experiment 1, confirming that without the workload limitation, the VTAS would outperform the NN algorithm. As has been mentioned repeatedly, given the context of altruistic participation, the performance of the VTAS is better than the one obtained from the NN, as it enforces participation limits and, consequently, reduces user fatigue.

Table 4.6 shows the MEAN $\pm$ SD of the realised tasks for the VTAS, LNN, and NN algorithms across different User Availability Probabilities. The high-

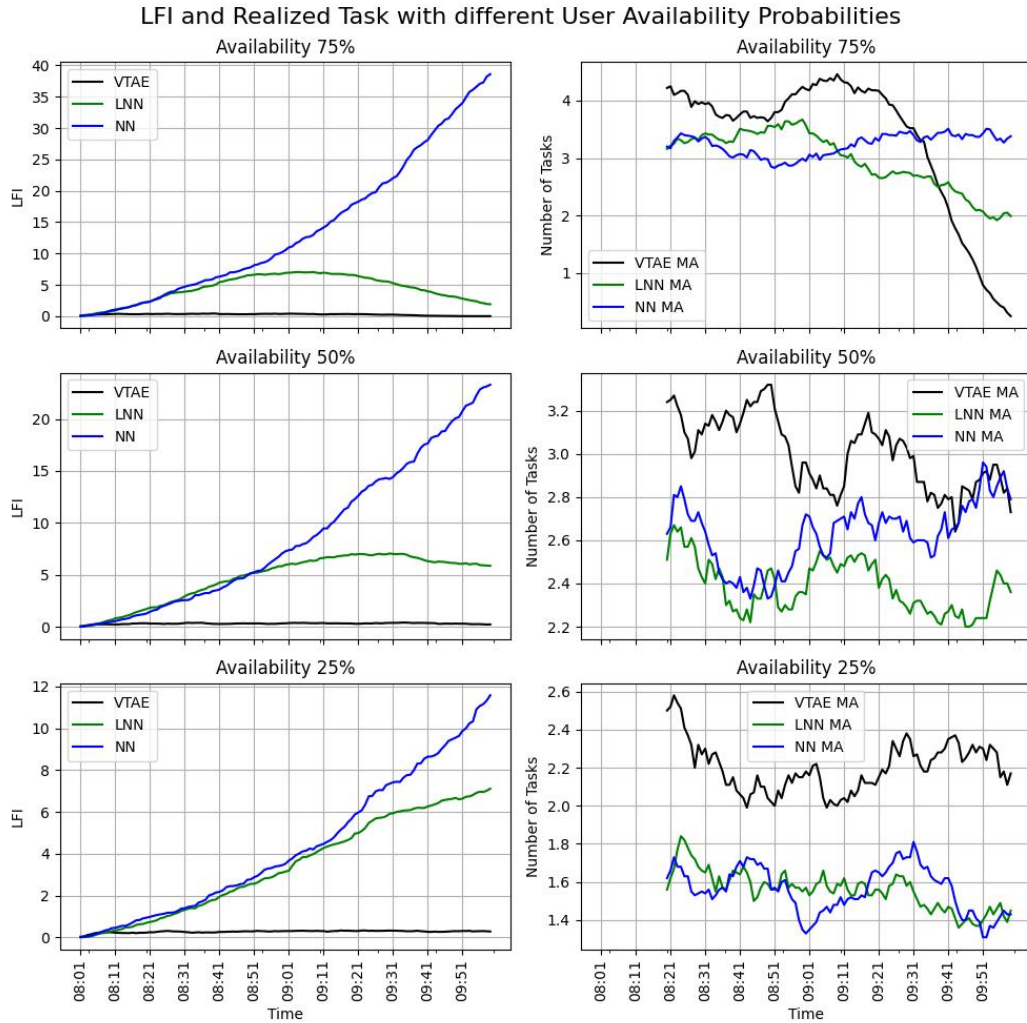


Figure 4.11: The LFI (Eq. 4.4 left figures) and realised task (Eq. 4.3 right figures) over the first slot across different User Availability Probabilities. The algorithms used are the VTAS, the LNN, and the NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used.

lighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability. These values support what was previously mentioned. Given the decline experienced by the VTAS with 75% of availability, the NN outperforms the VTAS in this metric. However, with the other tested probabilities, the VTAS significantly outperforms both the NN and the LNN.

		MEAN $\pm$ SD (RT)		
		User Available Probabilities		
		75%	50%	25%
Slot 1	VTAS	3.16 $\pm$ 1.71	2.99$\pm$0.81	2.22$\pm$0.81
	LNN	2.90 $\pm$ 0.91	2.37 $\pm$ 0.70	1.54 $\pm$ 0.57
	NN	3.22$\pm$0.83	2.64 $\pm$ 0.77	1.55 $\pm$ 0.56
Slot 2	VTAS	3.13 $\pm$ 1.78	2.99$\pm$0.89	1.92$\pm$0.67
	LNN	2.84 $\pm$ 0.96	2.30 $\pm$ 0.73	1.52 $\pm$ 0.52
	NN	3.20$\pm$0.85	2.58 $\pm$ 0.73	1.45 $\pm$ 0.51

Table 4.6: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the VTAS, LNN, and NN algorithms across different User Availability Probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability.

The left-hand side of Figure 4.11 shows the LFI metric throughout the first slot of the campaign using the VTAS, LNN, and NN algorithms. Since the campaign consists of two slots and the task completion objectives were reset at the beginning of each slot, only the results from the first slot are presented. The behaviour of the LNN is similar to that observed in the previous experiment, with one exception: in the 25% scenario, it exhibits exponential behaviour. This is because the workload limitation has not yet influenced the task distribution enough. As a result, few cells have reached the task limit of M_s tasks per slot, and the LFI metric has not started to stabilise. An indication of this is that the TCR of the LNN in this scenario does not exceed 50%. The VTAS, on the other hand, maintains a low LFI value throughout the campaign in all cases.

Table 4.7 presents the MEAN $\pm$ SD of the LFI metric for the VTAS, LNN, and NN algorithms under different User Availability Probabilities. The highlighted results correspond to the algorithms with the lowest mean in each slot for each User Availability Probability. These values support what was previously mentioned. The mean values of the NN in this metric are considerably high. However, the VTAS maintains low LFI values in all cases, indicating that the task distribution across the cells is equitable.

The visual representation of the LFI for all tested probabilities and al-

		MEAN $\pm$ SD (LFI)		
		User Available Probabilities		
		75%	50%	25%
Slot 1	VTAS	0.26$\pm$0.13	0.32$\pm$0.06	0.28$\pm$0.05
	LNN	4.38 $\pm$ 2.13	4.70 $\pm$ 2.27	3.44 $\pm$ 2.33
	NN	13.80 $\pm$ 11.34	8.75 $\pm$ 7.05	4.37 $\pm$ 3.34
Slot 2	VTAS	0.26$\pm$0.14	0.36$\pm$0.09	0.33$\pm$0.07
	LNN	4.51 $\pm$ 2.25	4.35 $\pm$ 2.22	3.33 $\pm$ 2.24
	NN	12.31 $\pm$ 10.25	8.99 $\pm$ 7.40	3.88 $\pm$ 3.11

Table 4.7: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the VTAS, LNN, and NN algorithms across different user availability probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each tested User Availability Probability.

gorithms is provided in Appendix E, Section E.2. These examples have been excluded from this section, as the results are consistent with those obtained in the previous experiment, and the visual representation of the LFI is considered unnecessary. Nevertheless, the reader is encouraged to consult the appendix for a more thorough and complete overview of the results obtained in this experiment.

The results of Experiment 2 demonstrate the efficacy of the VTAS algorithm in managing varying levels of User Availability Probability. The VTAS consistently outperforms both the LNN and NN algorithms in terms of TCR and Realised Tasks (RT), with the exception of the scenario where the User Availability Probability is 75%, in which the NN algorithm exceeds the VTAS due to the absence of workload constraints. However, adhering to workload restrictions is considered more advantageous than achieving a higher TCR and RT metric in the case of a Citizen Science domain. Hence, the VTAS exhibits superior performance with respect to the adherence to workload limitations. These findings confirm that the VTAS successfully achieves Goal 2 across varying User Availability Probabilities. Additionally, the VTAS maintains a low Load Fairness Index (LFI) across all tested probabilities, signifying an equitable task distribution. In contrast, the NN algorithm shows a significantly higher LFI. Thus, the effectiveness of the VTAS in relation to Goal

3 is likewise affirmed. In conclusion, these findings underscore the robustness of the VTAS strategy in ensuring equitable task allocation and maintaining high performance, even under challenging conditions of user availability.

4.3.3 Experiment 3: User Task Completion Probability

This experiment assesses the effectiveness of the VTAS algorithm under varying User Task Completion Probabilities. Campaigns with high User Task Completion Probabilities correspond to higher user commitment to their accepted tasks, resulting in a higher number of realised tasks, which facilitates the achievement of task-completion goals. In contrast, when the User Availability Probability is low, user commitment to their accepted tasks decreases, so more tasks are accepted and not realised, leading to a lower number of realised tasks and reducing the likelihood of successfully achieving the task-completion goal. This section examines how effectively the VTAS strategy satisfies the task-completion goal under different User Task Completion Probabilities (Goal 2), including scenarios where achieving the goal may not be feasible (i.e., in cases of low commitment). In such instances, ensuring an equitable distribution of tasks becomes critical. Therefore, the VTAS's ability to maintain an equitable task distribution was also evaluated (Goal 3).

In this subsection, the performance of the VTAS algorithm, in relation to the defined goals, was evaluated under User Task Completion Probabilities of 75% (high commitment), 50%, and 25% (low commitment). In all cases, the campaign configuration and task-completion goal were defined using the same parameters specified in the previous section. For these campaigns, the User Availability Probability was set at 50%, reflecting a medium level of user motivation, while the number of users involved was set at 50. The performance of the VTAS algorithm was compared against the baseline algorithms, LNN and NN. For each user task completion level, five simulations were conducted, and the results presented in this subsection represent the mean values obtained from these five executions.

The Task Completion Rate (TCR) metric is provided in Table 4.8. The VTAS algorithm consistently outperforms the LNN and NN algorithms, for

all the probabilities tested. Additionally, it can be observed that, in this case, the workload limitation is not reached, as the TCR does not reach 100% for any of the probabilities.

		TCR		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	94.52%	66.22%	34.41%
	LNN	74.95%	51.33%	25.05%
	NN	83.40%	49.10%	23.62%
Slot 2	VTAS	95.43%	62.45%	31.33%
	LNN	73.46%	52.45%	24.63%
	NN	82.29%	50.43%	26.22%

Table 4.8: TCR evaluation over the different User Task Completion Probability using the VTAS, LNN, and NN. The highlighted results correspond to the algorithms with the highest TCR in each slot for each tested User Availability Probability.

The right-hand side of the Figure 4.12 shows the realised task (Eq. 4.3) through the first slot of the campaign with the VTAS, LNN and NN algorithms. This figure shows that the VTAS consistently surpasses the number of realised tasks for the NN and LNN algorithms.

Table 4.9 shows the $\text{MEAN} \pm \text{SD}$ of the realised tasks for the VTAS, LNN, and NN algorithms across different User Task Completion Probabilities. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Task Completion Probability. These values support what was previously mentioned. The VTAS outperforms the NN and LNN algorithms in the number of realised tasks for all the probabilities tested.

The left-hand side of Figure 4.12 shows the LFI metric (Eq. 4.4) throughout the first slot of the campaign for the different probabilities tested with the VTAS, LNN, and NN algorithms. The VTAS maintains a low LFI value throughout the campaign, indicating that the task distribution across the cells is equitable. In contrast, the NN algorithm exhibits an exponential behaviour. In the case of a 25% and 50% probability, the LNN also shows exponential behaviour.

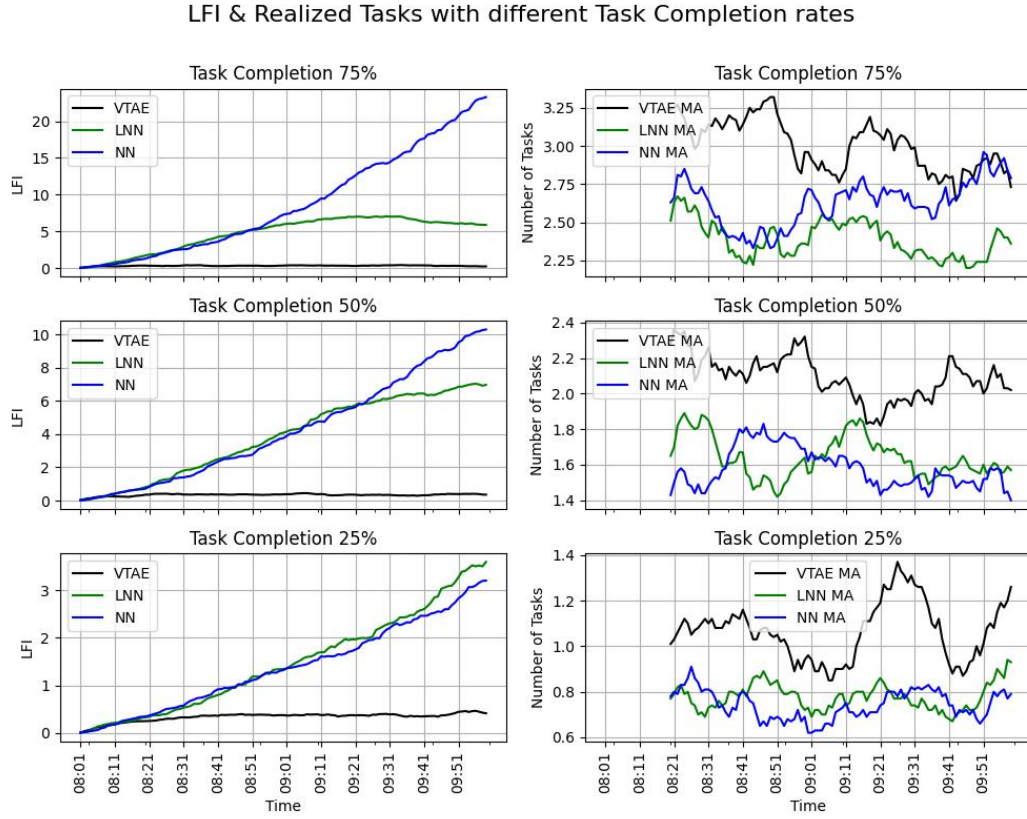


Figure 4.12: The LFI (Eq. 4.4 - left figures) and realised task (Eq. 4.3 - right figures) over the first slot across different User Task Completion Probabilities. The task allocation strategies used are the VTAS, the LNN, and the NN algorithms. To illustrate the evolution of the task realised, the moving average of 20 values was used. The results show the mean of five executions.

Table 4.10 presents the $\text{MEAN} \pm \text{SD}$ of the LFI metric for the VTAS, LNN, and NN algorithms under different User Task Completion Probability. The highlighted results correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability. These values support what was previously mentioned. Given the exponential behaviour of the NN, the mean values of this metric are considerably high. However, the VTAS maintains low LFI values in all cases, indicating that the task distribution across the cells is equitable.

The visual representation of the LFI for all tested probabilities and algorithms is provided in Appendix E, Section E.3. These examples have been

		MEAN $\pm$ SD (RT)		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	2.99$\pm$0.81	2.09$\pm$0.75	1.09$\pm$0.48
	LNN	2.37 $\pm$ 0.70	1.62 $\pm$ 0.56	0.79 $\pm$ 0.37
	NN	2.64 $\pm$ 0.77	1.55 $\pm$ 0.57	0.75 $\pm$ 0.39
Slot 2	VTAS	2.99$\pm$0.89	1.96$\pm$0.69	0.98$\pm$0.50
	LNN	2.30 $\pm$ 0.73	1.64 $\pm$ 0.58	0.77 $\pm$ 0.39
	NN	2.58 $\pm$ 0.73	1.58 $\pm$ 0.61	0.82 $\pm$ 0.36

Table 4.9: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each Task Completion Probability.

		MEAN $\pm$ SD (LFI)		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	0.32$\pm$0.06	0.34$\pm$0.07	0.33$\pm$0.09
	LNN	4.70 $\pm$ 2.27	3.83 $\pm$ 2.35	1.48 $\pm$ 1.04
	NN	8.75 $\pm$ 7.05	4.26 $\pm$ 3.15	1.38 $\pm$ 0.91
Slot 2	VTAS	0.36$\pm$0.09	0.38$\pm$0.10	0.38$\pm$0.12
	LNN	4.35 $\pm$ 2.22	3.65 $\pm$ 2.35	1.36 $\pm$ 0.93
	NN	8.99 $\pm$ 7.40	4.62 $\pm$ 3.51	1.69 $\pm$ 1.20

Table 4.10: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted values correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability.

excluded from this section, as the results are consistent with those obtained in the previous experiments, and the visual representation of the LFI is considered unnecessary. Nevertheless, the reader is encouraged to consult the appendix for a more thorough and complete overview of the results obtained in this experiment.

The results of Experiment 3 demonstrate the efficacy of the VTAS algorithm in managing varying levels of User Task Completion Probability. The VTAS consistently outperforms both the LNN and NN algorithms in terms

of TCR and Realised Tasks (RT) in all the cases. In the case of this experiment the workload limitation is not achieved, so it is clear that the VTAS exceeds the behaviour of the LNN and the NN algorithm in terms of achieving the task-completion goal across varying User Task Completion Probabilities (Goal 2). Additionally, the VTAS maintains a low Load Fairness Index (LFI) across all tested probabilities, signifying an equitable task distribution. In contrast, the NN and LNN algorithm shows a significantly higher LFI, attributable to its exponential behaviour. Thus, the effectiveness of the VTAS in relation to Goal 3 is likewise affirmed. In conclusion, these findings underscore the robustness of the VTAS algorithm in ensuring equitable task allocation and maintaining high performance, even under challenging conditions of user commitment with their accepted task.

4.3.4 Conclusions of the evaluation of the VTAS Algorithm in the Simulation Use Case

During this section, the effectiveness of the VTAS was evaluated in relation to Goal 3 and Goal 2 through three experiments with significantly different characteristics, compared to two baselines: LNN and NN. In all these experiments, the same task-completion goal and campaign configuration were defined, with only the customizable variables being modified. The first experiment involved observing the behaviour of the VTAS with different numbers of users; the second focused on evaluating the performance of the VTAS with varying probabilities of user availability; and the third analysed the VTAS's performance with different probabilities of task completion. In other words, the first experiment assessed the effectiveness of the VTAS in relation to different levels of individual workload, the second focused on the system's effectiveness in varying user motivation levels, and the third experiment measured the VTAS's performance based on user commitment to the tasks they had accepted. In all the experiments, there were scenarios where achieving the task-completion goal was straightforward, as well as cases where it was considerably complex. In all cases, the behaviour of the VTAS was assessed considering its envisaged

main application within the Citizen Science domain, which is the main focus of this thesis.

The results of the three experiments were consistent. The VTAS outperformed both the NN and LNN in all metrics when the workload limitation was not present. In scenarios where this limitation prevented the completion of additional tasks in the VTAS, the NN achieved better performance in terms of TCR and realised tasks (RT). However, it was demonstrated that in the absence of this limitation, the VTAS would outperform the NN. As mentioned in the experiments, it is preferable to respect the workload limitation, even if it results in lower performance in the TCR and realised tasks (RT) metrics. This is because, in the context of Citizen Science, it is crucial that users do not feel overwhelmed, which makes workload limitation fundamental, in addition to being an integral part of Goal 2. Therefore, even in cases where the VTAS does not achieve the best values in TCR and realised tasks (RT)-i.e., when the workload limitation is involved-its behaviour remains superior to the baselines. In the remaining cases, the VTAS showed the best performance in these metrics. Thus, the effectiveness of the VTAS in all the experiments regarding Goal 2 is demonstrated.

With respect to Goal 3, this is measured using the LFI metric. In all the evaluated cases, the LFI of the VTAS was lower than that of the baselines, demonstrating that the VTAS maintains an equitable distribution of tasks across cells. In contrast, both the NN and the LNN exhibited exponential behaviour in LFI, indicating an inequitable distribution of tasks. Consequently, the VTAS meets Goal 3 in all the evaluated cases.

Overall, the VTAS algorithm demonstrated superior effectiveness in ensuring high performance and equitable task allocation across different scenarios, achieving in all of them the Goal 2 and Goal 3. It consistently met task-completion goals while maintaining an equitable workload distribution, proving to be a reliable solution for Spatial Crowdsourcing applications in a Citizen Science domain. The algorithm's adaptability to varying crowd sizes, user availability, and task-completion probabilities highlights its versatility and robustness, making it well-suited to address real-world challenges.

4.4 Deusto Use Case

This section evaluates the VTAS using a real-world scenario involving real users from the University of Deusto¹. In this experiment, participants were tasked with capturing photos of the University of Deusto, in concrete in the Campus of Bilbao, to create a collage. This use case does not originate from a genuine Citizen Science initiative, as it lacks a community-driven need. Instead, it simulates one in which the goal is creating a pictographic representation of the university and its surroundings. Since both the Spatial Crowdsourcing model (detailed at Chapter 3) and the proposed strategy, the VTAS, are agnostic to the specifics of the Citizen Science initiative, they can be effectively applied in this context. Consequently, each task, in this case, involves taking a photo at the designated location. As previously mentioned, this use case provided an opportunity to assess the effectiveness of the VTAS algorithm in fostering altruistic participation within the context of a real-world experiment, as well as to evaluate user adoption and acceptance (Goal 1). Additionally, Goal 2 and Goal 3 were evaluated in this use case to assess the effectiveness of the VTAS in a real-world environment.

For this use case, the MVE engine integrating the VTAS strategy and offering its functionality through a RESTful API was also integrated with a Telegram² bot, specifically developed to interact directly with users. The bot is also available in the GitHub repository³. It was designed to provide users with cell recommendations, allowing them to accept or reject the suggestions. Additionally, the bot offers users information about the campaign's progress, including visual updates, and guides them through the steps of the interactive workflow.

A Telegram bot is an automated program designed to operate within the Telegram messenger application, allowing users to interact with it through messages and commands. These bots can perform a variety of tasks, such as sending notifications, managing groups, processing payments, or integrating with other services. In this case, the Telegram bot was used to interact with

¹<https://www.deusto.es/en/home>

²<https://web.telegram.org/a/>

³<https://github.com/mpuerta004/RecommenderSystem>

users and provide them with cell recommendations, giving them the option to accept or reject these suggestions. Additionally, the bot offered information about the campaign's progress through visual updates and guided users through an interactive workflow. A key feature of the bot was requesting users' location, both to recommend tasks nearby and to ensure that the submitted photos matched the assigned cell. This made the interaction with the system more user-friendly and seamless. Through text commands, users could perform actions such as requesting recommendations, accepting or rejecting suggestions, viewing a map composed of photos, and submitting images related to the recommended cells they had accepted. The bot interpreted these commands and responded efficiently, enhancing the user experience. A detailed explanation of this bot can be found in Appendix C.

For this experiment, 12 participants were recruited using a convenience sampling method from the Deusto University staff. Each participant booked two hours of time, during which they received a training presentation lasting thirty minutes. The remainder of the time was allocated for conducting the experiment. During the presentation, the goals of the experiment and the functionality of the Telegram bot were explained. Upon completion of the training, participants began the campaign.

The campaign defined for this use case included the following parameters: The experiment was conducted on February 23 of the 2024, from 10:25 to 12:00, centred at the University of Deusto's campus at Bilbao (longitude: -2.9396777560573013, latitude: 43.271157163906985). The parameter C_d was set to 0.05km, and the campaign's radius was 0.18km. The M_s value was set to 4, therefore, the expected individual effort is 7 tasks per person. Certain cells were excluded from the campaign due to inaccessibility. The campaign consisted of a single slot. A visualization of the campaign is presented in Figure 4.13, which also includes the final collage created by the volunteers. The campaign concluded either when users stopped receiving cell-recommendations for the workload limitation or when the campaign time expired.

The aim of this experiment with the VTAS strategy was to increase user acceptance and adoption of Spatial Crowdsourcing technology. Two methods were employed to evaluate the VTAS:



Figure 4.13: The final collage of photos obtained in the Deusto use case using the VTAS algorithm.

- **Simulation vs real-world Evaluation:** To validate the simulation and understand the differences it may have compared to real-world experimentation, the campaign was simulated using the VTAS, LNN, and NN algorithms. These simulations were compared with the real use case to understand the differences and similarities between both.
- **Survey Evaluation:** This evaluation assesses user acceptance and adoption at the conclusion of the experiment of the Spatial Crowdsourcing technologies apply to Citizen Science domain. At the end of the campaign, all participants completed a questionnaire designed to evaluate their acceptance and adoption of the Spatial Crowdsourcing tools, as well as to capture other relevant parameters for understanding users' perspectives. This survey inspired by the TAM (Davis et al., 1989) and SUS (Grier et al., 2013) metrics can be found in Appendix B.

The following of this section is organized as follows: Subsection 4.4.1 presents the simulation vs real-world evaluation of the Deusto use case. Sub-

section 4.4.2 presents the survey evaluation. Finally, Subsection 4.4.3 presents the conclusions of the Deusto use case.

4.4.1 Simulation vs real-world Evaluation

In this section the real-world campaign, and its comparison with the simulated campaign using the VTAS, LNN and NN is presented. This section aims to validate the simulation and understand the differences it may have compared to real-world experimentation.

In the real-world experiment, the task allocation process began at 10:25 and concluded at 11:15, as users stopped receiving cell recommendations once the tasks-completion goal was met before the campaign ended. The goal, based on the variable M_s , was to obtain 84 images, which translates to an individual effort of 7 tasks per person. However, this goal was exceeded, resulting in a total of 92 photos collected by the end of the experiment.

Two factors contributed to exceeding the workload limitation: (1) When a cell recommendation is generated for a user, another participant may accept it before the first user has a chance to evaluate it. If both users accept the recommendation, the total number of tasks can exceed the workload limit, even if the original recommendation respected the limit. (2) The system accepts photos taken in the wrong cell, even when they were supposed to be captured in designated areas. This flexibility was introduced to prevent user frustration, particularly in cases where some cells had difficult accessibility or when GPS calibration was inaccurate on mobile devices.

These factors are illustrated in the final collage (Figure. 4.13), where one cell in the top left corner contains only three photos. This is because the only accessible location within the cell can be easily confused with an adjacent cell if GPS calibration fails. While these conditions led to exceeding the workload limit, the overall effort remained controlled, as evidenced by users stopping to receive cell recommendations before the campaign ended.

To compare the effectiveness of this experiment with the simulated one, a simulation of this campaign was recreated using the baselines (LNN and NN) and the VTAS. For this simulation the variable User Task Completion

Probability was set at 75%, the User Available Probability was set at 50% and the number of users was set to 12.

Table 4.11 shows the Task Completion Ratio (TCR) (Eq. 4.2) metric for this campaign, using the four algorithms. This table show that the real-world VTAS has the highest TCR, while the VTAS simulation has the highest TCR among the simulations. The real-world VTAS exceeded the workload limitation, indicating a higher-than-expected level of user engagement and motivation during the simulation. The real-world VTAS concluded at 11:15, surpassing the tasks-completion target, while the simulations had not met this goal by the end of the campaign at 12:00. This discrepancy may be due to the simulation itself. The time intervals defined for determining the pace of activities in the simulation were set to replicate users following a routine. As a result, it was not designed to simulate activity as rapid as that observed in real life. In other words, within this campaign, users can complete multiple tasks within a single minute, whereas in the simulation, each task requires at least two minutes due to the way action times have been modelled. This difference in timing may be the cause of the variation in the TCR. However, over time, it is expected that participation will stabilise, gradually adopting a pattern more similar to the simulated behaviour. Nevertheless, this assumption has not been tested in this experiment. The only definitive conclusion that can be drawn is that real-world behaviour differs from the simulated model, as evidenced by the sudden spike in activity.

		TCR
Real-world	Deusto Campaign	111,90%
Simulations	VTAS	73,81%
	LNN	70,24%
	NN	57,14%

Table 4.11: TCR metric obtained in the Deusto experiment and in the simulations of this campaign using the VTAS, LNN, and NN algorithms.

The right-hand side of Figure 4.14 illustrates the moving average of the number of realised tasks per minute (Equation 4.3) throughout the experiment. To illustrate the evolution of the task realised, the moving average of

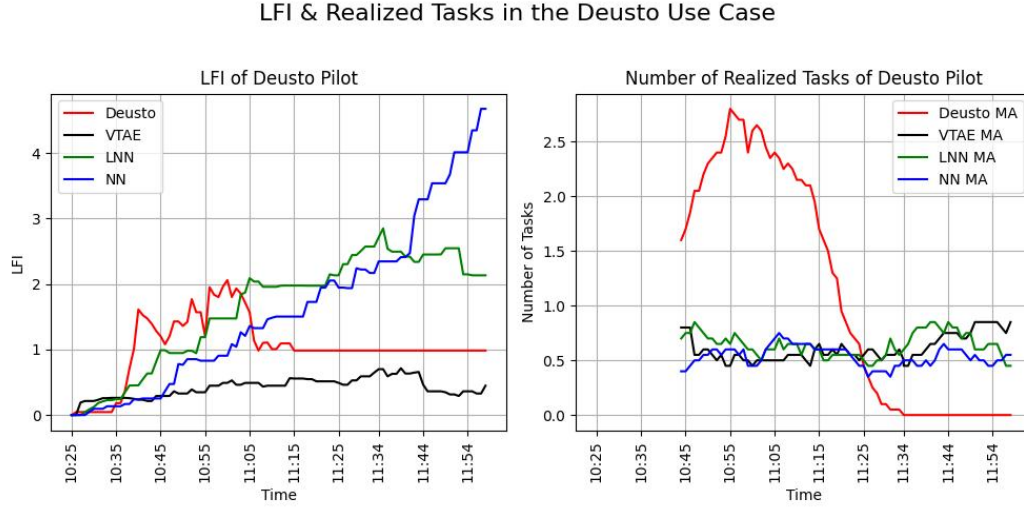


Figure 4.14: The LFI metric (Eq. 4.4- left picture) and Realised tasks(Eq. 4.3 - right picture) in the Deusto real-world campaign (referred to Real-world VTAS) and the results obtained in its simulation using the VTAS, LNN, and NN. To illustrate the realised task over time, the moving average (MA) of 20 values was used in the right hand side pictures.

20 values was used. This data includes results from the real-world VTAS experiment, labelled in the figure as Deusto, as well as the simulated LNN, VTAS, and NN algorithms. From the figure, it is evident that the real-world VTAS campaign shows a sharp peak in the number of realised tasks per minute, reaching a moving average of 2.75 tasks per minute. This peak occurs before the rate drastically decreases around 11:15, when users complete the task-completion goal of the campaign. In contrast, none of the simulated algorithms surpass a value of 1.0 realised tasks per minute. Furthermore, the simulated algorithms display very uniform behaviour, with no significant differences between them, making it challenging to identify which algorithm performs better based solely on this graph. As a result, a clear comparison between the performance of the simulated algorithms and the real-world VTAS experiment's behaviour cannot be drawn from this data. Despite all this, the data demonstrate that the VTAS algorithm, in a real-world context, still adheres to a workload limitation even while exceeding the task-completion goal of the campaign. This indicates that the VTAS meets Goal 2 in a real-world

environment.

The $\text{MEAN} \pm \text{SD}$ of this metric, presented in the last column of Table 4.12, supports these observations. Both the MEAN and SD are significantly higher in the case of the real-world VTAS compared to the simulated algorithms. In the real experiment, the mean is 1.0, whereas in the simulated algorithms it does not exceed 0.7. Regarding the SD, the value for the real-world VTAS is 2.41, far higher than the maximum of 0.72 recorded for the simulated algorithms. These differences are considerable, as shown in the figure. It is also observed that the $\text{MEAN} \pm \text{SD}$ values for the three simulated algorithms are very similar, indicating that there are no significant differences between them. The remarkable peak observed in the real-world VTAS experiment can be attributed to the high motivation and strong commitment of the users to complete the task objective within a short period of time.

The left-hand side of the figure 4.14 illustrates the LFI (Eq. 4.4) throughout the campaign for the real-world VTAS, VTAS, LNN and NN. This figure illustrates the consistently low LFI of the VTAS compared to the other algorithms, including the results from the real-world VTAS. The LFI of the real-world VTAS campaign exhibited a peak at nearly 2 before declining rapidly. The LNN and NN algorithms exhibit the highest LFI, characterized by exponential behaviour that indicates a non-equitable distribution of tasks across cells. Again, for this metric, the LFI behaviour of the VTAS simulation differs from that of the real-world VTAS. One of the factors that has contributed to the increase in LFI in the case of the real-world VTAS is the design of the Telegram bot, which allows the duplication of certain tasks and, consequently, exceeds the workload limitation. This duplication of multiple tasks significantly increases the metric. Another contributing factor is the limited accessibility of the upper-left cell, which has prevented users from completing all the tasks associated with that cell. Furthermore, this situation has led to some photos from that cell being associated with the adjacent cell, further increasing the LFI value. Despite these factors, the LFI of the real-world VTAS remains within manageable levels, as the metric decreases rapidly after the peak. This indicates that the VTAS algorithm is capable of maintaining the LFI under controlled values in a real-world environment, achieving Goal 3.

Table 4.12 presents the MEAN $\pm$ SD of LFI metric (Eq. 4.4). This table verifies the observations: the simulated VTAS has the lowest mean and SD, while the NN has the highest MEAN $\pm$ SD. Similarly, it is confirmed that the MEAN $\pm$ SD of the real-world VTAS are the second-lowest among the algorithms tested. In addition, the elevated LFI of the real-world VTAS can be attributed to the large number of tasks completed in a short time during the campaign, as well as the bot’s capability to exceed the workload limitation.

		MEAN $\pm$ SD (LFI)	MEAN $\pm$ SD (RT)
Real-world	real-world VTAS	1.03 $\pm$ 0.48	0.97$\pm$1.55
	VTAS	0.42$\pm$0.15	0.65 $\pm$ 0.87
Simulations	LNN	1.64 $\pm$ 0.85	0.62 $\pm$ 0.76
	NN	1.64 $\pm$ 1.27	0.51 $\pm$ 0.68

Table 4.12: MEAN $\pm$ SD of the LFI and RT metrics obtained in the Deusto experiment (real-world VTAS) and in the simulations of this campaign using the VTAS, LNN, and NN algorithms.

In conclusion, the experiment was significantly influenced by several factors, including the high motivation and engagement of users, who displayed strong determination to complete the task within a short time. The progression of tasks revealed an unexpectedly high level of user involvement and commitment, likely driven by the novelty effect or the ‘Wow’ effect. Despite this heightened participation, the LFI observed in the real-world VTAS remained within manageable levels. While certain aspects of the simulation align with the real-world VTAS behaviour, an analysis of the metrics in Tables 4.11 and 4.12 and in the Figure 4.14 indicates that the simulation’s behaviour does not accurately replicate human behaviour. This discrepancy stems from the inherent complexity of human behaviours, as well as from the design of the simulation and the defined time sets. However, the conclusion is that the simulation cannot be fully replicated within a simulated environment. Consequently, the simulated behaviour diverges from the behaviour observed in the algorithm, but both have similarities in maintaining the LFI under controlled values, archiving the Goal 3. In addition, the data demonstrate that the VTAS strategy, in a real-world context, achieves the task-completion goal

of the campaign while still adhering to a workload limitation. This indicates that the VTAS meets Goal 2 in a real-world environment.

4.4.2 Survey Evaluation

This section describes the evaluation collected through the survey realised at the end of the campaign to evaluate user adoption and acceptance. As mentioned above, 12 users participated and were recruited using a convenience sampling method among the university staff. The questionnaire consisted of 27 questions, and its goal was to evaluate the users' acceptance and adoption of the system and their perception of it (Goal 1).

	%	N. of people
Age		
19-35	25%	3
36-50	75%	9
Others	0%	0
Gender		
Female	25%	3
Male	75%	9
Other	0%	0
Level of education		
PhD	58.3%	7
Undergraduate	8.3%	1
Master's degree	33.3%	4
Others	0 %	0
Technological knowledge		
Intermediate Level	8.3%	1
Advanced Level	91.7%	11
Others	0	0
Work Status		
Unemployed	8.3%	1
Student	8.3%	1
Employed	83.3%	10
Others	0 %	0

Table 4.13: Demographic information of the Deusto use case.

The survey starts by requesting the demographic information, as outlined

in the Table 4.13. A significant majority of participants are employees of the university (83.3%), males (75%), with a high level of technological abilities (91.7%), and fall within the 36-50 age range (75%). An advanced level of technological proficiency indicates that the participant engages in activities such as making purchases online and using spreadsheets, while an intermediate level reflects tasks like browsing websites and watching videos on platforms like YouTube. The convenience sampling method employed for recruitment focused on the university staff, which led to a relatively homogeneous participant pool, as revealed by the demographic information.

Question 1.1: I think this activity was interesting to me.

Options:	Disagree	1	2	3	4	5	Agree
Answers:		0	0	1	5	6	
Mean± SD: 4.41 ± 0.65							

Question 1.2: I think this activity was useful to me.

Options:	Disagree	1	2	3	4	5	Agree
Answers:		0	1	6	2	3	
Mean±SD: 3.58 ± 0.95							

Question 1.3: I think that this activity dealt with an important topic for me.

Options:	Disagree	1	2	3	4	5	Agree
Answers:		0	1	3	3	5	
Mean±SD: 4.0 ± 1.0							

Question 1.4: Comments or recommendations regarding this activity that you would like to share.

Final Score of participant's interest level:
Mean± SD: 3.99 ± 0.87

Table 4.14: Inquiries about the participant's interest level in the Deusto experiment. This table shows the users' questions, possible options, and answers. Question where answers is a scale of the 1 to 5, where 1 is the user strongly disagrees (low satisfied) and 5 is the user strongly agrees (highly satisfied).

Following the collection of demographic information, participants responded to several questions to assess their satisfaction and interest in the activity. The four questions related to this topic, possible options, and results are dis-

played in Table 4.14. The results of Questions 1.1, 1.2, and 1.3 show that the activity was enjoyable and valuable for the participants as the mean of the results are 4.41, 3.58 and 4 over 5, with 1 indicating the user strongly disagrees (unsatisfied) and 5 indicating the user strongly agrees (highly satisfied). In the case of the open question (Question 1.4), some participants expressed confusion regarding the goal of the activity, and they encouraged us to improve this part of the presentation that was give before the campaign to contextualize it. Another recommendation proposes creating challenges for each task or gamifying the bot. These comments reflect that users are interested in the experiment but see possible improvements. It is a good insight that proves that they see potential and possible applications. To facilitate comparisons between the results of this study and those of future strategies, we calculated the overall mean of the means and standard deviations of all the questions in this section which are not open answers. Specifically, for this section os questions, which assesses participants' level of interest, the overall mean of the individual question means is 3.99 (where 1 represents strong disagreement or dissatisfaction, and 5 represents strong agreement or high satisfaction), with a standard deviation of 0.87. This result suggests that participants were highly satisfied with the activity, reflecting a positive perception of the experiment. Such a high level of satisfaction is an encouraging indicator for the future development of the system, as it demonstrates user engagement and highlights the potential for further enhancements.

The subsequent section of questions focused on system adoption and potential future use. These questions, possible options, and the answers are represented in Table 4.15. Based on Questions 2.1, 2.3, and 2.5, the future adoption of the system by the participants is assessed. With free-text, Questions 2.2 and 2.3, users expressed what they need to ensure future adoption and their desire for improvement. Among these responses, some users expressed concerns about the bot's battery consumption, while others suggested incorporating rewards or challenges for completing tasks or enabling the bot for additional purposes. These comments indicate a desire for further enhancements, reflecting the users' interest in future improvements and use. The user adoption is positive, with many looking forward to more advanced features in

Question 2.1: Do you see yourself continuing to use the recommender system and in the future?

Options: Yes Maybe No

Answers: 5 5 2

Question 2.2: If no, what would be the primary reason for discontinuing the use of the recommender system?

Question 2.3: How likely are you to adopt the telegram bot or participate in the next campaign?

Options: Disagree 1 2 3 4 5 Agree

Answers: 1 0 1 2 8

Mean± SD: 4.33 ± 1.18

Question 2.4: What factors influence your intention to adopt the bot services in your daily life?

Question 2.5: Would you recommend the recommendation system to others? Or to use in other applications or propose?

Options: Yes Maybe No

Answers: 7 5 0

Final Score of the user's adoption:

Mean± SD: 4.0 ± 1.20

Table 4.15: Inquiries concerning the system's adoption by participants and their perception of it. This table shows the users' questions, possible options, and answers. Question where answers is a scale of the 1 to 5, where 1 is the user strongly disagrees and 5 is the user strongly agrees.

the system. The final score for this section is 4.0 ± 1.20 , indicating a high level of adoption of the system and a strong likelihood of future use. To calculate this metric, numerical values were assigned to the responses: 'Yes' was given a score of 5, 'Maybe' a score of 3, and 'No' a score of 1. These values were then evaluated to determine the final result. This high level of satisfaction is a positive indicator of the system's future adoption.

The subsequent section of questions focused on the system's customization and users' perceptions. This set comprised five questions, possible options and the answers can be shown in Table 4.16. The scores slightly exceed the mid-

Question 3.1: How personalised do you find the recommendations provided by the Telegram bot?

Options: Not at all 1 2 3 4 5 Extremely
Answers: 0 2 4 3 3
 Mean± SD: 3.58±1.04

Question 3.2: To what extent do you feel that the recommendations align with your preferences?

Options: Not at all 1 2 3 4 5 Completely
Answers: 0 2 4 4 2
 Mean± SD: 3.5±0.96

Question 3.3: How do you perceive the role of personalised recommendations in enhancing your overall experience with the Telegram bot?

Options: Little 1 2 3 4 5 Extremely
Answers: 1 0 5 5 1
 Mean± SD: 3.42±0.95

Question 3.4: Are there any specific instances where you felt the recommendations were particularly well-tailored to your needs?

Question 3.5: What improvements, if any, would you suggest to enhance the personalization of recommendations provided by the Telegram bot?

Final Score of users' perceived customization of the system:
 Mean± SD: 3.5±0.98

Table 4.16: Questions related to the users' perceived customization of the system. This table shows the users' questions, possible options, and answers. Question where answers is a scale of the 1 to 5, where 1 is the user strongly disagrees and 5 is the user strongly agrees.

point, suggesting that users perceive a degree of adaptation and customization within the system. These results indicate that the system's customization features are progressing in the right direction and hold promise for further refinement and enhancement. The relatively low scores were anticipated, as the only factor used for personalization was the user's position. Consequently, the outcome aligns with the expectations. The final pair of questions in this section delved deeper into the user perception of personalization. Users perceived

the recommendations as highly random. Valuable feedback was gathered to create future improvements, such as the ability to specify a preferred area of interest and system customization options to request recommendations for nearby locations due to tiredness or time constraints. These ideas point towards more dynamic and user-centred customization features. In conclusion, users perceive some personalization, but in the future, they will want more customization and flexible interfaces that allow them to define their preferences. The final score for this section is 3.5 ± 0.98 , indicating that participants perceived a moderate to high level of customization in the system.

Question 4.1: What works well?

Question 4.2: What does not work?

Question 4.3: What could be improved/added?

Question 4.4: What is missing?

Question 4.5: How would you improve it?

Table 4.17: Feedback about the Telegram Bot.

The subsequent section focused on a series of open questions about the Telegram bot’s performance. The aim was to understand the bot’s usability strengths and weaknesses to assess whether this interaction mechanism would be valid for further experiments. All of these questions were open questions that allowed users to express their opinions freely. These questions are shown in Table 4.17. Several users reported occasional moments of slow system performance and longer-than-expected map loading times. This issue was attributed to the computational constraints (latency throughput) of the device running the RESTful API to connect the VTAS, a known issue that users were alerted about beforehand. In terms of further improvements, suggestions included developing a better Telegram keyboard system for potential replies, which could streamline interactions and make the bot more user-friendly. Another suggestion was to indicate the user’s position within the interaction to enhance the user experience. Even though users pointed out certain areas for improvement, all of them concluded that the system worked properly.

Question 5.1: Is there anything else you would like to share about your experience with the recommender system?

Question 5.2: What works well?

Question 5.3: What could be improved/added?

Question 5.4: What is missing?

Question 5.5: How would you improve it?

Table 4.18: Recommendations and feedback information.

The final set of questions was posed in the context of the recommendation of the VTAS, shown in Table 4.18. Some questions are similar, yet their aims differ, as they assess various aspects of how users receive and perceive the recommendations. Users provided valuable feedback and identified several areas for improvement, which can guide the system’s future development. All of these questions were free-text to allow users to express their opinions freely. Many users desired more personalised experiences, which could be achieved by explicitly definition of their interests. They suggested that allowing users to specify areas of interest or directly communicate their preferences would enhance the system’s relevance and usability. Furthermore, the feedback highlighted the potential for an engagement strategy, possibly through feedback mechanisms or gamification, to increase user interaction and satisfaction. Overall, the user feedback underscores the importance of customization and engagement strategies in enhancing the recommender system’s effectiveness.

The findings from the Deusto use case survey provide valuable insights into user perception, acceptance, adoption, and potential system enhancements. The demographic data revealed a homogeneous participant pool, primarily composed of technically proficient individuals. Survey results indicated significant interest in the activity and adoption, while users also expressed a continued desire for improvements. As the system currently lacks extensive customisation based on user features, the perception of its customisation was rated as average, as expected. However, users did not provide negative feedback regarding the cell-recommendation feature, suggesting a level of user comfort.

Despite user comments and suggestions, the overall perception reflects comfort with the system and a desire for sustained usage, indicating a positive outlook towards continued utilisation. This is supported by the quantitative data obtained from the survey, which demonstrated that the VTAS successfully meets Goal 1 in a real-world environment. However, although there were no direct negative comments, a recurring need for improvements and personalisation was noted in all the open responses, highlighting users' desire for a strategy that adapts to their needs. This should be considered in future developments of the solution.

4.4.3 Conclusions of the evaluation of the VTAS in the Deusto Use Case

The Deusto use case provided valuable insights into the real-world application of the VTAS strategy. This study simulated a Citizen Science initiative in which users participated in a data collection task (taking photographs) to create a representative collage of the University of Deusto. To facilitate the campaign, a Telegram bot was implemented to guide and interact with users throughout the process. The VTAS dynamically assigned image capture tasks across different cells, allowing participants to select their preferred locations and submit photos via the Telegram bot, which generated the collage in real-time.

First, the efficiency of the VTAS was evaluated in a real environment to achieve Goal 2 and Goal 3. The real-world VTAS experiment demonstrated high levels of user motivation and commitment, resulting in a strong Task Completion Rate (TCR - Eq. 4.2) and a substantial number of successfully Realised Tasks (RT - Eq. 4.3). Despite minor challenges, such as occasional task duplication and GPS inaccuracies, the system effectively achieves the task-completion goal. Although the VTAS slightly exceeded the predefined workload limitation and the task-completion goal, it maintained controlled effort levels by halting data collection once all cells accumulated just over M_s photos. This mechanism ensured that the VTAS not only met the task-completion goal of the campaign but also imposed a clear limit on user effort,

thereby fulfilling Goal 2. Furthermore, the rapid completion of the assigned tasks underscored the strong engagement and enthusiasm of participants. The initial surge in user participation significantly impacted the Load Fairness Index (LFI - Eq. 4.4) metric, causing fluctuations in workload distribution over the cells. However, over time, the VTAS effectively stabilized this metric, achieving an equitable task allocation. This demonstrated the system's ability to regulate participation and meet Goal 3.

Additionally, data from the real-world experiment was compared with a simulated version of the same campaign using VTAS, LNN, and NN to validate the simulations and understand their differences from real-world behaviour. The comparison of these four campaigns (the real-world campaign and the three simulated ones) revealed that the simulation did not accurately replicate human behaviour. This discrepancy arises from the inherent complexity of human behaviours, as well as the design of the simulation and the defined time parameters. However, the VTAS simulation in a real environment exhibited similarities to the simulated versions in certain aspects; for instance, it imposed workload limitations and maintained the LFI metric within controllable levels. Therefore, while the simulation does not fully replicate human behaviour, it provides valuable insights into the evaluated strategy's performance within a controlled environment.

Finally, to assess the effectiveness of the VTAS in achieving Goal 1, a questionnaire based on the Technology Acceptance Model (TAM) (Davis et al., 1989) and the System Usability Scale (SUS) (Grier et al., 2013) was used to evaluate user adoption and acceptance. A summary of the quantitative results from the questionnaire is presented in Table 4.19. These results indicate that users are comfortable with the system, accept it, and show good adoption. However, the qualitative findings highlight the need for improvements in user integration to ensure their comfort. This urgent need for enhancements could lead to users abandoning the system if proper measures are not implemented. Therefore, although the VTAS meets Goal 1, it is essential to enhance the solution to sustain user engagement and motivation.

Section	MEAN $\pm$ SD
Satisfaction and interest (Acceptance)	3.99 $\pm$ 0.87
Adoption	4.0 $\pm$ 1.20
Perceived customization	3.5 $\pm$ 0.98

Table 4.19: The MEAN $\pm$ SD of each quantitative section of the questionnaire.

4.5 Summary and Conclusions

This chapter presented the Volunteer Task Allocation Strategy (VTAS), a novel spatio-temporal task allocation strategy designed to address the challenges of task allocation in Spatial Crowdsourcing (SC) within the context of altruistic participation. The VTAS aims to enhance user perception, optimize task allocation, and maintain an equitable distribution of tasks across spatial locations. The effectiveness of the VTAS was evaluated through both simulated and real-world use cases. The evaluation of the VTAS algorithm in both the simulated and real-world Deusto use cases has demonstrated its effectiveness in achieving the primary goals defined in this thesis: equitable task distribution (Goal 3) and maintaining an optimal workload balance for participants (Goal 2). Additionally, in the real-world scenario, user acceptance and engagement were assessed, validating the users' acceptance and adoption (Goal 1).

In the simulated use case, the VTAS consistently outperformed the baseline algorithms (LNN and NN) across different experimental conditions. It demonstrated its ability to maintain high Task Completion Rates (TCR - Eq. 4.2). Even in scenarios where workload limitations affected performance metrics, the VTAS remained the preferred approach, as it prioritized preventing user overload, a crucial aspect in Citizen Science applications. Moreover, the algorithm successfully minimized the Load Fairness Index (LFI - Eq. 4.4), achieving a more equitable allocation of tasks across different cells. These results confirm the VTAS's robustness and versatility in adapting to diverse conditions such as varying user availability and commitment levels.

The Deusto use case further validated the VTAS's real-world applicability. The integration with a Telegram bot facilitated user participation in a Citizen

Science initiative, demonstrating high engagement levels and successful task completion. The comparison between the real-world and simulated scenarios revealed that the VTAS maintained its efficiency in practical applications. Despite minor operational challenges, the system effectively regulated task assignments, ensured task completion, and stabilized workload distribution over time. The positive reception among participants, as assessed through a TAM and SUS-based questionnaire, reinforced the users' acceptance and adoption of the system. However, feedback from participants in the real-world experiment indicated a need for technological improvements, particularly in refining the task assignment mechanisms and enhancing the Telegram bot interface. These suggestions highlight opportunities for further refinement, ensuring an even more seamless and engaging user experience in future applications.

Furthermore, the comparative analysis between the real-world and simulation experiments highlighted the adaptability of the VTAS in various contexts. While the simulation provided a controlled environment to systematically test different parameters, the real-world experiment introduced unpredictable variables such as user behaviour and external interferences. However, even though the comparison between the real-world experiment and the simulations revealed notable differences in user behaviour and system performance, the VTAS's performance remained consistent, demonstrating its robustness and practical applicability.

Overall, the VTAS algorithm has proven to be a reliable and adaptable solution for Spatial Crowdsourcing applications within the Citizen Science domain. It effectively balances task allocation, optimizes task completion goal, and enhances user experience, making it a valuable tool for large-scale participatory data collection. Future research could explore further optimizations and extensions of the VTAS to enhance its adaptability across different domains and real-world constraints.

In relation to the goals established for this solution (Goal 1, Goal 2, Goal 3), it can be concluded that Goal 2 and 3 are satisfactorily achieved. However, the fulfilment of Goal 1 is less evident. The constant feedback regarding the need to improve the solution's personalization and integration suggests that it is not perfect and has room for optimization. Moreover, this feedback

highlights the importance of ensuring that users perceive the solution as non-arbitrary and tailored to their specific needs. For this reason, the following chapter will propose a new solution to address these shortcomings and better fulfil Goal 1.

We are what we repeatedly do. Excellence, then, is not an act, but a habit.

Aristóteles

CHAPTER

5

Bio-inspired Volunteer Task Allocation Strategy: Bio-VTAS

Spatial Crowdsourcing, in its traditional form, is not well-suited for applications within Citizen Science due to a fundamental mismatch between their underlying principles—particularly in terms of incentives and user engagement. While Spatial Crowdsourcing typically depends on financial compensation to drive participation, Citizen Science relies on volunteers who are motivated by altruism, curiosity, or a sense of community. **This fundamental difference necessitates a rethinking of task allocation strategies, shifting away from purely economic models towards user-centred approaches that prioritise engagement based on individual preferences and intrinsic motivation.** As previously discussed, strategies designed for this context must integrate users into the decision-making process while maintaining a balance between task efficiency and a positive user experience. Ensuring long-term engagement requires not only avoiding financial incentives but also preventing cognitive

overload by respecting user limits. In response to these challenges, this thesis introduces three collaborative strategies aimed at improving the acceptance and adoption of Spatial Crowdsourcing in Citizen Science while maintaining efficiency in task execution.

In the previous chapter, the first of these strategies, VTAS, was introduced and evaluated based on three key goals that validate our hypothesis: Goal 1, which aims to enhance user perception by increasing acceptance and adoption; Goal 2, which seeks to optimise task allocation by maximising task completion without exceeding workload limitations; and Goal 3, which ensures equitable task distribution. The evaluation showed that VTAS successfully achieved Goal 3, as reflected in its low Load Fairness Index (LFI) value, indicating a balanced task distribution across all cases. Additionally, it excelled in the Task Completion Rate (TCR) metric, fully accomplishing Goal 2 and surpassing other algorithms as long as workload limits were not exceeded. Even when these limits were reached, VTAS demonstrated a performance more suitable for the Citizen Science domain than purely numerical optimisation approaches. Regarding Goal 1, users expressed acceptance and a willingness to adopt the system. However, an analysis of their responses revealed that, although users did not voice outright dissatisfaction, they did not feel completely comfortable with the system and identified the need for additional features to maintain their commitment. As a result, while the VTAS achieved Goal 1, users suggested that enhancements were necessary to sustain engagement and prevent potential demotivation.

For this reason, this chapter introduces a new Spatial Crowdsourcing strategy called the Bio-inspired Volunteer Task Allocation Strategy (Bio-VTAS), which aims to actively engage users in the decision-making process, fostering a more user-centric solution compared to VTAS and further enhancing its performance concerning Goal 1. Unlike VTAS, which primarily considered the user's position, Bio-VTAS incorporates their daily routines to recommend tasks that are more suitable and context-aware. The intended behaviour of this new approach is to match or even surpass the results achieved with the previous strategy while maintaining the results for Goal 2 and Goal 3 and providing a more personalised and user-centric experience (improving

Goal 1). However, given this shift towards a stronger emphasis on user involvement, it is possible that improvements in Goal 1 may come at the expense of Goal 2 and Goal 3. In summary, **although VTAS already demonstrated a viable balance aligned with our hypothesis, Bio-VTAS seeks to refine this balance further by integrating users more deeply into the task allocation process.** While this may lead to a slight decrease in performance regarding Goal 2 and Goal 3, it is expected to significantly enhance user engagement and satisfaction, leading to better long-term adoption of the algorithm.

For further details, the Spatial Crowdsourcing model supporting these strategies is outlined in Chapter 3, which discusses the spatial and temporal constraints, task completion objectives, and workload limitations inherent to Spatial Crowdsourcing campaigns.

As in the last chapter, this chapter provides a comprehensive explanation of the Bio-VTAS and its evaluation. The evaluation is based on two primary approaches: a simulation and a real-world assessment. The chapter is organised as follows: Section 5.1 presents the proposed collaborative solution, Bio-VTAS. Since the baselines used to compare the algorithm, the metrics employed, and the use cases considered for the evaluation of this solution are the same as those described in the previous chapter, this explanation will be omitted in the current chapter. For a detailed description of these elements, it is recommended to refer to Section 4.2 of the previous chapter. The only change introduced in this chapter is that the VTAS will now be used as an additional baseline alongside the LNN and the NN. Section 5.2 details the simulation use case, which assesses Bio-VTAS's performance in a controlled environment, while Section 5.3 presents a real-world use case evaluation. Finally, Section 5.4 provides a summary of the key findings of this study. To better understand the structure of this chapter and its relationship with the previous ones, it is recommended to refer to Figure 3.5. This image provides an outline of the sections and key concepts already explained in Chapter 4, which will not be detailed again in this chapter. In particular, aspects such as the baselines, evaluation metrics, and use case details have been previously defined and are assumed in this chapter to avoid unnecessary repetition.

Once more the reader is invited to review the glossaries **Key Concepts** and **Mathematical Notation** for understanding the definitions and mathematical notation used in the descriptions of this chapter.

5.1 Algorithm Overview

This section describe a novel approach, the Bio-Inspired Volunteer Task Allocation Strategy (Bio-VTAS). The goals of this solution mirror those of the VTAS, specifically Goal 1: Enhance User Perception, Goal 2: Optimised Task Allocation Solution and Goal 3: Create an equitable distribution of tasks across all cells. These goals are explained in detail in Chapter 4, Section 4.1. Bio-VTAS seeks to overcome the limitations of the VTAS respecting the Goal 1 while maintaining the same tasks-completion and task distribution results. Hence, Bio-VTAS has to enhance the user perception and improve user integration into the decision-making process. Like the VTAS, this solution is built upon the Spatial Crowdsourcing model described in Chapter 3.

Since the goals remain unchanged, this strategy continues to function as a recommendation strategy that suggests cells. The tasks associated with each generated cell-recommendation maintain the same states as those in the previous solution. So, this solution can also manage accepted and non-completed tasks to ensure that the campaign’s task completion goal can be met (Goal 2). Similarly, the user interaction loop with the system, explained at Figure 4.1, remains unchanged. The distinguishing feature of this solution lies in how cell-recommendations are generated.

This solution is based on a response threshold module inspired by swarm intelligence (Bonabeau et al., 1997), as observed in social insects like ants and bees. Drawing from the principles of division of labour in these species, Bio-VTAS optimizes task allocation among users and promotes specialization. The algorithm dynamically distributes tasks based on users’ participation patterns. When a user completes multiple tasks in a specific cell, they become specialized in that cell, increasing their likelihood of being assigned future tasks there. Conversely, they are less likely to receive tasks in cells where they lack

specialization. The algorithm learns these spatial participation routines to recommend tasks within a user's established specialization zone, adapting to their preferences. By tailoring task assignments to users' unique behaviours and schedules, this bio-inspired approach ensures efficient task distribution and successful completion across the entire campaign. The idea is for users to follow their routine while receiving task recommendations based on their specialization. Through the interaction process, the recommendations generated by this algorithm will gradually reflect the user's routine, making their participation in the CS campaign easier as it aligns with their daily activities.

All the variables used in the description of the Bio-VTAS solution are summarized in Table 5.1, and key definitions of this thesis are provided in the glossary titled **Key Concepts**.

This specialization of a user u in the cell j is represented by the variable $\theta_{u,j}^t$, which defines the **threshold (an inverse measure of specialization)** of user u in the cell j at time t . This threshold must always fall between a maximum value, θ_{max} , and a minimum value, θ_{min} . A higher threshold of a user in a concrete cell indicates a lack of specialization in this cell, while lower values mean greater specialization. The Bio-VTAS relies on a learning process that adjusts users' thresholds for each cell according to their activity. When a user completes a task in a cell, their threshold for that cell decreases by ξ_0 (Eq. 5.1), and thresholds for neighbourhood cells decrease by ξ_1 (Eq. 5.2). The neighbourhood of a cell j , defined as $n(j)$, consists of all cells whose centres are within a distance of less than $M * C_d$ from the centre of a given cell. Where C_d is the distance between two adjacent cells of the campaign and M is a fixed number that define the radius of the neighbourhood of a cell. Figure 5.1 illustrates an example of the neighbourhood (purple cells) surrounding the orange cell. Additionally, a forgetting rate φ increases the threshold over time if the user does not complete tasks in that cell or a cell of the neighbourhood, allowing for changes in the routine of the user (Eq. 5.3). Furthermore, the threshold variable is not reset at the conclusion of each slot, as it represents a learning process that evolves over time. Hence, the update of thresholds for the user u at time t when this user realised a task in the cell j is calculated by the following equations:

Variable	Definition
j, l, r	A cell of the campaign.
$\theta_{i,j}^t$	Threshold of user i to realize the task of cell j at time t .
θ_{max}	Maximum value that the Threshold can have.
θ_{min}	Minimum value that the Threshold can have.
ξ_0	Learning coefficient associated with the task a user realizes.
ξ_1	Learning coefficient for surrounding tasks related to the realised one.
C_d	Distance between the centres of two adjacent cells.
$n(j)$	Cells in the surroundings of cell j .
C_d	Distance between the centres of two adjacent cells.
M	Variable that define that define the radius of the neighbourhood of a cell.
φ	Forgetting coefficient.
L	Set of Cells of a campaign.
M_s	Maximum number of completed tasks allowed for each cell within each slot.
P_j^t	Priority of the cell j at time t (described at Eq. 4.1).
Pos_i^t	Position of user i at time t .
$d(x, y)$	Distance from location x to location y .
α, β	Positive variables.
m	Number of cells of a campaign.
n	Number of slots of a campaign.
S	Set of slots of a campaign.
S_i	The i -th slot of a campaign.

Table 5.1: Glossary of **Mathematical Notation** used in the Bio-VTAS algorithm description. This table lists variables used in the Bio-VTAS formulation and only the variables that will be used in this chapter are included; for the other variables, see the glossary **Mathematical Notation**.

$$\theta_{u,j}^{t+1} \rightarrow \theta_{u,j}^t - \xi_0 \quad \forall r \in L \quad (5.1)$$

$$\theta_{u,r}^{t+1} \rightarrow \theta_{u,n(j)}^t - \xi_1 \quad \forall r \in n(j) \subset L \quad (5.2)$$

$$\theta_{u,l}^{t+1} \rightarrow \theta_{u,l}^t + \varphi \quad \forall l \in L - \{n(j) \cup j\} \quad (5.3)$$

Where u is the user, j, l and r are cells of the campaign, L the set of cells

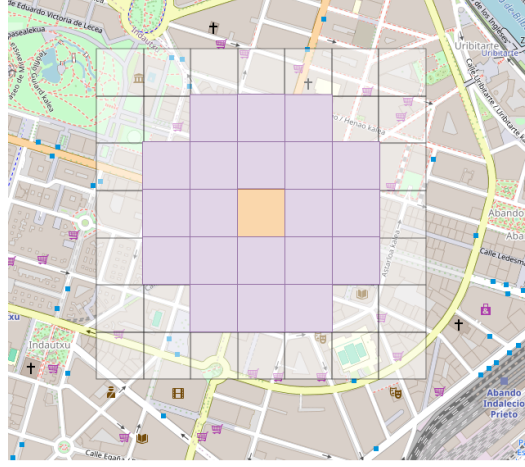


Figure 5.1: An example of the neighbouring cells (purple ones) of the orange cell. The white cells are not neighbours of the orange cell but they are still part of the campaign. In the case of this Figure, M is set at 3 to facilitate the visualization.

of the campaign, $n(j)$ the set of cell of the neighbourhood of cell j , ξ_0 and ξ_1 are positive variables that decrease the threshold in the cell j and in the cells respectively, and always $\xi_0 > \xi_1$. The φ is the forgetting coefficient, and $\theta_{u,j}^t$ is the threshold of user i for cell j at time t .

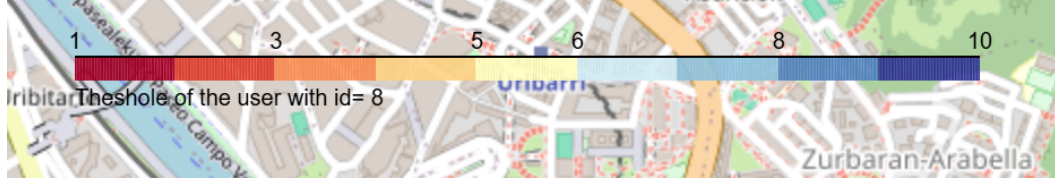


Figure 5.2: The threshold colour scale that represents the user's threshold value for each cell. Based on this colour scale, red indicates a minimum threshold value, signifying high specialization, while blue indicates the maximum threshold value, signifying no specialization.

A new visualization has been created to display the threshold of each cell for a user at a concrete moment of the campaign, as shown in Figure 5.3. This visualization uses a new colour scale, called threshold colour scale (see Figure 5.2), to represent a user's specialization for all cells of the campaign. For this figure, the θ_{max} is set at 10, and the θ_{min} is set at 1. Red indicates a minimum threshold value, signifying high specialization, while blue indicates the max-

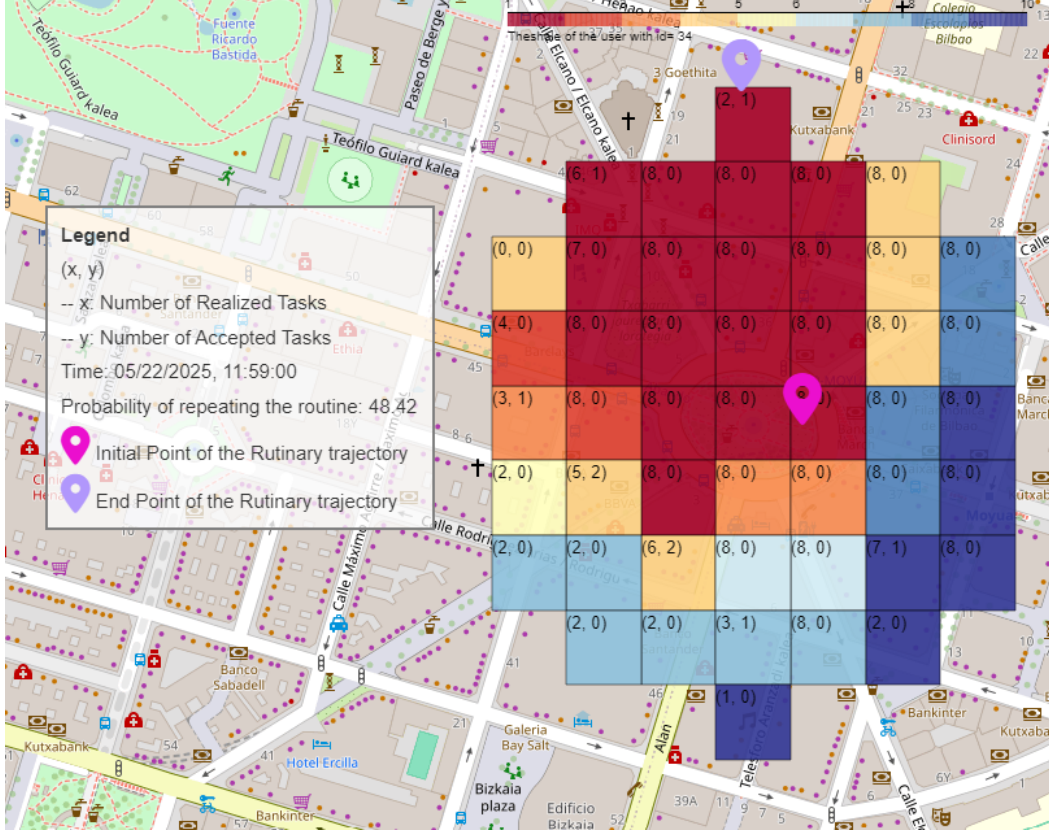


Figure 5.3: Visualization of the specialization of a user in the cells of a campaign. The threshold colour scale represents the user’s threshold value for each cell. In this case, the user has an associated routine and a 48.42% probability of repeating this trajectory. In addition, the θ_{max} is set at 10 and the θ_{min} is set at 1.

imum threshold value, signifying no specialization. Although visualizations can be generated at any moment during the campaign, the most insightful images are those created at the campaign’s conclusion, revealing the users’ final specialization. This final visualization reflects the learning process conducted by Bio-VTAS based on the user behavioural pattern throughout the campaign.

The recommendations of Bio-VTAS are based on user thresholds and also take into account the stimulus generated by each cell. The stimulus quantifies how much a cell requires a task, considering various campaign features, such as the number of completed tasks and the M_s variable. This metric allows the

algorithm to compare the relative necessity of different cells. In this model, the stimulus is defined in the same way as the priority established in Eq. 4.1 in the previous chapter. Reusing this metric ensures an equitable distribution of tasks across cells, a crucial aspect of the VTAS solution's functionality. An automated, scheduled process evaluates the stimulus of all cells at regular intervals (in the case of this work every two minutes), ensuring that the recommender system consistently operates with up-to-date values.

The likelihood of a user completing a task in a concrete cell determines how the algorithm selects cell-recommendations. Individuals engage in task execution when the cell's stimulus exceeds their thresholds. This probability simulate this swarm behaviour, so this probability depends on three factors: the cell's stimulus level, the user's threshold (the expertise level per cell in the campaign), and the distance between the user and the cell. Hence, the probability of an individual u engaging in a task in cell j is defined as in the following equation:

$$\text{Pro}_{u,j}^t = \frac{P_j^{t^2}}{P_j^{t^2} + \alpha * \theta_{u,j}^t + \beta * d(Pos_u^t, j)^2} \quad \forall t; \forall j \in L \quad (5.4)$$

Where α and β are positive variables that determine the influence of the user's threshold and the distance between the user and the cell, respectively; P_j^t is the priority of the cell j at time t ; $\theta_{u,j}^t$ is the threshold of user u for cell j at time t ; and $d(Pos_u^t, j)$ is the distance between the user's position (defined at Pos_u^t) and the cell j . This strategy identifies the cells where the user is most likely to participate with higher probability by applying the Eq. 5.4. Hence, a lower threshold for one cell, that is, a bigger user specialization, means a higher probability for this user to be recommended this cell.

As in the last strategy explained in the previous chapter, this strategy suggests N cells to the user, and this variable is set at 3 for our experimentation to limit the cognitive overload for the user (Bollen et al., 2010; Beierle et al., 2017). So, the three cells whose $\text{Pro}_{u,j}^t$ is greatest than the others are the cell-recommendations for that user. In addition, in this case, the variable M is set at 3.

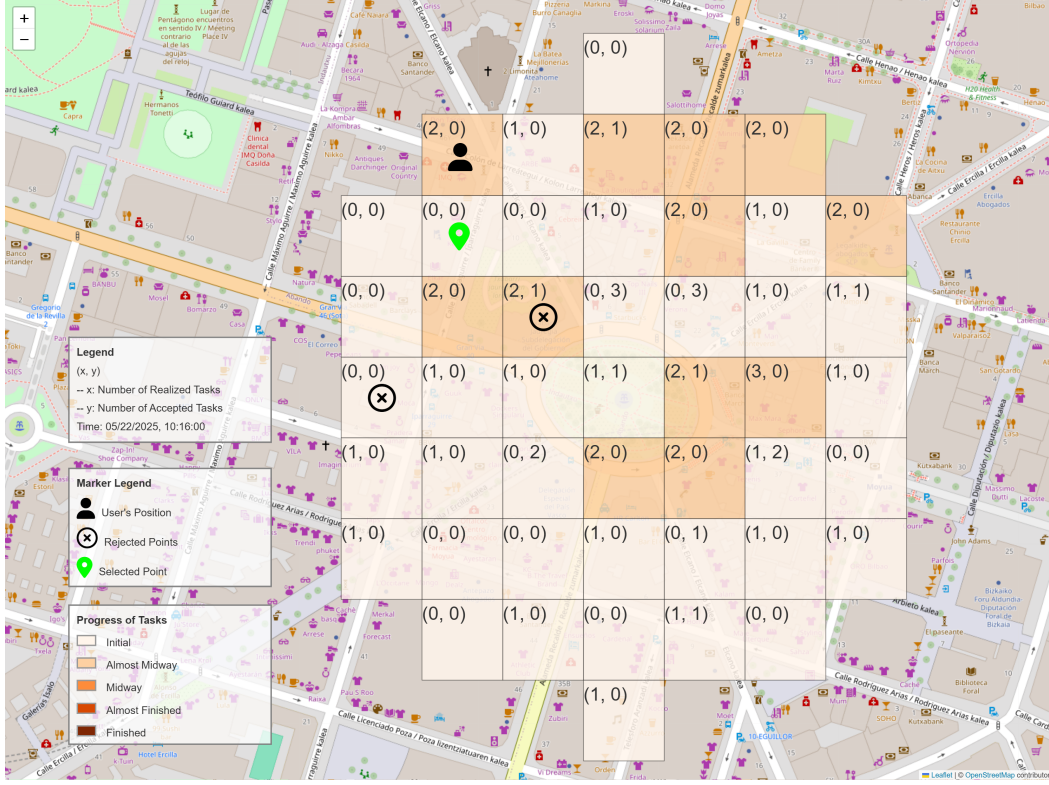


Figure 5.4: An example of a cell-recommendation using the Bio-VTAS algorithm with the completion colour scale. The user, represented by the person icon, is recommended three nearby cells, indicated by the green and cross icons, that have the highest task priority. Each cell has associated with two number in brackets that represent the number of tasks realised and the number of tasks accepted in the cell.

As a consequence of this, over time, the recommender process enables users to specialize in specific areas of the campaign based on their routines. The inclusion of a forgetting factor allows this specialization to evolve, enabling to adapt the recommendations if users' routines change. Thus, the recommendation algorithm tailors its suggestions to the user's routine, prioritizing cells where the user is most likely to participate. This approach balances user personalization with overall task efficiency.

Figure 5.4 shows a cell-recommendation generated for a user using the Bio-VTAS strategy. In this campaign, the M_s parameter is set to 8, and the colour scale is based on this value. Each cell in the image is labelled with a

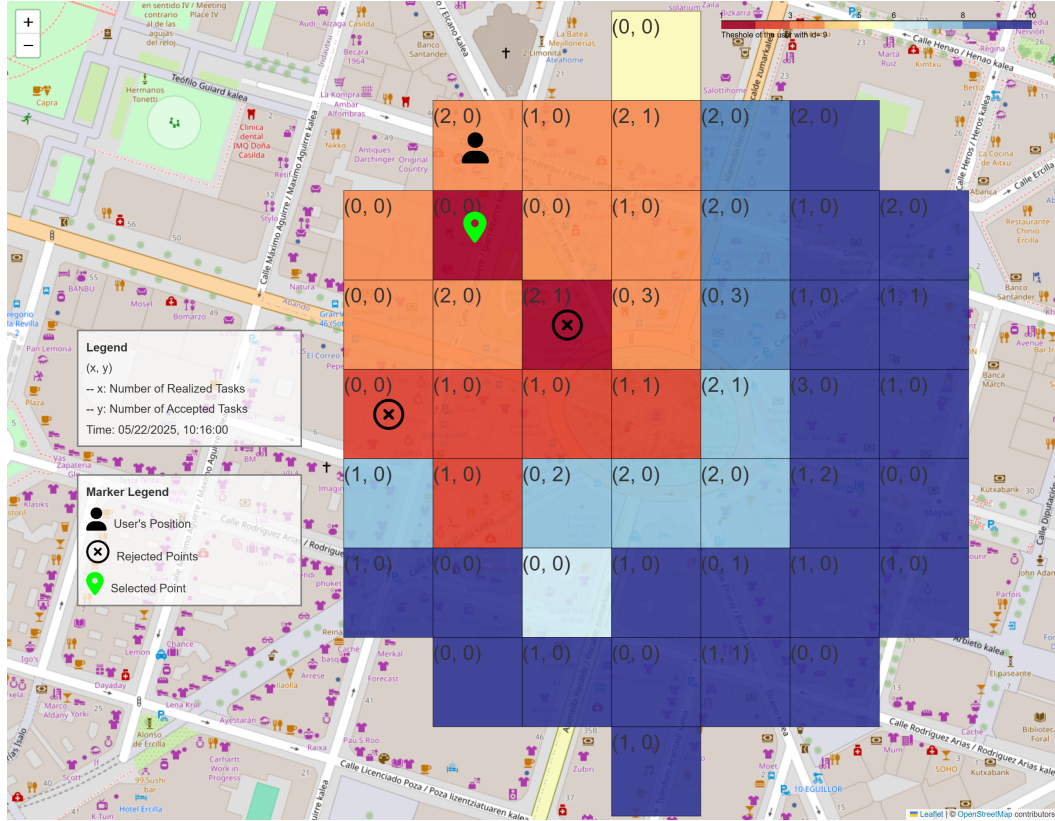


Figure 5.5: An example of a cell-recommendation using the Bio-VTAS algorithm for a user with the threshold colour scale (see Figure 5.2). The user, represented by the person icon, is recommended three cells, indicated by the green and cross icons, that have the highest task priority.

tuple (x, y) , where the first number represents the tasks completed in the cell, and the second represents the tasks accepted in the cell. The user, represented by the person icon, is recommended three cells, indicated by the green and cross icons, that have the highest probability. This example can also be seen with the threshold colour scale (see Figure 5.2) in Figure 5.5.

Bio-VTAS strategy for Spatial Crowdsourcing governs the MVE recommender, published at GitHub¹. Such MVE provides a RESTful API which supports the creation of campaigns, based on the Spatial Crowdsourcing model described in Chapter 3. The MVE handles task allocation by efficiently matching available users with tasks via the Bio-VTAS, regardless of task type or the

¹<https://github.com/mpuerta004/RecommenderSystem>

specific Citizen Science initiative. This ensures broad adaptability and inclusivity in task assignments.

5.2 Simulation Use Case

In this section, we present the simulation use case. The simulation models the task allocation problem based on the Spatial Crowdsourcing model defined in Chapter 3. The goal of the study is to assess the behaviour and performance of the Bio-VTAS under various conditions. The simulation allows for the observation of which behavioural variables of the simulated users have a significant impact on achieving the task-completion goal of the campaign (Goal 2) and highlights the differences between the Bio-VTAS and the baseline models used (VTAS, LNN and NN). Additionally, it primarily provides data on how effectively the Bio-VTAS distributes tasks (Goal 3). Furthermore, in cases where the task-completion goal cannot be fully achieved, it is essential and necessary for the algorithm to generate an equitable distribution of tasks, thereby preserving the usability and validity of the collected data. This aspect is also assessed in the simulation.

This simulation models the task allocation problem to evaluate the effectiveness of the Bio-VTAS solution. For that, it replicates user behaviour in the context of altruistic participation, where users adhere to their daily routines. The system accounts for a range of behaviours, from predictable patterns to completely random activities. This simulation setup is identical to the one used in Chapter 4, and it is explained in Section 4.3.

The experiments to evaluate the effectiveness of the Bio-VTAS were based on the same three customized variables described in Section 4.3 (Chapter 4): 1) *Number of Users*; 2) *User Availability Probability* and 3) *User Task Completion Probability*. The campaign configuration used in all experiments realised in this section matches that described in Chapter 4, Section 4.3. This configuration included two slots, each lasting 2 hours. The campaign was located in central Bilbao, at a longitude of -2.9349889768009008 and latitude of 43.26302147532318, with a radius of 0.2km. The campaign began at 08:00:00 and ended on the same day at 12:00:00. The C_d was 0.05km, and M_s was eight

tasks. The threshold for each cell with a user, $\theta_{u,j}^t$, ranges from a maximum of 10 to a minimum of 1. A visualization of this campaign is shown in Figure 5.6.

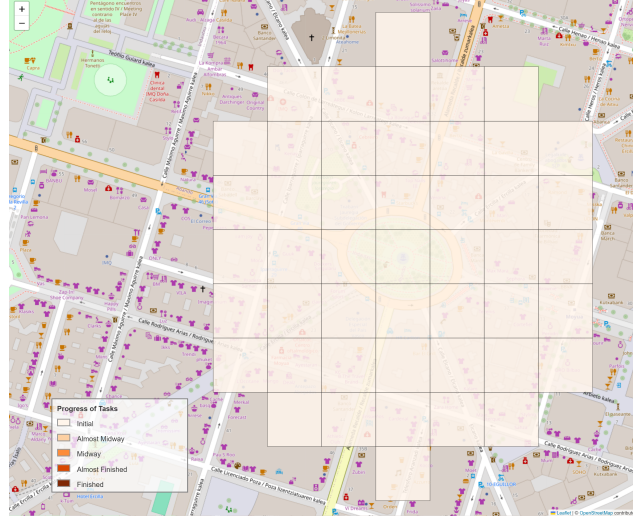


Figure 5.6: Visualization of the campaign configuration used to evaluate the effectiveness of the Bio-VTAS solution in this use case.

Similarly, the values of the variables for the Bio-VTAS algorithm have been defined and established for all the experiments to be conducted in this use case. However, in the Citizen Science campaigns that are carried out, these variables should be adjusted to meet the specific needs of each campaign. For example, if the forgetting factor needs to be higher in scenarios with a large number of users, this knowledge could be applied to improve the efficiency of the algorithm in such campaigns. Nevertheless, since defining these values can be complex for campaign designers, and it is not possible to predict with certainty how users will respond in a particular context, it has been decided to set these values for all algorithms. This approach allows the evaluation of the effectiveness of the Bio-VTAS algorithm even if it does not have a fully optimized configuration. Hence, the θ_{max} is set at 10.0, θ_{min} is set at 1.0, the φ is set at 0.005, α, β are set at 2.0, ξ_0 is set at 2.0 and ξ_1 is set at 1.0 for all the experiments realised in this use case.

The remainder of this section is organised as follows: Subsection 5.2.1 presents the first experiment, where the Bio-VTAS is evaluated with differ-

ent numbers of users to assess its effectiveness. Subsection 5.2.2 presents the second experiment, in which different User Availability Probabilities are tested to evaluate the Bio-VTAS’s performance. Subsection 5.2.3 presents the third experiment, where different User Task Completion Probabilities are tested to further evaluate the Bio-VTAS. Finally, Subsection 5.2.4 presents the conclusions of the simulation use case.

5.2.1 Experiment 1: Crowd Size Variation

This experiment evaluates Bio-VTAS across different crowd sizes—100, 75, 50, and 25 users—ranging from low to high individual effort. The analysis examines how Bio-VTAS manages task completion under varying participation levels (Goal 2), including scenarios where full completion is unattainable. In such cases, ensuring an equitable task distribution becomes essential (Goal 3). The campaign settings and task objectives remained consistent with those in the previous chapter. All campaigns assumed a 75% User Task Completion Probability, reflecting high commitment, and a 50% User Availability Probability. Bio-VTAS was compared against VTAS, LNN, and NN baselines, with each algorithm tested five times per scenario, averaging the results for analysis.

		TCR			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	100%	100%	94.52%	50.48%
	LNN	100%	98.56%	74.95%	40.59%
	NN	160.96%	121.22%	83.40%	39.89%
	Bio-VTAS	100%	99.9%	85%	44.47%
Slot 2	VTAS	100%	100%	95.43%	49.15%
	LNN	100%	98.67%	73.46%	39.521%
	NN	157.13%	122.02%	82.29%	41.12%
	Bio-VTAS	100%	99.79%	81.38%	46.01%

Table 5.2: TCR evaluation over the different number of users (100, 75, 50, and 25) using the VTAS, Bio-VTAS, LNN, and NN task allocation algorithms. The results are the mean of five executions. The bold values are the high results obtained in each slot for each user count.

The Task Completion Rate (TCR) metric is presented in Table 5.2. The results are consistent with those obtained in the previous chapter, as Bio-VTAS does not outperform VTAS when the workload limit is not imposed (25 and 50 users), nor does it surpass NN when this limit is in place (100 and 75 users). Additionally, when the workload limitation is not reached, Bio-VTAS achieves higher values than NN in almost all cases. Furthermore, since Bio-VTAS adheres to the workload limitation, just as VTAS does, we can conclude that Bio-VTAS outperforms both NN and LNN. Moreover, we can also infer that the TCR of Bio-VTAS could surpass NN if it were not constrained by the workload limitation. However, although VTAS still outperforms Bio-VTAS, the latter achieves the second-best results and surpasses both LNN and NN.

The right-hand side of Figure 5.7 illustrates the evolution of Realised Tasks (Eq. 4.3) per minute for the first slot of the tested campaign using the Bio-VTAS, VTAS, LNN, and NN algorithms. A moving average¹ of 20 values was applied to highlight the trend in the evolution of realised tasks. The results represent the mean of five executions. The graph demonstrates that the VTAS consistently achieves a higher number of realised tasks compared to Bio-VTAS, LNN, and NN. This superiority is maintained throughout most of the slot, except toward the end in scenarios with 100 and 75 users, where the VTAS graph shows a sharp decline. This drop occurs because the task-completion objectives have been fully met or are nearly achieved, causing users to stop receiving cell-recommendations as a mechanism to prevent overload. Consequently, the number of realised tasks decreases due to the enforcement of the workload limitation. A similar decline is observed with the Bio-VTAS and LNN algorithms in the same scenarios (100 and 75 users) for the same reason. Table 5.2 confirms this behaviour, as the Task Completion Rate (TCR) for VTAS, Bio-VTAS, and LNN in these scenarios reaches or approaches 100%. The graph further shows that the Bio-VTAS consistently performs better than

¹A Moving Average (MA) is a simple way to analyse data trends over time by smoothing out short-term fluctuations. It works by calculating the average of a set of values over a specific period and then continuously updating this average as new data points are added. For example, in a 7-day moving average, you take the average of the last 7 days of data. The next day, you drop the oldest value and add the newest one, recalculating the average. This helps reveal overall trends by reducing the impact of daily ups and downs.

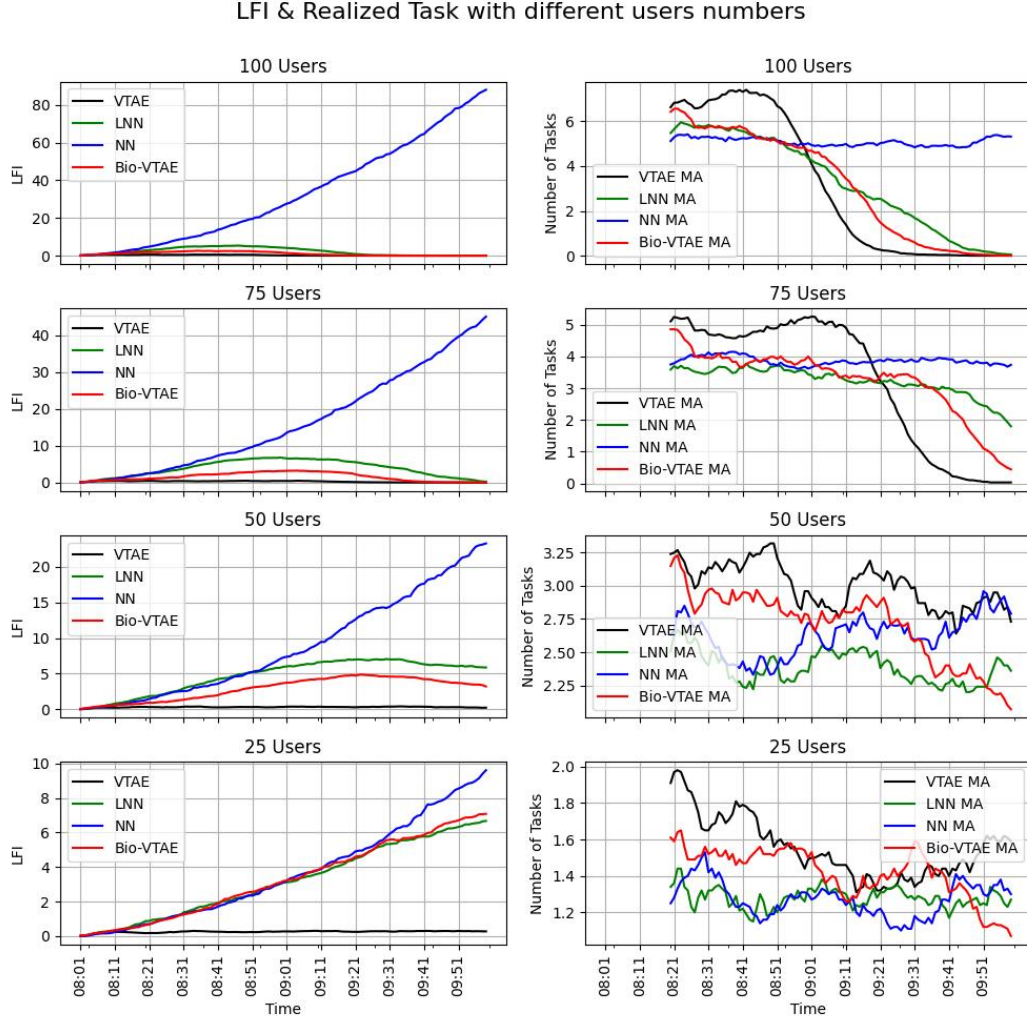


Figure 5.7: The LFI metric (Eq. 4.4 left-hand side pictures) and the realised task per minute (Eq. 4.3 right hand-side figures) over the first slot of the campaign across different users count (100, 75, 50, and 25 Users). The task allocation algorithms used are the VTAS, Bio-VTAS, LNN, and NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results presented are the mean of five executions.

LNN but does not surpass VTAS, with its curve positioned between those of VTAS and LNN.

Table 5.3 presents the $\text{MEAN} \pm \text{SD}$ of the realised tasks (RT) metric. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count, supporting the previously mentioned findings.

VTAS outperforms the NN algorithm in the number of realised tasks only in the cases of 25 and 50 users, meaning when the workload limitation is not reached. In the other two tested cases, NN outperforms all other algorithms in this metric, exhibiting the same behaviour observed in the TCR metric. These results are consistent with the results obtained in the previous chapter. Regarding the Bio-VTAS, this algorithm consistently performs better than the LNN algorithm, being the second-best performer in almost all cases. However, the Bio-VTAS does not surpass the VTAS algorithm. Therefore, Bio-VTAS achieves the second-best performance in this metric.

		MEAN $\pm$ SD (RT)			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	3.16 $\pm$ 3.38	3.16 $\pm$ 2.42	2.99$\pm$0.81	1.59$\pm$0.00
	LNN	3.16 $\pm$ 2.40	3.11 $\pm$ 1.08	2.37 $\pm$ 0.70	1.28 $\pm$ 0.56
	NN	5.06$\pm$0.99	3.83$\pm$0.77	2.64 $\pm$ 0.77	1.26 $\pm$ 0.48
	Bio-VTAS	3.16 $\pm$ 2.83	3.16 $\pm$ 1.71	2.69 $\pm$ 0.84	1.41 $\pm$ 0.51
Slot 2	VTAS	3.13 $\pm$ 3.27	3.13 $\pm$ 2.33	2.99$\pm$0.89	1.54$\pm$0.58
	LNN	3.13 $\pm$ 2.44	3.09 $\pm$ 1.11	2.30 $\pm$ 0.73	1.24 $\pm$ 0.47
	NN	4.92$\pm$1.20	3.82$\pm$0.96	2.58 $\pm$ 0.73	1.29 $\pm$ 0.57
	Bio-VTAS	3.13 $\pm$ 2.46	3.13 $\pm$ 1.58	2.55 $\pm$ 0.84	1.44 $\pm$ 0.56

Table 5.3: The MEAN $\pm$ SD of the Realised tasks (RT - Eq. 4.3) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count.

The left-hand side of Figure 5.7 illustrates the evolution of LFI (Load Fairness Index - Eq. 4.4) per minute for the first slot of the tested campaigns using the Bio-VTAS, VTAS, LNN, and NN algorithms. The results represent the mean of five executions. In this picture only the data from the first slot is presented. The first observation is the exponential behaviour of the NN algorithm for all the campaigns. As observed in the previous chapter, the LFI of the VTAS remains minimal. The LFI of Bio-VTAS exhibits similar behaviour, with values consistently between those of LNN and VTAS. This indicates that VTAS maintains the lowest LFI, followed by Bio-VTAS and

then LNN. For the scenario with 25 users, the LFI of LNN and Bio-VTAS display similar patterns. This outcome is expected for Bio-VTAS, as the algorithm regularises the task distribution as the target number of completed tasks is reached. Therefore, Bio-VTAS is only recommended for campaigns with a sufficient number of participants to meet the required task-completion goal. Under these conditions, the LFI is minimal, making Bio-VTAS and LNN effective choices for ensuring equitable task distribution.

		MEAN $\pm$ SD (LFI)			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	0.22$\pm$0.22	0.24$\pm$0.17	0.32$\pm$ 0.06	0.26$\pm$0.05
	LNN	2.23 $\pm$ 1.99	3.83 $\pm$ 2.21	4.70 $\pm$ 2.27	3.20 $\pm$ 2.10
	NN	32.55 $\pm$ 26.57	16.23 $\pm$ 13.32	8.75 $\pm$ 7.05	3.64 $\pm$ 2.81
	Bio-VTAS	1.04 $\pm$ 0.95	1.52 $\pm$ 1.12	2.90 $\pm$ 1.55	3.29 $\pm$ 2.23
Slot 2	VTAS	0.22$\pm$0.23	0.25$\pm$0.18	0.36$\pm$0.09	0.28$\pm$0.07
	LNN	1.96 $\pm$ 1.87	3.54 $\pm$ 2.22	4.35 $\pm$ 2.22	3.08 $\pm$ 2.04
	NN	28.91 $\pm$ 25.09	15.12 $\pm$ 12.82	8.99 $\pm$ 7.40	3.64 $\pm$ 3.01
	Bio-VTAS	1.80 $\pm$ 1.71	2.74 $\pm$ 2.04	4.50 $\pm$ 2.50	4.13 $\pm$ 2.93

Table 5.4: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each user count.

The MEAN $\pm$ SD of this metric is shown in Table 5.4. The highlighted results indicate the algorithms with the lowest mean for each user count and slot. These findings reinforce previous observations: the NN algorithm consistently exhibits significantly higher mean LFI values due to its exponential behaviour, whereas the VTAS algorithm maintains lower LFI values across all cases. Additionally, the Bio-VTAS algorithm demonstrates a notably higher LFI compared to VTAS. A key result in this table is the increase in LFI of the Bio-VTAS in the second slot. This could be attributed to users becoming more specialized, which results in their areas of operation becoming more defined. If there are campaign cells outside the users' specialized zones, these cells will be less likely to be recommended, leading to a less equitable task

distribution. Another possibility there are more users specialised in one cell than in others, allowing this cells to reach the M_s task threshold earlier. This creates an inequitable distribution of tasks among the cells. As long as the workload limit is not reached, those cells will remain unassigned. Another possible explanation is that the Bio-VTAS requires more time to learn users' movement patterns. Both explanations align with the expected behaviour of the algorithm. However, it is important to note that this increase in LFI is not observed in the $\text{MEAN} \pm \text{SD}$ of the RT metric (Table 5.3), where the Bio-VTAS shows consistent behaviour across both slots, achieving a similar number of tasks in each. This further supports the idea that user specialisation may be driving the LFI increase in the second slot. These results confirm that the Bio-VTAS is a suitable choice when it is known that the campaign will meet its task completion objectives.

Figure 5.8 illustrates the distribution of tasks among the cells at the end of the first slot of the campaign for Bio-VTAS with 25 users. The results show the number of realised tasks and accepted tasks for each cell, and with the completion colour scale developed, the distribution of the task can be shown. A large dispersion of colours can be observed, indicating that the Bio-VTAS does not exhibit an equitable distribution of tasks. While this chapter presents the representation for 25 users, the corresponding images for the other scenarios in this experiment can be found in Appendix E, Section E.1.

The results of Experiment 1 highlight the robustness of the VTAS algorithm in managing varying crowd sizes. The Bio-VTAS consistently achieves a higher TCR compared to the LNN and NN algorithms when workload limitations are not exceeded. Furthermore, the Bio-VTAS outperforms the LNN and NN algorithms in the total number of realised tasks, except in scenarios involving 100 and 75 users where the workload limitation is reached. These metrics collectively demonstrate that, if not for the workload limitation, the Bio-VTAS would also surpass the NN in the remaining cases. However, the Bio-VTAS strictly adheres to the imposed workload limitation, ensuring it operates within a defined task threshold. Consequently, while it does not exceed the TCR or the total number of tasks performed in all scenarios, its

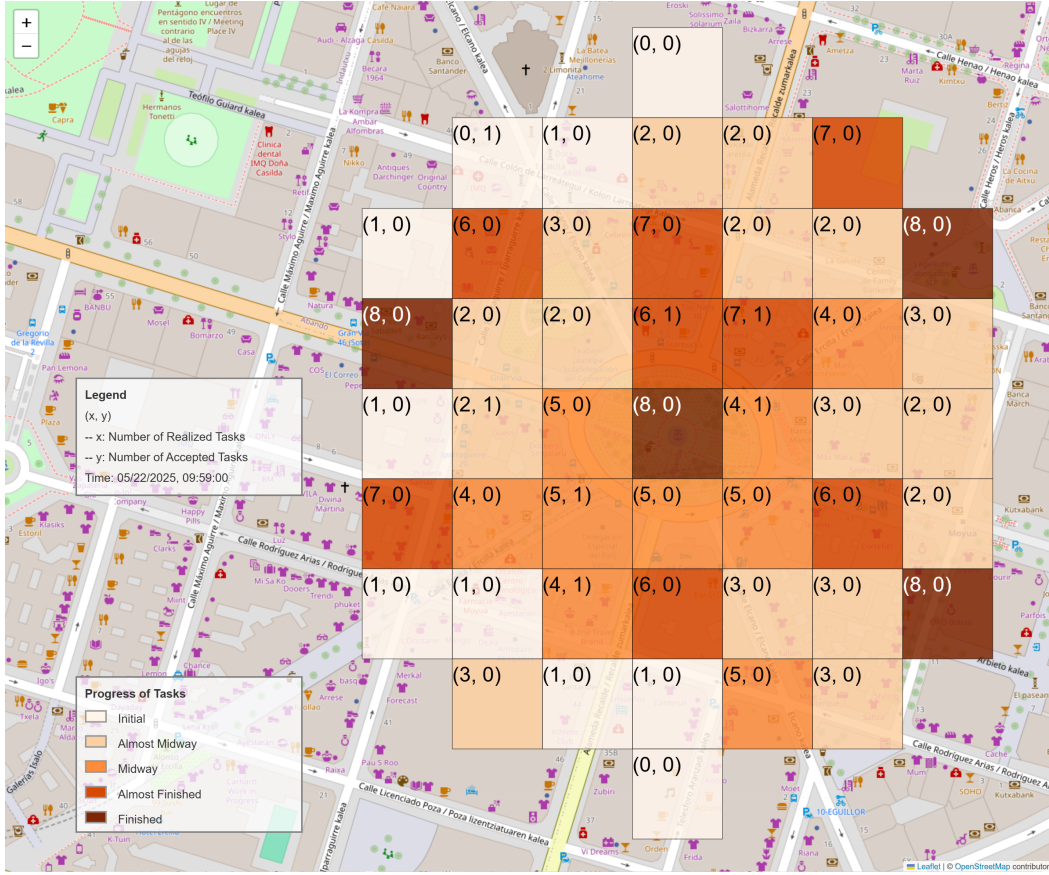


Figure 5.8: The task completion campaign involved 25 users using the Bio-VTAS algorithm. This image illustrates the distribution of tasks among the cells at the end of the first slot of the campaign.

overall performance remains superior to that of the NN. This outcome aligns with the Citizen Science context of this thesis, where preventing the overburdening of users with excessive tasks is a critical requirement that algorithms must fulfil. In conclusion, the Bio-VTAS algorithm proves to be more suitable for this context than the LNN and NN, achieving better results than the baseline algorithms in fulfilling the Goal 2, but not outperforming the VTAS results. Regarding Goal 3, Bio-VTAS is more unstable, and its performance is similar to, though slightly better than, the baseline algorithms LNN and NN. In contrast, VTAS maintains a low Load Fairness Index (LFI) across all tested scenarios, indicating an equitable task distribution even under varying levels of user effort, thereby fulfilling Goal 3. To conclude, while the

Bio-VTAS algorithm is suitable for this context and achieves better results than the baseline algorithms (LNN and NN) in addressing Goal 3, it does not outperform VTAS.

As Bio-VTAS is a more user-centric solution, it was expected that the optimisation results for the campaign requirements (Goal 1 and Goal 3) would slightly decline. Therefore, these results were anticipated. The key point is that, although this solution may not be the best among the proposed options, it still meets the goals in relation to the defined baselines (LNN and NN). Consequently, for the tested goals (Goal 2 and Goal 3), this solution continues to support our hypothesis while also being more user-centric and better integrating users, as intended by this new strategy.

5.2.2 Experiment 2: User Availability Probability

This experiment evaluates Bio-VTAS under different User Availability Probabilities—75% (high motivation and participation), 50%, and 25% (low motivation and participation). It examines how well Bio-VTAS supports task completion at varying participation levels (Goal 2), including cases where achieving the objective may not be possible. In such scenarios, ensuring an equitable distribution of tasks becomes essential (Goal 3). The campaign configuration and task completion objectives remained consistent with those outlined in the previous section. For all scenarios, the User Task Completion Probability was set at 75%, ensuring a high level of commitment, while the total number of users was fixed at 50. Bio-VTAS was compared against VTAS, LNN, and NN baselines, with each scenario being run five times per algorithm, and the final results representing the average values from these simulations.

The Task Completion Ratio (TCR) metric is presented in Table 5.5. As observed in the previous experiment with this metric, the highlighted results remain unchanged with the inclusion of Bio-VTAS. This is because Bio-VTAS does not improve upon the results obtained by VTAS when the workload limitation is not imposed (50% and 25%) nor those achieved by NN when the workload limitation is in place (75%). However, Bio-VTAS is the second-best

		TCR		
		User available rate		
		75%	50%	25%
Slot 1	VTAS	100%	94.52%	70.27%
	LNN	91.7%	74.95%	48.62%
	NN	101.9%	83.40%	49.04%
	Bio-VTAS	97.50%	85%	59.15%
Slot 2	VTAS	100%	95.43%	61.38%
	LNN	90.7%	73.46%	48.62%
	NN	102.2%	82.29%	46.33%
	Bio-VTAS	96.81%	81.38%	58.46%

Table 5.5: The TCR performed across different User Availability Probabilities using the VTAS, Bio-VTAS, LNN, and NN task allocation algorithms. The reported results represent the mean of five executions. The highest values indicate the best outcomes achieved in each category for the tested User Availability Probability.

algorithm in almost all cases where the workload limitation is not a factor. Another important observation is that Bio-VTAS respects workload limitations. Therefore, as seen in the previous chapter with VTAS, we can conclude that Bio-VTAS outperforms LNN and NN in terms of TCR effectiveness. Nevertheless, it does not surpass VTAS in any case.

The right-hand side of Figure 5.9 illustrates the Realised Tasks during the first slot of the campaign using the Bio-VTAS, VTAS, LNN, and NN algorithms. This figure shows that VTAS consistently achieves a higher number of realised tasks compared to Bio-VTAS, NN, and LNN. The only exception occurs towards the end of the 75% case, where the number of realised tasks for VTAS declines due to the workload limitation imposed by the campaign. A similar decline is also observed for Bio-VTAS in the 75% cases for the same reason. Additionally, in all other tested cases, Bio-VTAS outperforms the NN and LNN strategies. These results remained consistent in the last experiment, where Bio-VTAS ranked as the second-best algorithm for this metric, consistently outperforming NN and LNN for the same reasons observed in previous experiments.

Table 5.6 shows the MEAN $\pm$ SD of the Realised Tasks for the Bio-VTAS,

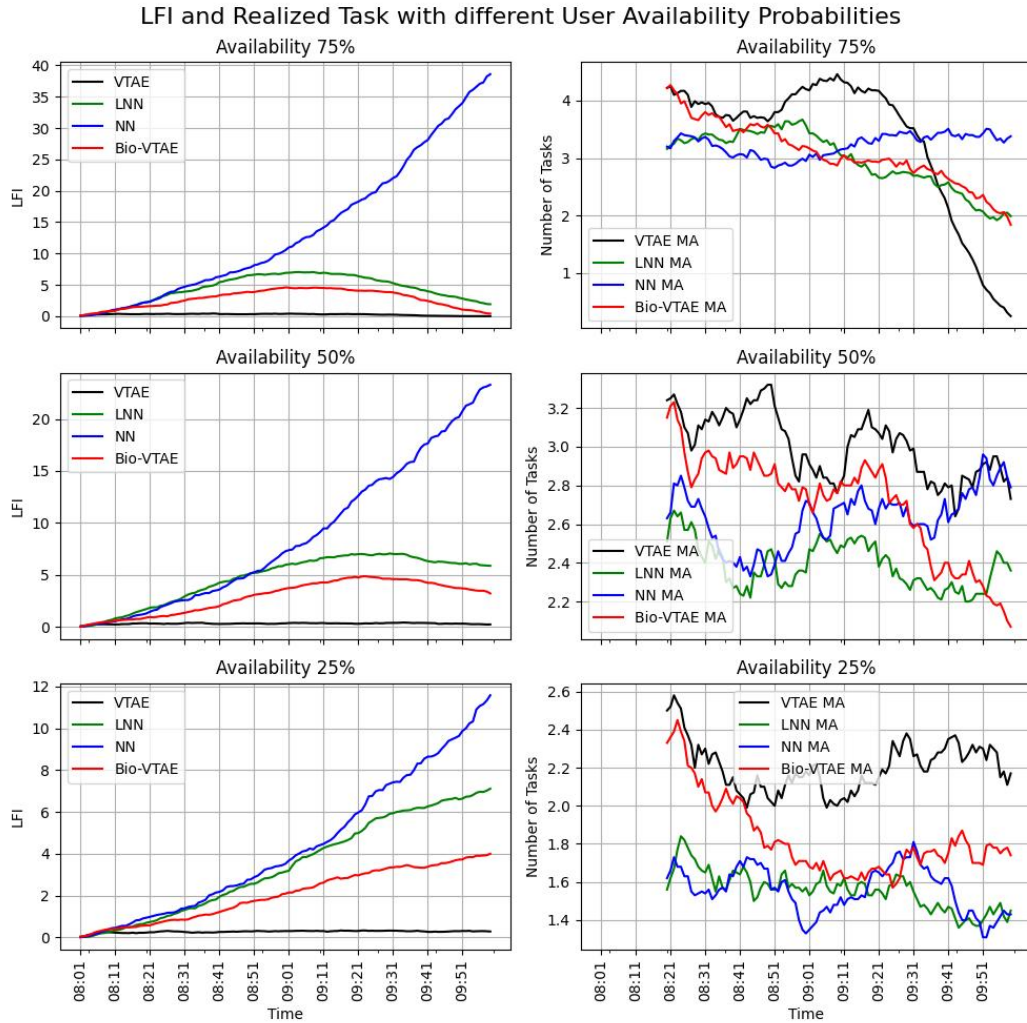


Figure 5.9: The LFI (Eq. 4.4 left figures) and Realised Task (Eq. 4.3 right figures) over the first slot of the campaign across different user availability probabilities. The task allocation strategies used are the VTAS, Bio-VTAS, LNN, and NN algorithms. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results show the mean of five executions.

VTAS, LNN, and NN algorithms across different User Availability Probabilities. The highlighted results correspond to the algorithms with the highest mean in each slot for each tested User Availability Probability. These values support what was previously mentioned. Given the decline experienced by the VTAS and Bio-VTAS with 75% of availability, the NN outperforms

them in this metric. However, with the other tested probabilities, the VTAS significantly outperforms the NN, the Bio-VTAS, and the LNN; and the Bio-VTAS outperform the NN and LNN strategies. As mentioned in the previous paragraphs, this was also observed in the TCR metric. The Bio-VTAS consistently outperforms the LNN and NN due to the Citizen Science domain of this thesis, while the VTAS surpasses the Bio-VTAS in every case. The behaviour of the Bio-VTAS is similar to that of the VTAS, but with a slightly lower performance, since its decline with 75% of availability is less pronounced.

		MEAN $\pm$ SD (RT)		
		User available rate		
		75%	50%	25%
Slot 1	VTAS	3.16 $\pm$ 1.71	2.99$\pm$0.81	2.22$\pm$0.81
	LNN	2.90 $\pm$ 0.91	2.37 $\pm$ 0.70	1.54 $\pm$ 0.57
	NN	3.22$\pm$0.83	2.64 $\pm$ 0.77	1.55 $\pm$ 0.56
	Bio-VTAS	3.08 $\pm$ 1.16	2.69 $\pm$ 0.84	1.87 $\pm$ 0.70
Slot 2	VTAS	3.13 $\pm$ 1.78	2.99$\pm$0.89	1.92$\pm$0.67
	LNN	2.84 $\pm$ 0.96	2.30 $\pm$ 0.73	1.52 $\pm$ 0.52
	NN	3.20$\pm$0.85	2.58 $\pm$ 0.73	1.45 $\pm$ 0.51
	Bio-VTAS	3.03 $\pm$ 1.07	2.55 $\pm$ 0.84	1.83 $\pm$ 0.65

Table 5.6: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user availability probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability.

The left-hand side of Figure 5.9 shows the LFI metric for the first slot of the campaign, using the Bio-VTAS, VTAS, LNN, and NN algorithms. One key observation, consistent with the previous experiment, is that the LFI metric of the NN algorithm exhibits an exponential behaviour. In the case of a 25% probability, both the LNN and Bio-VTAS also show exponential behaviours. This occurs because the LNN and Bio-VTAS achieve a more equitable distribution (resulting in low LFI) when they are close to the workload limitation. However, with 25% availability, this limit is still far from being reached. An example of this can be seen with 75% availability, where the Bio-VTAS and LNN algorithms approach the workload limitation near the end of the slot,

causing the LFI to decrease. In contrast, the VTAS algorithm consistently maintains a low LFI throughout the campaign. The Bio-VTAS was the second best-performing algorithm in this metric, consistently outperforming the NN and LNN strategies.

		MEAN $\pm$ SD (LFI)		
		User available rate		
		75%	50%	25%
Slot 1	VTAS	0.26$\pm$0.13	0.32$\pm$0.06	0.28$\pm$0.05
	LNN	4.38 $\pm$ 2.13	4.70 $\pm$ 2.27	3.44 $\pm$ 2.33
	NN	13.80 $\pm$ 11.34	8.75 $\pm$ 7.05	4.37 $\pm$ 3.34
	Bio-VTAS	2.77 $\pm$ 1.41	2.90 $\pm$ 1.55	2.07 $\pm$ 1.23
Slot 2	VTAS	0.26$\pm$0.14	0.36$\pm$0.09	0.33$\pm$0.07
	LNN	4.51 $\pm$ 2.25	4.35 $\pm$ 2.22	3.33 $\pm$ 2.24
	NN	12.31 $\pm$ 10.25	8.99 $\pm$ 7.40	3.88 $\pm$ 3.11
	Bio-VTAS	4.27 $\pm$ 2.34	4.50 $\pm$ 2.50	3.84 $\pm$ 2.24

Table 5.7: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different user availability probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each tested User Availability Probability.

Table 5.7 presents the MEAN $\pm$ SD of the LFI metric for the Bio-VTAS, VTAS, LNN, and NN algorithms under different User Availability Probabilities. The highlighted results correspond to the algorithms with the lowest mean in each slot for each User Availability Probability. These values support what was previously mentioned. Given the exponential behaviour of the NN, the mean values of this metric are considerably high. However, the VTAS maintains low LFI values in all cases, indicating that the task distribution across the cells is equitable. The Bio-VTAS algorithm is the second with the lowest MEAN $\pm$ SD. Indeed, as in the previous experiment (Subsection 5.2.1), its values increase significantly in the second slot. As explained earlier, this could be due to users having already specialised; therefore, if several users focus their routines on the same cell, the algorithm fails to achieve an equitable distribution of tasks. The LFI metric is only regulated and decreases with this algorithm when the campaign is close to the tasks-completion goal.

In summary, this experiment has evaluated the performance of four task allocation algorithms: VTAS, Bio-VTAS, LNN, and NN, under different User Availability Probabilities (75%, 50%, and 25%). Bio-VTAS exhibits behaviour that is very similar to that of VTAS. However, Bio-VTAS demonstrates less efficient performance across all metrics than the VTAS. Despite this, its performance consistently surpasses that of LNN and NN related to Goal 2 and Goal 3. As mentioned earlier, this demonstrates the efficiency of Bio-VTAS. By outperforming the NN and LNN baselines, it confirms that, in relation to the examined objectives (Goal 2 and Goal 3), Bio-VTAS supports the proposed hypothesis. Similarly, it was expected that Bio-VTAS would not achieve the same level of efficiency in these goals as VTAS, given that it is a more user-centric solution. Its aim is to achieve a more balanced trade-off between the campaign’s technical objectives (Goal 2 and Goal 3) and the user perspective (Goal 1). However, the main limitation of Bio-VTAS is that the LFI tends to increase over time. This indicates that users tend to specialise in specific cells, reducing the equitable distribution of tasks. This limitation makes Bio-VTAS suitable only in scenarios where it can be ensured that the campaign will achieve its task-completion goal.

5.2.3 Experiment 3: User Task Completion Probability

This experiment examines Bio-VTAS’s performance across different User Task Completion Probabilities—75% (high commitment), 50%, and 25% (low commitment). It evaluates how well the strategy achieves the task completion objective (Goal 2), even in situations where lower commitment levels may hinder overall success. In such cases, maintaining a equitable task distribution becomes essential, leading to an additional assessment of Bio-VTAS’s ability to ensure equitable allocation (Goal 3). The campaign parameters and task completion objectives remained consistent with those established in the previous section. In all scenarios, the User Availability Probability was set at 50%, representing a moderate level of participation, while the total number of users was fixed at 50. Bio-VTAS was benchmarked against VTAS, LNN, and NN baselines, with each scenario being executed five times per algorithm.

The results presented in this section reflect the average values obtained from these simulations.

		TCR		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	94.52%	66.22%	34.41%
	LNN	74.95%	51.33%	25.05%
	NN	83.40%	49.10%	23.62%
	Bio-VTAS	85.00%	61.6%	27.29%
Slot 2	VTAS	95.43%	62.45%	31.33%
	LNN	73.46%	52.45%	24.63%
	NN	82.29%	50.43%	26.22%
	Bio-VTAS	81.38%	56.60%	26.17%

Table 5.8: The TCR metric conducted across different task completion probabilities using the VTAS, Bio-VTAS, LNN, and NN algorithms. The highest values represent the best results achieved in each category for the tested User Task Completion Probability.

The Task Completion Rate (TCR) metric is provided in Table 5.8. The VTAS algorithm consistently outperforms the Bio-VTAS, LNN and NN algorithms for all the probabilities tested. Additionally, it can be observed that, in this case, the workload limitation is not reached, as the TCR does not reach 100% for any of the probabilities. Similarly, it is evident that the higher the User Task Completion Probability, the better the TCR. The VTAS algorithm is the most efficient strategy in all scenarios, while Bio-VTAS consistently ranks as the second most efficient. The NN and LNN algorithms exhibit comparable performance, remaining close to each other across all scenarios.

The right-hand side of Figure 5.10 shows the Realised Tasks (Eq. 4.3) through the first slot of the campaign with the Bio-VTAS, VTAS, LNN and NN algorithms. This figure shows that the VTAS consistently surpasses the number of realised tasks for all the algorithms. The Bio-VTAS algorithm is the second most efficient in all scenarios.

Table 5.9 shows the MEAN $\pm$ SD of the realised tasks for the Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probabilities. The highlighted results correspond to the algorithms with the highest

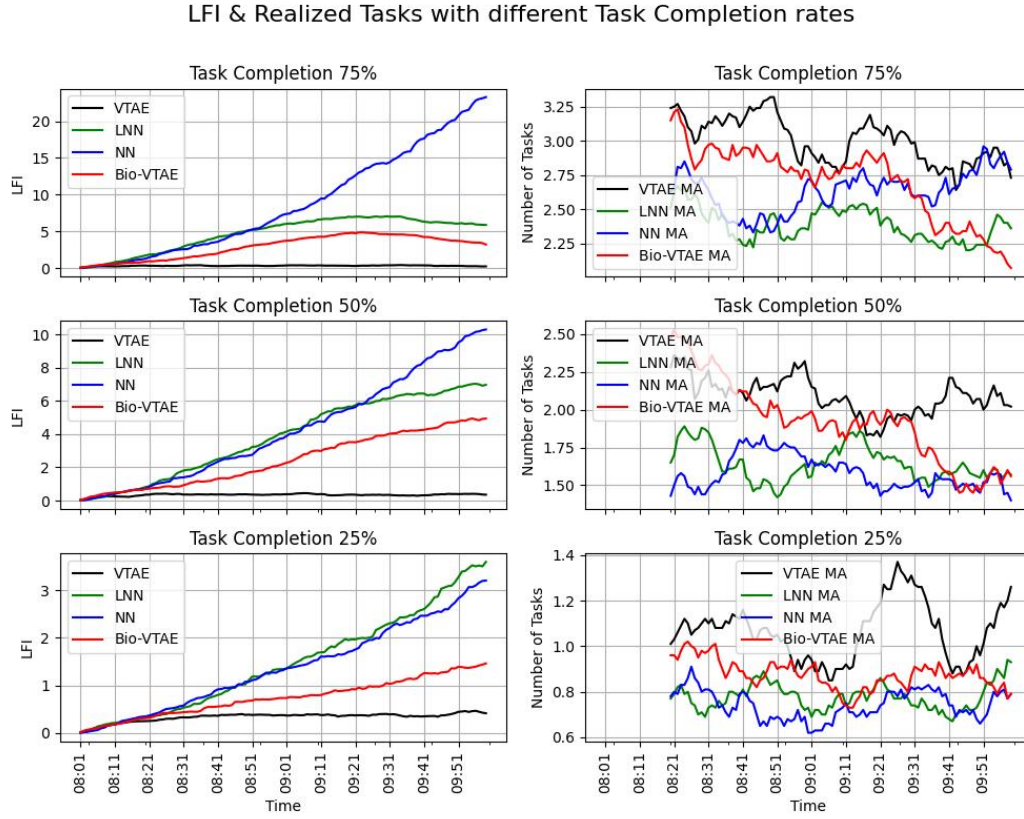


Figure 5.10: The LFI metric (Eq. 4.4 - left-hand figures) and the Realised Tasks (Eq. 4.3 - right-hand figures) over the first slot of the campaign across different User Task Completion Probabilities. The task allocation solutions used are the VTAS, Bio-VTAS, the LNN, and the NN algorithms. To illustrate the evolution of the Realised Task, the moving average of 20 values was used. The results show the mean of five executions.

mean in each slot for each User Task Completion Probability. These values support what was previously mentioned. The VTAS outperforms the NN, Bio-VTAS and LNN algorithms in the number of realised tasks for all the probabilities tested. The Bio-VTAS algorithm is the second most efficient in all scenarios.

The left-hand side of Figure 5.10 illustrates the LFI metric (Eq. 4.4) throughout the first slot of the campaign for various User Task Completion Probabilities tested with the Bio-VTAS, VTAS, LNN, and NN algorithms. In this picture only the data from the first slot is presented. The VTAS algorithm

		MEAN $\pm$ SD (RT)		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	2.99$\pm$0.81	2.09$\pm$0.75	1.09$\pm$0.48
	LNN	2.37 $\pm$ 0.70	1.62 $\pm$ 0.56	0.79 $\pm$ 0.37
	NN	2.64 $\pm$ 0.77	1.55 $\pm$ 0.57	0.75 $\pm$ 0.39
	Bio-VTAS	2.69 $\pm$ 0.84	1.95 $\pm$ 0.73	0.88 $\pm$ 0.42
Slot 2	VTAS	2.99$\pm$0.89	1.96$\pm$0.69	0.98$\pm$0.50
	LNN	2.30 $\pm$ 0.73	1.64 $\pm$ 0.58	0.77 $\pm$ 0.39
	NN	2.58 $\pm$ 0.73	1.58 $\pm$ 0.61	0.82 $\pm$ 0.36
	Bio-VTAS	2.55 $\pm$ 0.84	1.77 $\pm$ 0.64	0.83 $\pm$ 0.40

Table 5.9: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each Task Completion Probability.

consistently maintains a low LFI value throughout the campaign, indicating an equitable task distribution across the cells. In contrast, the NN algorithm exhibits an exponential behaviour in the LFI metric. Similarly, both the LNN and Bio-VTAS algorithms demonstrate exponential behaviour at 25% and 50% probabilities. However, at a 75% probability, this exponential behaviour diminishes as the task completion process approaches the workload limitation, although this limit is not reached within the scope of this experiment.

Table 5.10 presents the MEAN $\pm$ SD of the LFI metric for the Bio-VTAS, VTAS, LNN, and NN algorithms under different User Task Completion Probabilities. The highlighted results correspond to the algorithms with the lowest mean of LFI in each slot for each probability. These values further support the previous observations. Due to the exponential behaviour of the NN algorithm, its mean LFI values are significantly high. In contrast, the VTAS consistently maintains low LFI values across all scenarios, demonstrating a balanced task distribution across cells and efficient task completion. In the second slot, the LNN algorithm shows a slightly lower average than the Bio-VTAS, but the difference is minimal.

In this experiment, the User Task Completion Probability variable was

		MEAN $\pm$ SD (LFI)		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	0.32$\pm$0.06	0.34$\pm$0.07	0.33$\pm$0.09
	LNN	4.70 $\pm$ 2.27	3.83 $\pm$ 2.35	1.48 $\pm$ 1.04
	NN	8.75 $\pm$ 7.05	4.26 $\pm$ 3.15	1.38 $\pm$ 0.91
	Bio-VTAS	2.90 $\pm$ 1.55	2.43 $\pm$ 1.57	0.74 $\pm$ 0.39
Slot 2	VTAS	0.36$\pm$0.09	0.38$\pm$0.10	0.38$\pm$0.12
	LNN	4.35 $\pm$ 2.22	3.65 $\pm$ 2.35	1.36 $\pm$ 0.93
	NN	8.99 $\pm$ 7.40	4.62 $\pm$ 3.51	1.69 $\pm$ 1.20
	Bio-VTAS	4.50 $\pm$ 2.50	3.83 $\pm$ 2.48	1.54 $\pm$ 1.03

Table 5.10: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted values correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability.

tested to understand how user commitment to their accepted tasks affects the effectiveness of the Bio-VTAS. High User Task Completion Probability indicates strong user commitment, leading to more realised tasks. The study evaluated probabilities of 75%, 50%, and 25% using Bio-VTAS, VTAS, LNN, and NN algorithms, with 50 users and a User Availability Probability of 50%. Results averaged over five executions showed that VTAS consistently outperformed the other algorithms in Task Completion Rate (TCR) and realised tasks without reaching workload limitations. Bio-VTAS was the second most efficient, while NN and LNN had comparable performance. The LFI and realised tasks were analysed, showing VTAS's superior performance and equitable task distribution, as indicated by low Load Fairness Index (LFI) values. Bio-VTAS maintained the second-lowest LFI, demonstrating equitable task distribution and efficient task completion across different probabilities. A peculiarity of the Bio-VTAS is that in the second slot, the LFI increases, which indicates that the task distribution is not as equitable as in the VTAS. This is due to the specialisation of users, which leads to a non-equitable distribution of tasks across the cells. Therefore, despite not being the most outstanding performer, it can be concluded that, compared to the baseline algorithms

LNN and NN, Bio-VTAS represents a significant improvement and successfully achieves Goal 2 and Goal 3. As mentioned in previous experiments, it was expected that VTAS would outperform Bio-VTAS in relation to Goal 2 and Goal 3. Moreover, Bio-VTAS performs better than the baseline LNN and NN algorithms, despite the fact that the LFI metric regulates its performance, especially when the workload limitation is enforced. Therefore, although Bio-VTAS could be improved in these metrics, it remains the most user-centric solution among those tested in this thesis.

5.2.4 Conclusions of the evaluation of the Bio-VTAS Algorithm in the Simulation Use Case

The performance of the Bio-VTAS algorithm was evaluated in this use case across various scenarios, including different crowd sizes, User Availability Probabilities, and User task Completion Probabilities. The results indicate that while the Bio-VTAS algorithm consistently performs well, it does not surpass the VTAS algorithm in any of the tested metrics.

In scenarios with varying crowd sizes, the Bio-VTAS algorithm demonstrated a strong performance, particularly when the number of users was sufficient to meet the workload limitations. However, it was observed that the VTAS algorithm consistently achieved higher TCR and realised tasks compared to Bio-VTAS. The Bio-VTAS algorithm outperformed the LNN in all the cases and NN algorithms in cases where workload limitation was not achieved, indicating its robustness in task allocation.

When evaluating User Availability Probabilities, the VTAS algorithm outperformed Bio-VTAS in both TCR and realised tasks, especially when the workload limitation was not fully achieved. The Load Fairness Index (LFI) for Bio-VTAS increased over time, suggesting user specialisation and non-equitable task distribution in the second slot. This behaviour was not observed in the VTAS algorithm, which consistently maintained a low LFI, indicating an equitable task distribution across the cells.

In the context of User Task Completion Probabilities, the Bio-VTAS algorithm maintained its position as the second most efficient algorithm, fol-

lowing VTAS. The VTAS algorithm consistently achieved higher TCR and realised tasks demonstrated a balanced task distribution and efficient task completion across different probabilities. The LFI for Bio-VTAS was higher in the second slot, indicating a less equitable task distribution due to user specialisation.

Overall, the Bio-VTAS algorithm is a robust and reliable choice for task allocation, particularly in scenarios where the workload limitation is met. Although this algorithm does not surpass the performance of VTAS, it does demonstrate the ability to achieve Goal 2 and Goal 3 compared to the baseline algorithms LNN and NN. Being a more user-centric solution, it was expected that Bio-VTAS would see a decline in Goal 2 and Goal 3, which has been verified. Nevertheless, this solution demonstrates better performance in the context of Citizen Science compared to the baseline algorithms LNN and NN in all cases.

5.3 Deusto Use Case

This section evaluates the Bio-VTAS using a real-world scenario involving real users from the University of Deusto¹. In this experiment, participants were tasked with capturing photos of the University of Deusto to create a collage. This use case does not originate from a genuine Citizen Science initiative, as it lacks a community-driven need. Instead, it simulates one in which the goal is to create a pictographic representation of the university's campus and its surroundings. Since both the Spatial Crowdsourcing model and the proposed strategy for spatio-temporal task allocation are agnostic to the specifics of the Citizen Science initiative, they can be effectively applied in this context. Consequently, each task, in this case, involves taking a photo at the designated location. As previously mentioned, this use case provided an opportunity to assess the effectiveness of the Bio-VTAS algorithm in fostering altruistic participation within the context of a real-world experiment, as well

¹<https://www.deusto.es/en/home>

as to evaluate user adoption and acceptance (Goal 1). Additionally, Goal 2 and Goal 3 were evaluated in this use case to assess the effectiveness of the Bio-VTAS in a real-world environment.

For this use case, the Micro-Volunteering Engine (MVE) governed by the Bio-VTAS strategy, was integrated with a Telegram¹ bot, specifically developed to interact directly with users. The bot's functionality is explained in the previous chapter and more in detail in Appendix C.

For this experiment, 13 participants were recruited using a convenience sampling method from the Deusto University staff. Each participant booked two hours of time, during which they received a training presentation lasting a maximum of thirty minutes. The remainder of the time was allocated for conducting the experiment. During the presentation, the goals of the experiment and the functionality of the Telegram bot were explained. Upon completion of the training, participants began the experiment.

The experiment took place on July 19, 2024, from 11:00:00 to 13:00:00, centred at the University of Deusto (longitude: -2.9396777560573013, latitude: 43.271157163906985). The parameter C_d was set to 0.05km, and the campaign's radius was 0.18 km. Additionally, several cells were removed due to their inaccessibility. The cell distribution is shown in Figure 5.11, which also displays the final collage. The campaign concluded when users stopped receiving recommendations, it was either because the task-completion goal had been achieved or the allotted time had expired. Compared to the Deusto use case conducted to test the VTAS (Section 4.4), the current Deusto pilot involves more participants and one less cell, due to accessibility issues identified in the previous chapter. Consequently, the variable M_s has been increased to 5 in order to match the effort per person to the previous Deusto use case, so the campaign task completion goal is to obtain 100 photos.

Since the campaign was small in scale and the number of tasks assigned to each user was limited, the maximum threshold value was adjusted in this experiment to allow for user specialisation within the cells during a short campaign duration. For this reason, θ_{max} was set to 5.0. The remaining

¹<https://web.telegram.org/a/>

This section is organized as follows: Subsection 5.3.1 presents the simulation vs real-world evaluation of the Deusto use case. Subsection 5.3.2 presents the survey evaluation of the Deusto experiment. Finally, Subsection 5.3.3 presents the conclusions of the Deusto experiment.

5.3.1 Simulation vs real-world Evaluation

In this section, the Bio-VTAS algorithm is evaluated using a real-world scenario involving real users from the University of Deusto. The task allocation process began at 11:20 and concluded at 12:20, as users stopped receiving recommendations once the experiment's task-completion goal was achieved before the campaign ended. The objective, based on the variable M_s , was to collect 100 images; however, this goal was surpassed, with a total of 110 photos collected. Two factors contributed to this outcome: 1) When a recommendation is generated, another participant may have accepted it beforehand, creating the potential to exceed the workload limitation. 2) Due to GPS accuracy issues and the proximity of cells that were relatively close to each other. The location of the cell's center complicates taking a picture within its boundaries, as nearby locations can be easily associated with other cells. While these reasons facilitated participants in exceeding the limit, it is important to note that even though the workload limitation was surpassed, the overall effort remained constrained.

To compare the effectiveness of this experiment with the simulated one, a simulation of this campaign was recreated. A simulation of the baselines (LNN and NN) and the VTAS and Bio-VTAS were used for this proposal. Simulations was configured with a 50% User Available probability and a 75% User Task Completion Probability, with 13 users.

Table 5.11 presents the Task Completion Ratio (TCR) metric for this campaign, using the five algorithms: the real-world Bio-VTAS experiment as well as the simulated Bio-VTAS, LNN, VTAS, and NN algorithms. The table complements this analysis by summarizing the TCR metric for the real-world experiment and simulations. The real-world Deusto Campaign using Bio-VTAS achieves an TCR of 110%, indicating an exceptional task performance.

		TCR
Real-world	Real-world Bio-VTAS	110%
	Bio-VTAS	52%
	VTAS	43%
Simulations	LNN	47%
	NN	41%

Table 5.11: TCR metric obtained in the Deusto experiment and in the simulations of this campaign using the Bio-VTAS, VTAS, LNN, and NN algorithms.

However, the simulations fall significantly short, with Bio-VTAS achieving the highest TCR among them at 52%, followed by LNN (47%), VTAS (43%), and NN (41%). This highlights the gap between the real-world implementation and simulations, with Bio-VTAS consistently showing good performance but still underperformed compared to the real-world results.

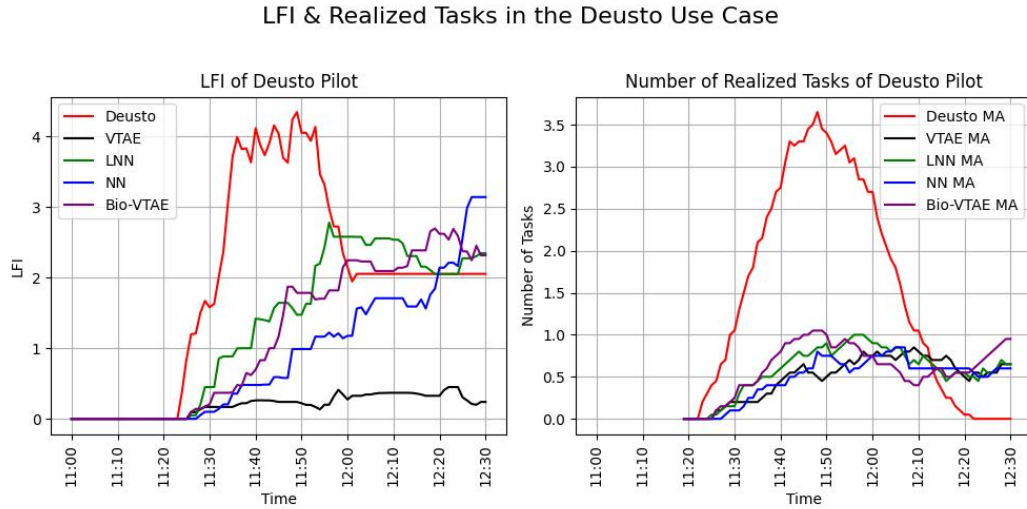


Figure 5.12: The LFI metric (Eq. 4.4- left picture) and Realised tasks (Eq. 4.3 - right picture) over time in the Deusto experiment and the result obtained in its simulation using the VTAS, LNN, NN and Bio-VTAS. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used.

The right-hand side of Figure 5.12 shows the moving average of the number of realised tasks per minute (Equation 4.3) throughout the experiment. To capture the task completion trend, a moving average of 20 values was applied. The data includes results from the real-world Bio-VTAS experiment, labelled

as Deusto in the Figure, as well as the simulated Bio-VTAS, LNN, VTAS, and NN algorithms. The real-world Bio-VTAS (Deusto line) data reaches a significantly higher peak in the number of tasks completed, particularly between 11:20 and 12:20, compared to the simulation results. This can be attributed to the high level of user participation, which allowed the campaign’s objective to be achieved in record time. Among the simulated algorithms, no clear difference is observed, suggesting that while these simulations reflect some dynamics of the realised tasks metric, none replicates the real-world performance. These findings are further supported by the $\text{MEAN} \pm \text{SD}$ values of this metric presented in the last column of Table 5.12. The real-world Bio-VTAS demonstrates a significantly higher mean RT compared to the simulations, with Bio-VTAS being the second-best performer.

		MEAN $\pm$ SD(LFI)	MEAN $\pm$ SD(RT)
Real-world	real-world Bio-VTAS	2.59 $\pm$ 0.97	1.58$\pm$1.92
	Bio-VTAS	1.61 $\pm$ 0.86	0.78 $\pm$ 0.79
	VTAS	0.26$\pm$0.10	0.64 $\pm$ 0.67
Simulations	LNN	1.77 $\pm$ 0.79	0.70 $\pm$ 0.85
	NN	1.28 $\pm$ 0.86	0.61 $\pm$ 0.83

Table 5.12: $\text{MEAN} \pm \text{SD}$ of the LFI and RT metrics obtained in the Deusto experiment (real-world Bio-VTAS) and in the simulations of this campaign using the Bio-VTAS, VTAS, LNN, and NN algorithms.

Left-hand side of the figure 5.12 illustrates the LFI (Eq. 4.4) throughout the experiment for the real-world Bio-VTAS, VTAS, LNN and NN. The Deusto real-world data shows the highest LFI, peaking sharply around 11:20 before declining steeply, as the graph begins to stabilize due to the workload limitation. This is due to the large number of tasks completed during this period, which led to an increase in the LFI. In contrast, the simulation results for the algorithms demonstrate a much more controlled and lower LFI throughout the experiment. The VTAS demonstrated a minimal LFI throughout the experiment, while the LNN, Bio-VTAS, and NN algorithms showed an increase in LFI after 12:10. This suggests that although the simulation algorithms capture some dynamics of the LFI, none replicate the drastic peaks

observed in the real-world experiment. These findings are further supported by the $\text{MEAN} \pm \text{SD}$ values of this metric presented in the third column of Table 5.12. The VTAS achieve the lowest LFI, while the Bio-VTAS and LNN algorithms show the highest LFI. This suggests that the Bio-VTAS and LNN algorithms are more prone to fluctuations in the LFI, while the VTAS algorithm is more stable.

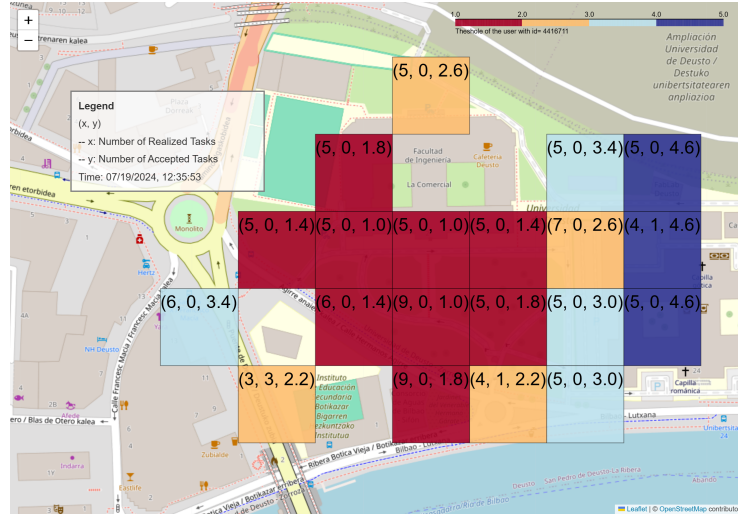


Figure 5.13: Figure of a user expertise level at the end of the Deusto use case. In this case, the user has a expertise in the red zone of the map. The three number of each cell represent the number of task realised $real_{j,i}^t$, number of task accepted $accept_{j,i}^t$ and the threshold of this use for this cell $\theta_{i,j}^t$. The threshold colour scale is based on the threshold value of the user for each cell.

At the end of the experiment, we can observe the level of specialization of the users. In general, they are divided into two groups: users who have specialized in one area and users who have not fully specialized. An example of a person who has specialized in an area can be seen in Figure 5.13, while a user who has not fully specialized is shown in Figure 5.14. The user who has specialized in an area has a higher number of tasks completed in that area. This strategy is designed for users who specialise, as they can benefit from receiving recommendations tailored to their own participation routine. In contrast, users who have not specialised will receive recommendations that may seem more random. In this case, the group of users with greater specialisation

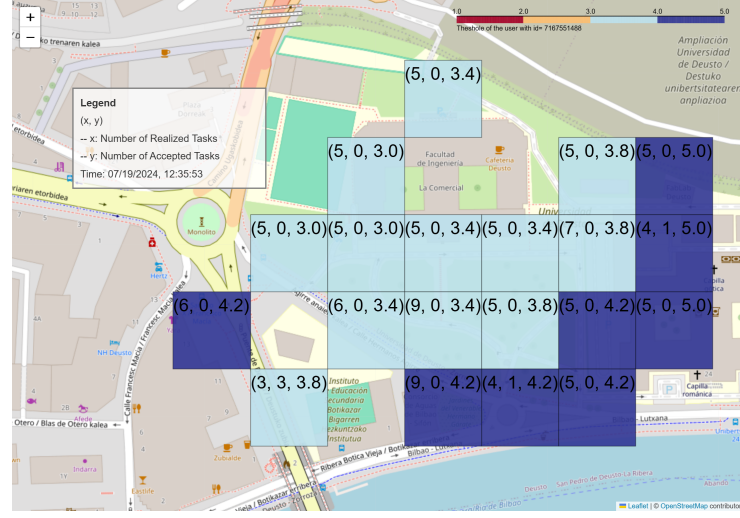


Figure 5.14: Figure of a user expertise level at the end of the Deusto use case. In this case, the user does not have any expertise in any part of the map. The three number of each cell represent the number of task realised $real_{j,i}^t$, number of task accepted $accep_{j,i}^t$ and the threshold of this use for this cell $\theta_{i,j}^t$. The threshold colour scale is based on the threshold value of the user for each cell.

has been able to complete more tasks than the group of non-specialised users and, furthermore, they will have interacted with a more personalised version compared to those who have not specialised. This is a key aspect of the Bio-VTAS user-centric solutions, as it allows users to interact with the system in a more personalised way, increasing users' motivation and engagement.

In conclusion, the experiment was notably shaped by various factors, including the high motivation and engagement of users, who demonstrated a strong commitment to completing the task within a short timeframe. The progression of tasks revealed an unexpectedly high level of user involvement, likely driven by the novelty or 'Wow' effect. Despite this increased participation, the LFI observed in the real-world Bio-VTAS remained within manageable levels. As demonstrated in the previous chapter, the analyses indicate that the simulation's behaviour does not fully replicate human behaviour. This discrepancy stems from the inherent complexity of human behaviour, which cannot be entirely captured in a simulated environment.

5.3.2 Survey Evaluation

This section presents the evaluation results obtained through a survey conducted at the end of the campaign. As mentioned earlier, 13 users participated in the study, recruited using a convenience sampling method among university staff. The survey, consisting of 27 questions, which are detailed in Appendix B. Its primary goal was to assess users' acceptance, adoption and perception of the system. The questionnaire was largely the same as the one used to evaluate the user perspective of the VTAS in Chapter 4, with the addition of three new questions aimed at comparing the experiment realised with the VTAS and with the Bio-VTAS. These additional questions were answered only by the five participants who took part in both the first and the second Deusto use case.

	%	N. of people
Age		
19-35	46.2%	6
36-50	46.2%	6
51-65	7.7%	1
Others	0%	0
Gender		
Female	23.1%	3
Male	76.9%	10
Other	0%	0
Level of education		
PhD	53.8%	7
Master's degree	46.2%	6
Others	0 %	0
Technological knowledge		
Advanced Level	100%	13
Others	0	0
Work Status		
Employed	100%	13
Others	0 %	0

Table 5.13: Demographic information of the Deusto use Case.

The survey begins by collecting demographic information, summarized in Table 5.13. A significant majority of participants are university employees

(100%), predominantly male (76.9%), possess high technological proficiency (100%), and fall within the 19–35 (46.2%) and 35–50 (46.2%) age range. Participants with advanced technological proficiency reported engaging in activities such as making online purchases and using spreadsheets, while those with intermediate proficiency described tasks like browsing websites and watching videos on platforms like YouTube. The convenience sampling method employed for recruitment focused exclusively on university staff, resulting in a relatively homogeneous participant pool, as reflected in the demographic data.

Question 1.1: I think this activity was interesting to me.

Options: Disagree 1 2 3 4 5 Agree
Answers: 0 0 3 5 5
 Mean± SD: 4.15±0.80

Question 1.2: I think this activity was useful to me.

Options: Disagree 1 2 3 4 5 Agree
Answers: 0 1 5 4 3
 Mean± SD: 3.62±0.95

Question 1.3: I think that this activity dealt with an important topic for me.

Options: Disagree 1 2 3 4 5 Agree
Answers: 0 1 7 2 3
 Mean± SD: 3.54±0.97

Question 1.4: Comments or recommendations regarding this activity that you would like to share.

Final Score of participant's interest level:
 Mean± SD: 3.77 ± 0.91

Table 5.14: Inquiries about the participant's interest level in the Deusto experiment. This table shows the users' questions, options, and answers.

Following the collection of demographic information, participants responded to several questions aimed at assessing their satisfaction and interest in the activity. The four questions related to this topic, possible options and results, are presented in Table 5.14. The results for Questions 1.1, 1.2, and

1.3 indicate that the activity was generally enjoyable and valuable for participants, with mean scores of 4.15, 3.62, and 3.54 out of 5, respectively. On this scale, a score of 1 represents strong disagreement (unsatisfied), while a score of 5 represents strong agreement. For the open-ended Question 1.4, some participants reflected that their lack of significant interest stemmed from the nature of the experiment as a proof of concept. They identified limited engagement with the experiment's objectives and a lack of customization in their interactions with the chatbot as key factors contributing to this sentiment. These comments highlight a broader issue: participants did not feel a sense of agency, which is a critical element for Citizen Science experiments. This lack of agency was expected, as the experiment was not a real Citizen Science initiative where participants were personally affected by the cause. Despite this limitation, the results were positive overall, demonstrating participants' interest in the system and its potential value. As in the survey realized with the VTAS, this section was designed to assess the participants' interest in the experiment and their perception of its value. To facilitate comparisons between the results of this study and those of future strategies, we calculated the overall mean of the means and standard deviations of all the questions in this section which are not open answers. This final score indicate that the participants found the activity interesting and useful, with a mean score of 3.77 out of 5. These results suggest that the participants were engaged and interested in the experiment, indicating that the Bio-VTAS has the potential to foster user engagement and interest (Goal 1).

The subsequent set of questions focused on system adoption and potential future use. These questions, possible options, and the number of users that vote for each option are represented in Table 5.15. Based on Question 2.1, 2.3, and 2.5, the future adoption of the system by the participants is proven. The analysis of the responses to the free-text questions, specifically Questions 2.2 and 2.4, confirm that users who do not intend to continue using the system attribute this decision to a lack of interest in the topic or in Citizen Science initiatives. In Question 2.4, with respect to the factors that influence their adoption, the lack of interest in the CS indicatives is the main answer. Another

Question 2.1: Do you see yourself continuing to use the recommender system and in the future?

Options: Yes Maybe No

Answers: 6 5 2

Question 2.2: If no, what would be the primary reason for discontinuing the use of the recommender system?

Question 2.3: How likely are you to adopt the telegram bot or participate in the next campaign?

Options: Disagree 1 2 3 4 5 Agree

Answers: 1 0 2 6 4

Mean±SD: 3.92±1.12

Question 2.4: What factors influence your intention to adopt the bot services in your daily life?

Question 2.5: Would you recommend the recommendation system to others? Or to use in other applications or propose?

Options: Yes Maybe No

Answers: 9 4 0

Final Score of the system's adoption:

Mean± SD: 3.97 ± 1.19

Table 5.15: Inquiries concerning the system's adoption by participants and their perception of it. This table shows the users' questions, options, and answers.

comment is to explore new places and the possibility of using the bot in other applications where users feel more comfortable. The final score¹ of this section is 3.97±1.19, indicating that users are likely to adopt the system and recommend it to others. These results suggest that the Bio-VTAS has the potential to foster user adoption and engagement (Goal 1).

The subsequent set of questions focused on the system's customization and users' perceptions. This set comprised five questions, whose answers can be shown in Table 5.16. The scores slightly exceed the midpoint, suggesting that users perceive a degree of adaptation and customization within the system.

¹To calculate this metric, numerical values were assigned to the responses: 'Yes' was given a score of 5, 'Maybe' a score of 3, and 'No' a score of 1.

Question 3.1: How personalised do you find the recommendations provided by the Telegram bot?

Options: Not at all 1 2 3 4 5 Extremely
Answers: 1 3 3 6 0
Mean± SD: 3.08±1.03

Question 3.2: To what extent do you feel that the recommendations align with your preferences?

Options: Not at all 1 2 3 4 5 Completely
Answers: 0 2 5 6 0
Mean± SD: 3.31±0.75

Question 3.3: How do you perceive the role of personalised recommendations in enhancing your overall experience with the Telegram bot?

Options: Little 1 2 3 4 5 Extremely
Answers: 0 2 3 6 2
Mean± SD: 3.62±0.96

Question 3.4: Are there any specific instances where you felt the recommendations were particularly well-tailored to your needs?

Question 3.5: What improvements, if any, would you suggest to enhance the personalization of recommendations provided by the Telegram bot?

Final Score of users' perceived customization of the system:
Mean± SD: 3.33±0.91

Table 5.16: Questions related to the users' perceived customization of the system. This table shows the users' questions, options, and answers.

These results indicate that the system's customization features are progressing in the right direction and hold promise for further refinement and enhancement. These scores were anticipated, as the Bio-VTAS is designed to adapt to the user's routine, but this use case is a proof of concept, and the user does not have a routine of use. The final pair of questions in this section delved deeper into the user perception of personalization. Valuable feedback was gathered to inform future improvements, such as the ability to specify a preferred area of interest and system customization options to request recommendations for

nearby locations due to tiredness or time constraints. These ideas point towards more dynamic and user-centred customization features. In conclusion, users perceive some personalization, but in the future, they will want more customization and flexible interfaces that allow them to define their actual preferences. The final score of this section is 3.34 ± 0.88 . This result indicates that the users perceive the system as customizable and that they are satisfied with the personalization of the recommendations. This result is consistent with the results obtained in the VTAS evaluation.

Question 4.1: What works well?

Question 4.2: What does not work?

Question 4.3: What could be improved/added?

Question 4.4: What is missing?

Question 4.5: How would you improve it?

Table 5.17: Feedback about the Telegram Bot

The next section of the questionnaire focused on evaluating the performance of the Telegram bot. The results, summarized in Table 5.17, indicate a generally positive user assessment, with participants praising its ease of use, speed, and variety of options. Despite the largely favourable feedback, some users noted a not appropriate time to arrive at the cell to realize the task as an area for improvement. Additionally, a suggestion was made to implement a feature for visualizing task progress, which will be considered for future updates. Despite this, all the comments are positive and highlight the Telegram bot functionalities.

The next set of questions pertains to the recommendations provided by the system and the users' perceptions regarding their effectiveness. These questions are presented in Table 5.18. When asked what features users felt were missing, most indicated that they did not feel anything was lacking. However, some users mentioned they would appreciate the ability to track task progress or view a list of completed tasks to monitor their personal achievements. All users agreed on the system's proper functionality. Regarding potential improvements, suggestions focused on creating a social networking feature,

Question 5.1: Is there anything else you would like to share about your experience with the recommender system?

Question 5.2: What works well?

Question 5.3: What could be improved/added?

Question 5.4: What is missing?

Question 5.5: How would you improve it?

Table 5.18: Recommendations and feedback information

enhancing response times for reaching the site, adding a task progress tracking feature and creating a tool to give feedback. These responses highlight the importance of incorporating engagement strategies to maintain user interest in the platform. Overall, the feedback was positive, with no significant negative comments or personalization needs identified.

The findings from this survey offer valuable insights into user perception and their acceptance and adoption and potential system enhancements. The demographic data revealed a homogeneous participant pool, primarily comprised of technically proficient individuals. The survey results indicate a generally positive user perception of the system, with high engagement levels and appreciation for its potential value. Participants found the activity interesting and useful, though some noted a lack of agency due to its proof-of-concept nature. Adoption rates appear promising, with many users open to future use, though interest in Citizen Science initiatives played a role in continued engagement. While the system’s customization features were acknowledged, users expressed a desire for greater personalization and flexibility. The Telegram bot received favourable feedback for its usability and speed, though areas for improvement, such as task timing and visualization features, were highlighted. Overall, the findings suggest a strong foundation with opportunities for further refinement and enhancement. These results prove the Bio-VTAS’s effectiveness in fostering user acceptance and adoption (Goal 1). A summary of the quantitative results obtained in this survey are show in Table 5.19.

Section	MEAN $\pm$ SD
Satisfaction and interest	3.77 $\pm$ 0.91
Adoption	3.97 $\pm$ 1.19
Perceived customization	3.34 $\pm$ 0.91

Table 5.19: The mean $\pm$ standard deviation (SD) of each quantifiable section of the questionnaire realised after the campaign conducted at the University of Deusto (Bilbao' Campus) using the Bio-VTAS approach.

5.3.2.1 Comparison of the Deusto Experiment with the First Evaluation

Question 5.1: Have you noticed any difference between the recommendations obtained in the first and second pilot?

Question 5.2: Which recommender do you find most usable in your day to day life?

Options:	VTAS	Bio-VTAS	I don't know
Answers:	0	4	1

Question 5.3: Explain the above answers.

Table 5.20: Recommendations and feedback information.

The final set of questions focused on the comparison between the first and second Deusto campaigns, with responses gathered exclusively from participants who took part in both campaigns. The questions and the distribution of responses for single-answer selections are detailed in Table 5.20. Of the five respondents, the majority reported observing a noticeable difference in behaviour between the two campaigns. Several participants noted that the recommendations provided by the new system were better aligned with their locations. Furthermore, in response to the final question, many participants highlighted that the second recommender system reduced the effort required to complete tasks and featured a more intuitive and user-friendly interface. These findings are highly promising, as they reflect improvements in both user experience and the personalization of recommendations.

In summary, the first pilot generated numerous suggestions and comments on areas requiring improvement, particularly in terms of recommendation per-

sonalization. In contrast, the second pilot elicited fewer such comments, with feedback being predominantly positive. Suggestions during the second pilot mainly pertained to engagement strategies and communication features that users proposed incorporating into the system. In terms of adaptation and adoption, the results of the second pilot are very similar to those obtained in the first (see Table 5.21). There is a slight difference in that the average values are slightly higher for VTAS. Apart from this, no significant differences were observed in the evaluation of the VTAS and Bio-VTAS systems respect the quantitative values. Regarding the non-quantitative values, it can be observed that with VTAS, users expressed an urgent need for improvements in many aspects, a concern that was not experienced with Bio-VTAS.

Section	Mean $\pm$ SD	
	VTAS	Bio-VTAS
Satisfaction and interest (acceptance)	3.99 ± 0.87	3.77 ± 0.91
Adoption	4.0 ± 1.20	3.97 ± 1.19
Perceived customization	3.5 ± 0.98	3.34 ± 0.91

Table 5.21: Comparison of the mean $\pm$ standard deviation (SD) for each quantitative section of the questionnaire realised after campaigns conducted at the University of Deusto (Bilbao Campus) using the VTAS and Bio-VTAS strategies.

In conclusion, it can be observed that the comments regarding improvements and suggestions for the Bio-VTAS are neither as numerous nor as diverse as those for the VTAS. Therefore, we can conclude that, in addition to achieving Goal 1, the Bio-VTAS improves upon the results obtained by the VTAS.

5.3.3 Conclusions of the evaluation of the Bio-VTAS algorithm in the Deusto Use Case

The Deusto Use Case evaluated the Bio-VTAS algorithm in a real-world scenario involving participants from the University of Deusto, who were tasked with capturing photos to create a collage. The experiment aimed to assess

the effectiveness of the Bio-VTAS in fostering user participation and its adoption in a real-world setting. The Bio-VTAS was integrated with a Telegram bot to provide cell-recommendations to users, who could accept or reject them. The experiment involved 13 participants and took place on July 19, 2024. The campaign's goal was to collect 100 photos, which was surpassed with a total of 110 photos collected.

The evaluation methods involved comparing real-world and simulated scenarios and surveys. In the real-world setting, the Bio-VTAS achieved a Task Completion Ratio (TCR) of 110%, which was significantly higher than in the simulated scenarios, where the highest TCR reached only 52% for the Bio-VTAS. The real-world experiment demonstrated better user participation and higher task completion rates compared to the simulations. Overall, the results of the experiment were strongly influenced by factors such as the users' high motivation and engagement, which drove their determination to complete tasks within a short time-frame. Task progression revealed an unexpectedly high level of user involvement and commitment, likely influenced by the novelty or "Wow" effect. Despite the increased participation, the LFI observed in the real-world Bio-VTAS remained within manageable levels. Although some behaviours observed in the simulation mirrored those seen in the real-world Bio-VTAS, it became clear that the simulation did not accurately replicate human behaviour. This discrepancy stems from the complex nature of human behaviour, which cannot be fully captured in a simulated environment.

The survey results indicated that participants found the activity interesting and valuable, with a general willingness to adopt the system in the future. However, some participants noted a lack of significant interest due to the experiment being a proof of concept rather than a genuine Citizen Science initiative. The comparison between the first and second Deusto campaigns showed noticeable improvements in user experience and recommendation personalization (better Goal 1). Participants reported that the new system better aligned with their locations and reduced the effort required to complete tasks. The feedback was predominantly positive, with suggestions focusing on engagement strategies and communication features.

Overall, the Bio-VTAS demonstrated strong real-world performance, but the experiment highlighted the challenges of replicating human behaviour in simulations. The feedback from participants underscored the importance of personalization, flexibility, and sustained engagement for the successful adoption of the system.

5.4 Summary and Conclusions

This thesis addresses the problem of spatial crowdsensing, specifically focusing on task allocation in the context of altruistic participation. Key challenges associated with altruistic participation include potential decreases in participation and loss of user interest due to work overload. Additionally, the typical approach of promoting participation through financial rewards cannot be used in this context.

In this context, the Bio-inspired Volunteer Task Allocation Strategy (Bio-VTAS) algorithm offers several benefits and addresses key challenges associated with task allocation in the context of altruistic participation. One of the primary benefits of the Bio-VTAS algorithm is its ability to enhance user experience by adopting a user-centric approach. This approach considers user preferences and workload limitations, preventing overloading and fostering sustained participation. Additionally, the Bio-VTAS algorithm optimizes task allocation by encouraging user specialization based on their participation routines, which further enhances engagement and efficiency. This strategy was designed to create a more user-centric approach than the one proposed in the previous chapter, aiming to balance improving user perception and task completion efficiency.

In this chapter, this strategy was explained and evaluated through two use cases. Each use case has several experiments to assess the Goal 1, Goal 2 and Goal 3. The first use case is a simulation that models the task allocation problem based on the Spatial Crowdsourcing model defined in Chapter 3. The second use case is a real-world experiment conducted at the University of Deusto. The results of these experiments provide valuable insights into the effectiveness of the Bio-VTAS algorithm in addressing the challenges of task

allocation in altruistic spatial crowdsensing contexts. This strategy aims to enhance the results obtained in the previous chapter respect the Goal 2 and Goal 3, as well as to improve users' perception of the system (Goal 1). This was the expected result; however, the Bio-VTAS only improves the user's perception of the system (Goal 1) but does not enhance the technical performance of the last strategy: the VTAS.

More in detail, the first use case demonstrates that Bio-VTAS achieves better results than LNN and NN in the TCR and RT metrics when workload limitations are not in effect. In these cases, Bio-VTAS successfully meets Goal 2. Similarly, although it does not outperform NN when workload limitations are applied, Bio-VTAS remains the preferable choice in the context of Citizen Science. This is because, in an altruistic participation setting, ensuring that users do not feel overburdened is more important than maximising task assignment efficiency. Overall, this confirms that Bio-VTAS successfully meets Goal 2. Moreover, Bio-VTAS outperforms the baseline algorithms, LNN and NN, in maintaining a low LFI, thereby achieving Goal 3. However, it exhibits an increase in this metric over time due to specialisation. Although Bio-VTAS surpasses the LNN and NN baselines, it does not outperform VTAS in terms of technical performance, which is directly related to Goal 2 and Goal 3. However, this outcome was expected. As a more user-centric solution, Bio-VTAS was anticipated to perform slightly lower in terms of the campaign's technical objectives. Nevertheless, it integrates users more deeply into the decision-making process, fostering a significantly more personalised and user-focused experience, which aligns with the needs of Spatial Crowdsourcing solutions in altruistic contexts.

In the second use case, user feedback on the system was gathered through questionnaires conducted at the end of a campaign involving real users using the Bio-VTAS. The results show that users were satisfied with their experience and expressed a willingness to participate in future campaigns, thereby achieving Goal 1 and surpassing the evaluation results of the VTAS. As a result, both user acceptance and adoption of the system have been validated (see Tabla 5.21). Additionally, feedback regarding potential improvements or issues was minimal and highly specific, primarily provided when directly asked.

This suggests that users were more satisfied with the Bio-VTAS compared to the VTAS.

In conclusion, while the Bio-VTAS does not outperform the VTAS in terms of achieving the goals outlined in Goal 2 and Goal 3, it does succeed in improving user perception of the system, thus fulfilling Goal 1. Therefore, it can be stated that the Bio-VTAS represents an improvement over the VTAS, as it enhances the user experience while delivering comparable technical performance.

In future developments, a solution should be designed that maintains a personalised strategy while also improving the technical objectives defined in Goal 2 and Goal 3. This would allow for a balanced approach, aligning task assignment efficiency with a positive user perception. Therefore, in the next chapter, this solution will be further refined to enhance technical performance without compromising user experience.

*Use what talents you possess: the woods would be very
silent if no birds sang there except those that sang best.*

Henry Van Dyke

CHAPTER

6

Hybrid Volunteer Task Allocation Strategy: Hybrid-VTAS

Citizen Science (CS) and Spatial Crowdsourcing (SC) operate under fundamentally different paradigms, particularly in how they engage and motivate participants. Traditional Spatial Crowdsourcing systems rely on financial incentives to encourage user contributions, whereas Citizen Science thrives on voluntary participation driven by curiosity, altruism, and a commitment to collective knowledge. This contrast presents a challenge when attempting to adapt Spatial Crowdsourcing methodologies for Citizen Science applications, as direct monetary compensation is neither a viable nor a desirable mechanism for sustaining engagement. **Instead, alternative task allocation strategies should be designed to balance task assignment optimisation with the intrinsic motivations of volunteers, ensuring that participation remains both meaningful and effective.** To address this, previous chapters have introduced Spatial Crowdsourcing strategies that shift away from financial incentives and instead focus on enhancing user involvement in the decision-making process.

These approaches have demonstrated that engagement can be improved by balancing task efficiency with user perspective, reducing cognitive overload, and respecting individual preferences. However, although these strategies have made notable advancements, further refinements are needed to develop a more user-centred solution that better balances system optimisation and user satisfaction.

In the previous chapters, the first two strategies, VTAS and Bio-VTAS, were introduced and evaluated against three key goals that serve to validate our hypothesis: Goal 1, which aims to enhance user perception by increasing acceptance and adoption; Goal 2, which seeks to optimise task allocation by maximising task completion without exceeding workload limitations; and Goal 3, which ensures equitable task distribution by guaranteeing an equitable allocation of tasks among cells. The first proposed solution, VTAS, introduced in Chapter 4, successfully accomplishes Goal 2 and Goal 3 by optimising task allocation and maintaining an equitable in task distribution. While users expressed both acceptance and adoption of the system, it became clear that a stronger emphasis on user involvement in the decision-making process was necessary. VTAS primarily relies on user proximity as a customisation parameter, limiting its ability to adapt to individual routines and preferences. To address this, the second proposed solution, Bio-VTAS, presented in Chapter 5, was designed to incorporate user specialisation based on participation patterns. By adapting to users' behaviours, Bio-VTAS provides more tailored task recommendations, thereby improving user perception and engagement (Goal 1). However, despite enhancing personalisation, Bio-VTAS does not surpass VTAS in achieving Goal 2 and Goal 3, although it demonstrates comparable performance across several observed metrics.

Both strategies align with Goal 1, as evidenced by survey results, which indicate similar levels of user acceptance and adoption. However, a deeper analysis of qualitative feedback reveals key distinctions. VTAS users expressed dissatisfaction with the system's perceived randomness, frequently requesting improvements such as engagement techniques or gamification elements to enhance their experience. In contrast, while Bio-VTAS users also highlighted areas for improvement, these concerns emerged primarily when expli-

citly prompted rather than as a widespread demand. This suggests that, overall, users experienced greater satisfaction with the system’s approach when using Bio-VTAS. Nevertheless, the lack of superior quantifiable outcomes indicates that, despite achieving greater personalisation, Bio-VTAS may not fully align with user preferences regarding its implementation.

To refine these strategies further and develop a more user-centric solution that achieve better results related to the Goal 2 and Goal 3, this chapter begins by investigating how users prefer to interact with the system in terms of personalisation, usage patterns, and engagement levels. To capture a broad range of perspectives, face-to-face interviews were conducted, focusing on five key aspects: *system usage*, *frequency of use*, *feedback*, *motivation*, and *overall impressions*. The full interview questions is provided in Appendix D. Insights from these interviews will inform the development of a hybrid solution, Hybrid-VTAS, which seeks to integrate the strengths of both VTAS and Bio-VTAS, balancing efficiency, fairness, and user engagement more effectively. This survey provides insights into how users want to interact with the system, extending beyond what is necessary to propose the strategy in this chapter. As mentioned in the limitations, the goal of this research is not to develop motivation and engagement tools, but to create Spatial Crowdsourcing strategies that allow users to feel part of the system balancing this with obtained an optimize and effective task allocation procedure. While integrating motivation and engagement mechanisms could enhance this solution in future work, this is beyond the scope of this thesis.

Therefore, this chapter begins by analysing the interview responses from participants involved in the second Deusto Use Case, which implemented the Bio-VTAS solution. The insights gathered from this analysis will inform the development of Hybrid-VTAS, an *ensemble algorithm* that merges the strengths of both VTAS and Bio-VTAS. While VTAS focused on equitable task distribution and task completion efficiency, and Bio-VTAS introduced user-centric adaptations for greater personalisation, Hybrid-VTAS seeks to refine the balance between these aspects. By aligning strategy creation with user preferences regarding their interaction with Spatial Crowdsourcing technology in the Citizen Science domain, this approach aims to enhance per-

formance across the three core goals. Ultimately, Hybrid-VTAS aspires to maintain efficiency and equitable while fostering active user participation by adapting to individual preferences and interaction patterns.

In addition to explaining the solution and how interviews guided the creation of the ensemble solution, this chapter also presents an evaluation of the proposed solution. In contrast to the previous two chapters, this solution will not be tested with a Deusto Use Case, i.e., with a real-world scenario. The evaluation relies on simulations, similar to those conducted in earlier chapters, to assess its performance. A real use case is considered unnecessary, as the adoption and acceptance assessments from previous chapters already verify the effectiveness of the proposed solution in relation to Goal 1, given that this new strategy is an ensemble approach.

The chapter is organized as follows: Section 6.1 presents the results of the face-to-face interview conducted to capture a wide range of perspectives regarding users interaction with Citizen Science enabling technologies. Section 6.2 introduces the proposed collaborative solution, Hybrid-VTAS, based on the results of the previous chapters and the face-to-face interviews. The evaluation metrics used in this section are consistent with those from the previous two chapters, and the baseline models for comparison include VTAS, Bio-VTAS, LNN, and NN algorithms. However, this section will focus on a single use case: the simulated use case employed in the previous chapters. Thus, section 6.3 outlines the simulation use case, which tests the performance of Hybrid-VTAS in a controlled environment. Finally, Section 6.4 summarizes the key findings of this study. Once more, the reader is invited to review Figure 3.5 for a comprehensive overview of the chapter structure and its connection to the previous ones. Concepts such as baselines, evaluation metrics, and use case details have already been introduced in Chapters 4 and 5 and will be referenced here without unnecessary repetition.

Be kindly reminded that the key definitions of this thesis are at glossary titled **Key Concepts** and the mathematical notation utilised is published at glossary titled **Mathematical Notation**.

6.1 Face-to-face interview Evaluation

This section describes the evaluation obtained from face-to-face interviews conducted individually with each participant at the completion of the Deusto Use Case, in which the Bio-VTAS was tested. The 13 users who participated in the Deusto use case with the Bio-VTAS were the same ones who conducted these interviews. As was described in the experimental procedure, these 13 users were recruited via a convenience sampling method from the university staff. The set of questions for the interview was developed based on feedback from the Deusto Use Case, which assessed the effectiveness of the VTAS and the Bio-VTAS (described in Section 4.4 y Section 5.3). During these evaluations, it became evident that participants had varied expectations regarding how the system should function. Consequently, the interview was tailored to capture these diverse perspectives in order to refine and enhance the user experience. The primary aim of these interviews was to gain a comprehensive understanding of different participant's unique perspective on interacting with the technology used in Citizen Science, ultimately enhancing both their engagement and comfort. The interview consisted of 20 questions, organized into five sections: *system usage*, *frequency of use*, *feedback*, *motivation*, and *interview conclusions*. The interview is provided in Appendix D.

The first block of questions is related to how users want to interact with the system. The questions are shown in Table 6.1. Regarding Question C1.1, participants expressed various ways they would integrate the system into their daily routines. Eight users out of thirteen explicitly mentioned their desire to use it to discover new areas, either during trips or as an occasional activity. On the other hand, five participants stated that they would also incorporate it into their daily routines, although several would combine this with a more exploratory use. A significant insight is that all respondents agreed that the level of usage would depend directly on how motivating they found the underlying Citizen Science campaigns and their alignment with the such campaign's goals.

Regarding preferences for selecting areas (Question C1.2), nine participants highlighted that they would prefer to explore nearby places, especially when

- Question C1.1:** How do you think you would use this system in your daily life? Would it be during your free time, while getting around the city, on your way to work, to explore new places, or a mix depending on the situation?
- Question C1.2:** When using the system, would you prefer to visit nearby places, go further away, pick a specific area yourself, or have no restrictions?
- Question C1.3:** Would you like the system to learn from your choices and suggest options based on what you’ve liked before?
- Question C1.4:** Do you prefer having multiple options to choose from?
- Question C1.5:** Do you think your preferences for distance and usage might change in the future?

Table 6.1: Table of questions related to how users want to interact with for spatial crowdsourcing technologies used in Citizen Science domain.

integrating the system into their daily routines. Some noted that the choice of closer locations would depend on factors such as weather or available time. Other five participants expressed interest in being able to customize the areas of exploration, selecting specific regions or applying thematic filters based on their objectives. An interesting point is that all users mentioned that they would like the system to dynamically adapt to their needs and behaviour, offering relevant recommendations based on their prior interactions.

Respecting the Question C1.3, all participants responded positively to the idea of the system learning from their preferences and providing pre-selection or filtering based on past interactions. This reflects a clear interest in intelligent systems that adapt to the user, simplifying and optimizing the experience.

Twelve participants indicated a preference for receiving a limited selection of relevant recommendations rather than all available ones as inferred from Question C1.4. Ten participants specified that the ideal number of options would be between two and three, while one participant appreciated the ability to choose how many options they wished to receive. However, one user stated that they would prefer access to all alternatives, as long as the system allows filtering based on their interests.

Nine participants believed that their usage preferences could change over time, depending on the situation or the context in which they use the system (Question C1.5). Some mentioned that usage might increase during leisure activities or tourism, while others indicated that if they lose interest, the frequency of use could decrease. On the other hand, four participants noted that their preferences would remain stable, although they emphasized the importance of the system allowing for adjustments to settings if necessary.

In general, the interviews results reflect a significant interest in a system that combines flexibility, personalization, and adaptability. Users value the ability to integrate the Spatial Crowdsourcing recommender into both their daily routines and specific moments of exploration, as long as it manages to motivate them. Additionally, there is a clear preference for personalised filters, limited options, and the system's ability to learn and evolve over time.

Question C2.1: How often do you think you would use this system?

Question C2.2: Would you like the system to notify you when it has something relevant for you, or would you rather decide when to use it yourself? Would you want to set your own rules for when you get notifications?

Question C2.3: Do you think how often you use the system might change in the future? What factors do you think could influence this change?

Table 6.2: Table of questions related to the frequency of use of the users.

The second block of questions is related to the frequency of use. The questions are shown in Table 6.2. In Question C2.1 related to the frequency of system use, the responses reflect a variety of behaviours, ranging from sporadic use to daily or weekly use. However, most participants indicated that they would use the system occasionally or under specific circumstances, such as whilst commuting to work or while exploring new places. This suggests that the system's use would be more related to contextual situations than to a fixed routine. Although some participants mentioned the possibility of daily use, this trend does not appear to be predominant, with most agreeing that their frequency of use would depend directly on their affinity with the

system's goals. This highlights that the system's functionality and its ability to integrate into daily activities would play a key role in fostering regular adoption.

Regarding notifications and time management for using the system (Question C2.2), two clear preferences emerge. On the one hand, a significant group of participants prefers to control when they want to receive recommendations, indicating that it would be more convenient to activate the system only when they are free and available. On the other hand, there are those who value the possibility of explicitly defining a notification pattern, but with a non-intrusive approach, to ensure that these notifications do not interfere with their daily activities. In general, there is a rejection of automatic notifications, since they are perceived as invasive, reflecting the need for personalization and flexibility in the system's design to adapt to users' individual schedules and preferences.

Regarding the potential evolution of usage frequency (Question C2.3), the responses indicate a mixed perception. A considerable number of participants believe that their frequency of use could decrease over time, especially if the system fails to maintain its relevance or interest in the long term. Key factors that could influence this evolution include how useful the system is in achieving the campaign's goals, the variety of activities offered, and the presence of incentives to motivate continued participation. Some participants expressed that if the activities become repetitive or unappealing, their interest in the system might decline. Conversely, if the system provides novel content and effective functionality and adjusts to their needs, there is a higher likelihood that usage will increase or remain consistent. This highlights the importance of ensuring a dynamic and personalised experience to maintain active user engagement over time.

In conclusion, the results suggest that system adoption depends on its ability to adapt to users' individual circumstances and needs, as well as its capacity to sustain interest and motivational strategies. While some users show a predisposition toward frequent use, the majority indicate that this would depend on the system's utility and design, underscoring the importance of prioritizing personalization, flexibility, and innovation in its development.

Question C3.1: Would you like to rate how much you liked the task, the distance, and the place?

Question C3.2: How would you prefer to give feedback? For example, a 5-star rating, a thumbs-up, or writing your own comments?

Question C3.3: How often would you like to give feedback? Every few days, once a month, or something else?

Table 6.3: Table of questions related to the feedback of the users.

The third block is related to how often and in what manner users want to provide feedback to the system. The questions are shown in Table 6.3. In the analysis of responses from Block 3, participants expressed diverse preferences regarding how to evaluate and provide feedback of the system. These responses offer key insights into the elements they consider relevant when interacting with the system and their expectations for the evaluation process.

Regarding what they would like to rate, most users agree on the importance of assessing aspects related to the task and location, such as the distance, site accessibility, task difficulty, and the overall experience. More subjective elements were also highlighted, such as how much they enjoyed the task or whether completing the task disrupted their routine. Some users even mentioned an interest in evaluating how much they learned from the activity or receiving feedback on the impact of their contribution to the project. However, not everyone shares the same enthusiasm for providing evaluations, as a minority expressed disinterest in participating in such feedback processes due to privacy concerns. Overall, there is an apparent demand for tools that allow users to evaluate both objective aspects (like distance or difficulty) and subjective ones (personal enjoyment and alignment with expectations).

Regarding the method of providing feedback (Question C3.1), the predominant preference is a star-based rating system, mentioned by the majority of users. However, some also value the option to include free text to complement their feedback, but only as an optional feature. A smaller number of participants prefer more straightforward options, such as thumbs up or down. This diversity in preferences suggests that the system should offer flexibility, allowing users to choose the method that best suits their interaction style,

whether it be a more structured approach (stars) or a more open one (free text). Furthermore, this also depends on the level of detail in the feedback that users are willing to provide to the system. In other words, some users may be more engaged and willing to invest more effort in offering feedback if it leads to a better interaction, while others may prefer simpler mechanisms, such as rating, to provide this information. However, one user explicitly expressed disinterest in any form of feedback, reinforcing the idea that this process should always be optional.

As for the frequency of providing feedback (Question C3.2), most participants suggested doing it immediately after completing a task or session. However, there is a common concern that this process should not feel mandatory or intrusive. Some users prefer having the freedom to decide when to give their opinion, while others proposed providing feedback only at key moments, such as when realizing a task for the first time or after accumulating multiple tasks. In this sense, several participants emphasized that feedback should be optional, accessible only when they wish to provide it, and designed in a way that is neither repetitive nor bothersome. This highlights the importance of respecting user autonomy by offering options that do not interrupt the core interaction with the system.

Finally, most participants indicated that they would like to define specific parameters at the beginning of their interaction with the system (Question C3.3). This includes aspects such as the frequency of notifications and feedback preferences. Some also emphasized the need for these settings to be adjustable over time, reflecting an expectation for adapting the system to users' changing needs. However, a minority stated that they do not consider it necessary to configure these options initially, which could reflect a preference for default feedback parameters that adjust automatically based on user interaction.

In conclusion, the analysis reveals that users value personalization, flexibility, and autonomy when interacting with the system and providing feedback. A design that allows users to evaluate various aspects (such as tasks, locations, and distance), offers multiple feedback options (stars, free text, or none), and avoids mandatory or intrusive processes is more likely to create

a positive and sustained perception. At the same time, there is a clear need to offer adjustable initial settings so each user can tailor the system to their usage style, reinforcing the idea of a user-centred approach that prioritizes individual preferences.

Question C4.1: What would keep you interested in using the system over time? For example:

- Tracking your participation and sharing it on social media (WhatsApp, Telegram, X, TikTok, Facebook, etc.)
- Competing with others on a ranking list
- A built-in social network
- Getting rewards like interesting activities or useful information.

Question C4.2: What would encourage you to try new ways of using the system beyond what you're used to? Or do you think your habits wouldn't change?

Table 6.4: Table of questions related to the motivation of the users.

The analysis of Block 4 reveals interesting patterns regarding the motivations and incentives that could sustain long-term user engagement with the system. The questions from Block 4 are presented in Table 6.4. Overall, the responses emphasize the importance of an approach that integrates social elements, individual recognition, and the overall relevance of the project's purpose.

In terms of factors that could motivate users to continue using the system (Question C4.1), a recurring theme is the importance of individual recognition and the visibility of personal contributions. Many participants mentioned that they would like to see how much they have contributed to the project, whether through participation statistics, rankings, or even symbolic rewards. Rankings, in particular, stand out as a tool that fosters a sense of friendly competition, which many users find motivating. Others highlighted the desire to access interesting insights, participate in engaging activities, or understand the social benefits generated by their contributions, suggesting that the system should effectively communicate the tangible impact of each completed task.

The social dimension also emerges as a strong driver of motivation. Some participants value the idea of sharing their achievements on social networks such as WhatsApp, Instagram, Facebook, or TikTok, while others suggested integrating a social network directly into the system’s interface. These social networks could include features such as interaction with friends, viewing shared activities, and exchanging opinions. For certain users, connecting with their social circle could be a key reason to remain active, reinforcing the role of community as part of the experience. Additionally, challenges and competitions between users or specific groups were also mentioned as appealing elements, reflecting a preference for social dynamics that blend collaboration and competitiveness.

On the other hand, some participants indicated that their primary motivation would stem from the alignment of the system with the project’s overarching purpose. For this group, understanding the positive impact of their contributions—whether in terms of social benefits or advancements in Citizen Science—is a decisive factor. This altruistic approach, where the value lies more in the collective outcome rather than in personal benefits, underscores the importance of effectively communicating the system’s ultimate goal and the achievements made possible through user participation.

Regarding what could motivate users to step out of their comfort zone (Question C4.2), the responses suggest that this would depend on specific and relevant incentives. Some participants mentioned that unique, challenging, or less-explored tasks could capture their interest, while others emphasized the impact of social recognition, such as reviews from other users or support from their close social circle. Other factors include symbolic rewards, such as interesting facts or curiosities about the places visited, as well as activities that align with their interests and do not overly disrupt their routines. It is clear that the time and effort required to participate are key considerations: well-designed activities that fit users’ schedules and preferences could make them more willing to engage in less familiar tasks.

In conclusion, users’ motivations combine intrinsic and extrinsic aspects. However, there is strong interest in the relevance and impact of the project, suggesting that communicating the social benefit should be central to

the design of effective Spatial Crowdsourcing strategies. On the other hand, social dynamics and mechanisms for personal recognition, such as rankings, rewards, and social networks, can be powerful tools to sustain interest and active participation. To encourage users to step out of their comfort zones, it will be crucial to balance offering engaging and novel activities with incentives and a design that respects users' limitations and preferences. This holistic approach could ensure both long-term retention and commitment to more challenging tasks.

- Question C5.1:** Out of everything we've talked about (how to use the system, giving feedback, etc.), what do you think is most important?
- Question C5.2:** Is there anything about using the system that we haven't discussed but should?
- Question C5.3:** If everything we talked about works as expected, would that make you more likely to keep using the system long-term?
- Question C5.4:** Do you think your usage habits will change over time? If so, how? For example, would you change how often you use it, how you give feedback, or what you use it for?
- Question C5.5:** Overall, do you see yourself using this system in the long run? Is there anything else you'd like to share about your experience or this interview?

Table 6.5: Table of questions related to the conclusion of the interview

The last block of questions is presented in Table 6.5. This set of questions were devised to conclude the interview assessing the potential adoption over time of the presented system. The analysis of Block 5 reveals a comprehensive view of users' expectations and concerns regarding the system's utility, interaction, and potential long-term use. The responses highlight the importance of personalization, both intrinsic and extrinsic motivation, and the system's alignment with individual needs and routines.

In terms of priorities, users agree that recommendations should be highly personalised. The system's adaptability, both in terms of content and interaction, is considered crucial. Several participants emphasized that recommendations should not be invasive and should be explained so that users can

understand the reason for their recommendations. This aspect of "explainability" not only reinforces trust in the system but also fosters a sense of connection and relevance with the assigned tasks. Another key element is the facility of use, with an emphasis on a clear and functional interface that allows users to interact without technical difficulties.

The social impact of the project also emerges as a fundamental aspect. Users rate as important the ability to see their participation reflected in a summary or final conclusion of the social benefits achieved through their efforts. Some suggested that this approach could be strengthened by showcasing tangible results of the Citizen Science project. Additionally, for specific users, the social component within the system-such as interactions with other participants or friends-is an important driver that could make the experience more prosperous and more engaging.

Regarding additional aspects that were not widely discussed, privacy emerges as a recurring concern. Several users expressed a preference for the system not to use their personal data invasively, emphasizing the need to safeguard private information. Innovative ideas, such as augmented reality, were also mentioned, although these suggestions were less common.

With regard to the system's long-term usage, participants are aware that their engagement may change over time. A majority acknowledged that initial enthusiasm could diminish. This pattern of declining use suggests that the system must offer continuous updates, novel activities, and motivational elements to sustain long-term interest. Personalized, non-intrusive notifications are seen as a key tool to remind users of their participation, especially during times when they may have lost interest. However, answers indicate that users' willingness to engage with the system largely depends on the specific nature and purpose of each Citizen Science project. If the project's goal is not relevant or fails to generate enough motivation, users may leave the system even if it is functionally robust.

When it comes to long-term use, opinions are mixed. While eight participants imagine themselves using the system over extended periods, whether continuously or in occasionally, four participants believe their use will be sporadic or even temporary. The diversity of these perspectives underscores

the importance of offering flexibility in the system’s design, ensuring it caters to both users seeking occasional interaction and those desiring more consistent engagement. It also highlights that user commitment depends not only on the platform itself but also on external factors, such as the Citizen Science project’s objectives and how motivating these are for users.

In summary, the system’s long-term success depend on its ability to balance three key factors: personalised recommendations and interactions, clear communication of both social impact and individual benefits, and mechanisms to sustain interest over time, such as introducing new incentives and activities. Privacy, ease of use, and non-intrusiveness are also essential components for fostering a positive experience. While opinions on sustained use are varied, it is clear that a system designed with these priorities in mind could maximize user engagement and satisfaction across different contexts and over time.

6.1.1 Conclusions of the face-to-face interviews

This section outlines the result of face-to-face interviews conducted with 13 participants of the real-world experiments conducted in the Deusto university using the Bio-VTAS solution. The analysis of the participants’ demographics is presented in the previous chapter, in Table 5.13. The interview aimed to understand participants’ perspectives on system usage, frequency of use, feedback, motivation, and overall conclusions.

Participants shared diverse preferences for how they would integrate the system into their daily routines, with many expressing a desire to discover new areas and customize their exploration to suit their individual needs and interests. However, all the different usage options were selected by at least one person, demonstrating the versatility that users desire in the system. As a consequence, they appreciated the system’s ability to learn from their preferences and provide relevant recommendations. Most participants preferred receiving a limited selection of options and believed their usage preferences might change over time.

Regarding the frequency of use, participants indicated that their usage would be more situational than part of a continuous routine with a prefer-

ence for controlling when to receive recommendations. They emphasized the importance of the need for dynamic and personalised experiences to maintain interest over time.

In terms of feedback, participants valued the ability to evaluate tasks and locations, with a preference for a star-based rating system complemented by optional free text. They highlighted the importance of non-intrusive and optional feedback processes, with settings adjustable over time.

Motivation to continue using the system was driven by individual recognition, social elements, and the project's relevance. Participants valued seeing their contributions, sharing achievements on social networks, and understanding the social benefits of their participation. They also appreciated challenges and competitions as motivating factors.

The overall conclusions emphasized the importance of personalised recommendations, clear communication of social impact, and mechanisms to adapt to all users preferences and characteristics and eventually sustain interest over time. Privacy, ease of use, and non-intrusiveness were also crucial for a positive user experience. While opinions on long-term use varied, the system's success would depend on balancing personalised interactions, social impact communication, and continuous engagement strategies.

6.2 Algorithm Overview

Based on the feedback obtained in the interviews, the idea is to create a collaborative solution, an ensemble solution, that integrates the strengths of the VTAS and Bio-VTAS algorithms. The insights gained from the interview's answers will be used to determine how to do this ensemble to align the algorithm's personalization with users' preferences regarding how they wish to interact with the system within Citizen Science technologies. The goal is to deliver an enhanced approach that ensures efficiency and equitable task distribution while actively engaging users by accommodating their preferences and participation habits. Thus, this section provides a comprehensive explanation of Hybrid-VTAS. This solution has to achieve the goals explained in Section 4.1: Goal 1) Enhance User Perception, Goal 2) Optimised Task Allocation

Solution and Goal 3) Create an equitable distribution of tasks across all cells. Firstly, we will analyse how the interviews results define the behaviour of the proposed ensemble algorithm. Subsequently, the Hybrid-VTAS will be explained in detail, including the implementation and specification aspects necessary for a clear understanding of the solution.

Based on the responses gathered from the interviews, it is evident that while the majority of users expressed a preference for using the system to explore new places, there is considerable diversity in the approaches among users. This confirms that the creation of an ensemble-type algorithm is an appropriate solution. For those users who are interested in exploratory use, the VTAS system proves to be efficient, as the cell-recommendations will be more widely distributed, thus avoiding dependence on the user's spatial routine. On the other hand, for users who prefer a more routine-based use, the Bio-VTAS system is a suitable alternative, as the cell-recommendations will be concentrated in areas where the user typically interacts, better adapting to their usage patterns. Since most of the users prefer an exploratory approach, the default strategy will be the VTAS. That is, during the first interaction of a user with the spatial Crowdsourcing technology, the VTAS strategy will be used. Ultimately, for each recommendation the user makes, the algorithm must determine, based on their interactions, whether to suggest a VTAS task or a Bio-VTAS task. In the case of the initial interactions with the user, where no information is available, VTAS will be used.

Furthermore, it is important to recall that users indicated a preference for a system that learns from their interactions in a non-intrusive manner, adapting to their needs and usage patterns without necessitating confirmation for each individual decision. This implies that the system must be capable of autonomously detecting changes in the each participant's usage pattern. Consequently, the proposed solution must be designed to transit automatically from an exploratory approach to a more routine-based one. The transition from the default VTAS to Bio-VTAS will be triggered when a high frequency of task acceptance is observed in the same area, suggesting that the user has begun to perform tasks in a routine manner or has specialized in a specific area. Conversely, the user will revert to an exploratory approach when their

task acceptance no longer adheres to a routine pattern, indicating that they no longer specialize in a particular area. Therefore, the specialization variable, or threshold $\theta_{u,j}^t$ (definition at Chapter 5; section 5.1), will act as the determining factor to indicate when a user has shifted from exploratory use to routine use, or vice versa. Additionally, based on the interview responses in terms of intrusiveness, this transition should occur automatically, without the need for user confirmation.

Finally, users also agreed in their responses that they would prefer a pre-selection of possible tasks, rather than being shown all available options. Consequently, the solution proposed in this chapter will take this preference into account, recommending only N cell-recommendations per request received.

All these insights, as described, define the behaviour of the proposed solution in this chapter. In the rest of this section, we will explain in detail how this solution has been developed, based on the *modus operandi* we have just described.

As in previous solutions, this algorithm will still be a recommendation algorithm that recommends N *cell-recommendations* when it receives a request from a user. Similarly, this algorithm reuses the task states that were proposed in Chapter 4. The Hybrid-VTAS is based on an ensemble algorithm that aims to achieve the effectiveness of VTAS with the personalization of Bio-VTAS. To do this, the cell-recommendations that the Bio-VTAS algorithm would generate are calculated. If for any of the recommended cells, the user's threshold for those cells (denoted by the variable $\theta_{u,j}^t$) is lower than a predefined value R, the user will receive those recommendations. Conversely, if the user's threshold for that recommendation is higher than R, the user will receive recommendations from the VTAS. The accepted cell-recommendations, created by the VTAS or by the Bio-VTAS, will update the user's threshold, just as was done with Bio-VTAS, using equations: Eq. 5.1, 5.2, and 5.3.

Thus, the algorithm waits until the user has completed a sufficient number of tasks to establish a specialization before sending recommendations based on their specialization. In this way, the Bio-VTAS algorithm can learn from the user's behaviour, while the VTAS ensures that task distribution remains equitable. Similarly, users with erratic or non-routine participation will only

receive recommendations from the VTAS, which prevents them from receiving recommendations based on a routine they do not follow. Moreover, this approach prevents specialization from resulting in the uneven distribution observed in the second slot of the simulations when Bio-VTAS was used. In that scenario, high specialization led to certain cells becoming overloaded with tasks due to their presence in more specialized areas than others. With the new solution, even erratically behaving (non-routine participation) users contribute to a more balanced task distribution. As a result, it is expected that the Hybrid-VTAS will achieve a lower LFI metric compared to Bio-VTAS.

As with the previous algorithms, the variable N is defined and fixed at three cell recommendations per request. Similarly, the variable R , which defines when a user is sufficiently specialised to be recommended to the Bio-VTAS, is set to a value of 3.0. This threshold helps determine whether a user is considered specialised or not. Therefore, when a user's threshold is below 3.0 in a given cell of the campaign, they will be recommended cells based on Bio-VTAS, whereas if it is 3.0 or higher, they will be recommended cells based on VTAS. The variables of Bio-VTAS chosen here are the same as those used in the previous chapter.

Figures 6.1 and 6.2 show cell-recommendations generated for a user using the Hybrid-VTAS algorithm. In this campaign, the M_s parameter is set to 8, and the completion colour scale is based on this value. Each cell in the image is labelled with a tuple (x, y) , where the first number represents the tasks completed in the cell for the current slot, and the second represents the tasks accepted in the cell. The user, represented by the person icon, is recommended three cells, indicated by the green (the accepted one) and cross icons, with the highest probability. In Figure 6.1, the user is recommended cells that are close to their current location, as the user has not yet specialized in any area, so the VTAS is the corresponding algorithm used throughout the Hybrid-VTAS. In contrast, in Figure 6.2, the user is recommended cells where they have specialized, as indicated by the high number of tasks completed in those cells compared to others in their surroundings.

The ensemble solution, Hybrid-VTAS strategy is integrated into the MVE component which exposes a well defined (OpenAPI 3.0-based) RESTful API

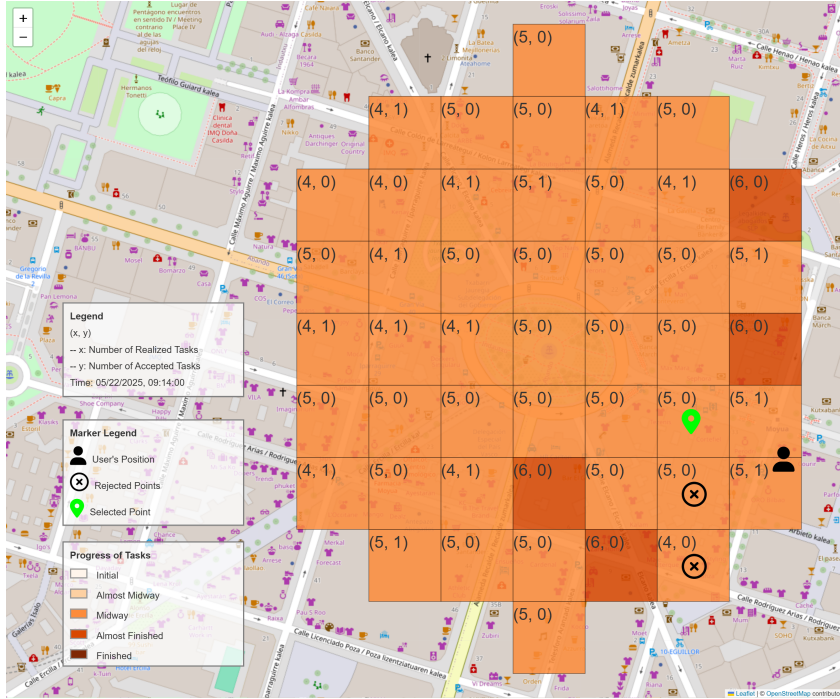


Figure 6.1: An example of N cell-recommendations using the Hybrid-VTAS algorithm. The user, represented by the person icon, is recommended three nearby cells, indicated by the green (the accepted one) and cross icons. In this example, the behaviour of VTAS is observed through the Hybrid-VTAS.

available on GitHub repository¹. Through this API, the Micro Volunteering Engine (MVE), supports the creation of campaigns based on the Spatial Crowdsourcing model described in Chapter 3. The MVE handles task allocation by efficiently matching available users with tasks via the Hybrid-VTAS strategy, regardless of task type or the specific Citizen Science initiative. This ensures broad adaptability and inclusivity in task assignments.

¹<https://github.com/mpuerta004/RecommenderSystem>

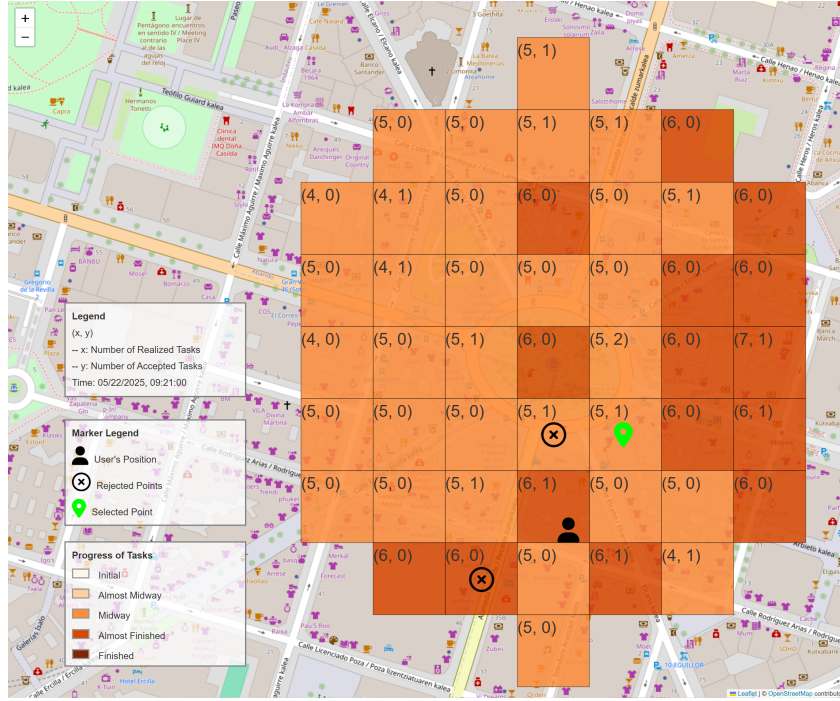


Figure 6.2: An example of N cell-recommendations using the Hybrid-VTAS algorithm. The user, represented by the person icon, is recommended three nearby cells, indicated by the green (the accepted one) and cross icons. In this example, the behaviour of Bio-VTAS is observed through the Hybrid-VTAS.

6.3 Simulated Use Case

This section presents the simulation use case, which investigates the task allocation problem within the framework of the Spatial Crowdsourcing model, as defined in Chapter 3. The primary objective of this study is to evaluate the performance and behaviour of the Hybrid-VTAS under varying conditions. The simulation enables a systematic analysis of the behavioural factors exhibited by simulated users that significantly influence the achievement of the campaign's task-completion objective (Goal 2). Furthermore, it facilitates a comparative assessment of the Hybrid-VTAS against baseline models, namely VTAS, Bio-VTAS, LNN, and NN, highlighting their respective differences. A key focus of this evaluation is the efficiency with which the Hybrid-VTAS allocates tasks (Goal 2). Additionally, in scenarios where the task-completion objective is not fully attained, it is imperative for the algorithm to ensure an

equitable distribution of tasks (Goal 3), thereby preserving the validity and usability of the collected data. This aspect is also rigorously examined within the simulation framework.

This way, this simulation models the task allocation problem to evaluate the effectiveness of the Hybrid-VTAS solution. It replicates user behaviour in the context of altruistic participation, where users adhere to their daily routines. The system accounts for a range of behaviours, from predictable patterns to completely random activities. This simulation setup is identical to the one used in Chapter 4, and it is explained in Section 4.3.

The experiments to evaluate the effectiveness of the Hybrid-VTAS were based on the same three customized variables described in Section 4.3 (Chapter 4): 1) *Number of Users*; 2) *User Availability Probability* and 3) *User Task Completion Probability*. The campaign configuration used in all experiments realised in this section matches that described in Chapter 4, Section 4.3. This configuration included two slots, each lasting 2 hours. The campaign was located in central Bilbao, at a longitude of -2.9349889768009008 and latitude of 43.26302147532318, with a radius of 0.2km. The campaign began at 08:00:00 and ended on the same day at 12:00:00. The C_d was 0.05km, and M_s was eight tasks. Similarly, the parameter values for the Bio-VTAS algorithm have been defined and standardized for all experiments conducted within this use case. The selected variables remain consistent with those outlined in Chapter 5, Section 5.2. Specifically, the threshold for each cell containing a user, $\theta_{u,j}^t$, is set within a range of 1 to 10, with 10 as the upper limit and 1 as the lower limit. The parameter φ is fixed at 0.005, while α and β are both assigned a value of 2.0. Additionally, ξ_0 is set at 2.0, and ξ_1 is maintained at 1.0 across all experiments conducted in this study. A visualization of this campaign is shown in Figure 6.3.

The remainder of this section is organized as follows: Subsection 6.3.1 presents the first experiment, where the Hybrid-VTAS is evaluated with different numbers of users to assess its effectiveness. Subsection 6.3.2 presents the second experiment, in which different User Availability Probabilities are tested to evaluate the Hybrid-VTAS's performance. Subsection 6.3.3 presents the third experiment, where different User Task Completion Probabilities are

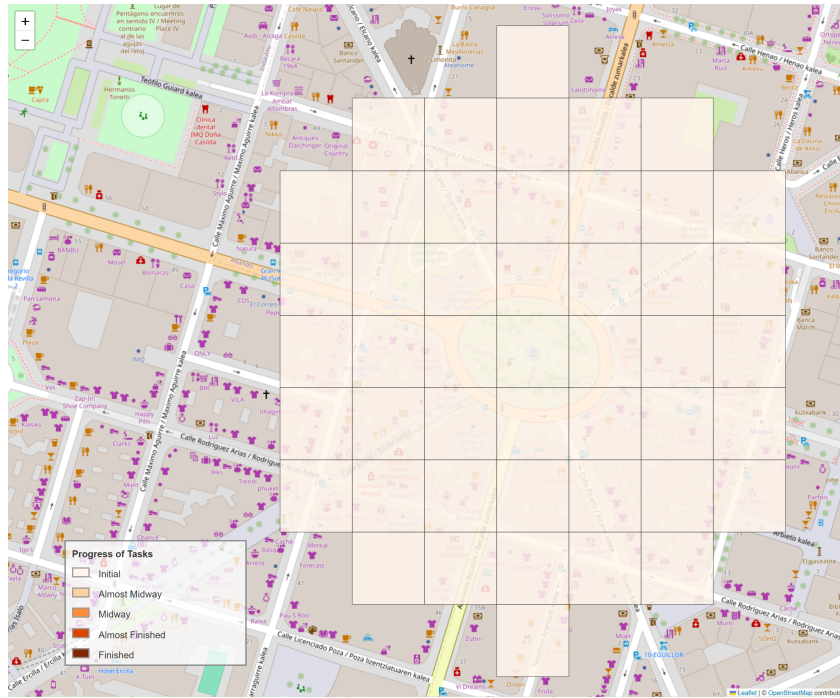


Figure 6.3: Visualization of the campaign configuration used to evaluate the effectiveness of the Hybrid-VTAS solution in this use case. This campaign is the same as the one used in Chapter 4, Section 4.3.

tested to further evaluate the Hybrid-VTAS. Finally, Subsection 6.3.4 provides the conclusions drawn from the experiments conducted in this use case.

6.3.1 Experiment 1: Crowd Size Variation

This experiment evaluates the adaptability of the Hybrid-VTAS in managing task allocation across varying group sizes—100, 75, 50, and 25 users—where smaller groups demand greater individual effort. The analysis focuses on the Hybrid-VTAS’ ability to facilitate task completion under varying crowd size (Goal 2) and assesses its capacity to maintain equitable task distribution in cases where full task completion is unattainable (Goal 3). To ensure consistency, the campaign parameters and objectives remained unchanged from the previous chapters. The simulations were conducted as with a User Task Completion Probability of 75%, indicating strong participant commitment, and a User Availability Probability of 50%. The Hybrid-VTAS was evaluated

against VTAS, Bio-VTAS, LNN, and NN models, with each scenario repeated five times per algorithm. The results presented reflect the average outcomes from these trials.

		TCR			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	100%	100%	94.52%	50.48%
	LNN	100%	98.56%	74.95%	40.59%
	NN	160.96%	121.22%	83.40%	39.89%
	Bio-VTAS	100%	99.9%	85%	44.47%
	Hybrid-VTAS	100%	100%	95.96%	49.89%
Slot 2	VTAS	100%	100%	95.43%	49.15%
	LNN	100%	98.67%	73.46%	39.52%
	NN	157.13%	122.02%	82.29%	41.12%
	Bio-VTAS	100%	99.79%	81.38%	46.01%
	Hybrid-VTAS	100%	99.95%	89.89%	49.31%

Table 6.6: TCR evaluation over the different number of users (100, 75, 50, and 25) using the VTAS, Bio-VTAS, Hybrid-VTAS, LNN, and NN task allocation algorithms. The results are the mean of five executions.

The Task Completion Rate (TCR) metric (Eq. 4.2) is presented in Table 6.6. The table shows that as the number of users increases, the TCR rises across all algorithms due to the reduced individual effort required with more participants. Hybrid-VTAS, Bio-VTAS, VTAS, and LNN incorporate workload constraints, ensuring the TCR remains capped at 100% to prevent user overload, thereby validating the proper implementation of workload limitations. In contrast, the NN algorithm lacks such constraints, exceeding workload limits when 75 and 100 users are involved. In these cases, NN achieves a higher TCR than all other algorithms. However, as previously highlighted, workload-limiting algorithms are more suitable in Citizen Science contexts, where preventing user overload is a key goal. This confirms that even when NN outperforms other algorithms in TCR, Hybrid-VTAS, VTAS, and Bio-VTAS remain the most suitable choices for task allocation in Citizen Science. For group sizes of 50 and 25 users—where workload limitations are not reached—Hybrid-VTAS and VTAS achieve the highest TCR, followed by Bio-VTAS,

LNN, and NN. Thus, Hybrid-VTAS and VTAS demonstrate the best overall performance in terms of the TCR metric.

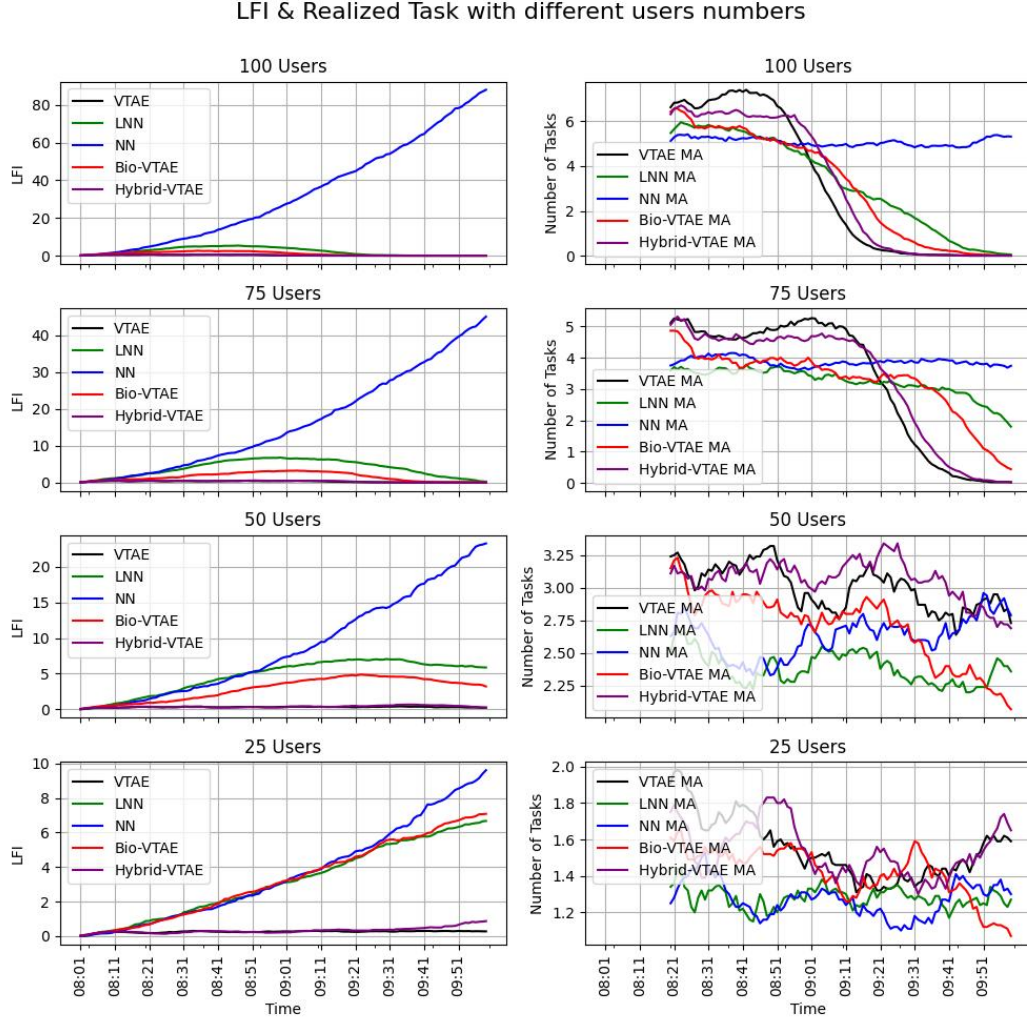


Figure 6.4: The LFI (Eq. 4.4 left-hand side pictures) and the Realised Task (Eq. 4.3 right hand-side figures) over the first slot across different users count (100, 75, 50, and 25 users). The task allocation algorithms used are the Hybrid-VTAS, VTAS, Bio-VTAS, LNN, and NN. To illustrate the evolution of the task realised, the moving average (MA) of 20 values was used. The results presented are the mean of five executions.

The right-hand side of Figure 6.4 illustrates the evolution of Realised Tasks

(RT - Eq. 4.3) per minute for the first slot of the tested campaign using the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms. A moving average¹ of 20 values was applied to emphasise the trend in the evolution of realised tasks. The results represent the mean of five executions. It can be observed that Bio-VTAS, Hybrid-VTAS, and VTAS exhibit a noticeable decrease towards the end of the first slot when considering 100 and 75 users. This behaviour is attributed to the workload limitation, which prevents the cells from processing more than M_s tasks per cell in each slot. This observation is further corroborated by the TCR values presented in Table 6.6, where Hybrid-VTAS, VTAS, and Bio-VTAS achieve a 100% TCR in scenarios involving 100 and 75 users. Moreover, it is evident that the performance of Hybrid-VTAS closely resembles that of VTAS. Additionally, these two algorithms consistently outperform all others, except when the workload limitation is applied. This confirms that both the TCR and the number of realised tasks would be higher if the workload limitation were not enforced.

Table 6.7 presents the mean $\pm$ standard deviation (MEAN $\pm$ SD) of the Realised Tasks (RT) metric. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count. These values support the previously mentioned findings. The VTAS and Hybrid-VTAS outperform the NN algorithm in terms of the number of realised tasks only in scenarios with 25 and 50 users—that is, when the workload limitation is not reached. In the other two cases tested, the NN algorithm surpasses all others in this metric, exhibiting the same behaviour observed in the TCR metric. Additionally, it can be noted that the values obtained for Hybrid-VTAS are slightly lower than those for VTAS, demonstrating that these two algorithms perform very similarly in terms of the number of tasks completed.

The left-hand side of Figure 6.4 illustrates the evolution of LFI (Load Fairness Index - Eq. 4.4) per minute for the first slot of the tested campaign using

¹A Moving Average (MA) is a simple way to analyse data trends over time by smoothing out short-term fluctuations. It works by calculating the average of a set of values over a specific period and then continuously updating this average as new data points are added. For example, in a 7-day moving average, you take the average of the last 7 days of data. The next day, you drop the oldest value and add the newest one, recalculating the average. This helps reveal overall trends by reducing the impact of daily ups and downs.

		MEAN $\pm$ SD (RT)			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	3.16 $\pm$ 3.38	3.16 $\pm$ 2.42	2.99$\pm$0.81	1.59$\pm$0.00
	LNN	3.16 $\pm$ 2.40	3.11 $\pm$ 1.08	2.37 $\pm$ 0.70	1.28 $\pm$ 0.56
	NN	5.06$\pm$0.99	3.83$\pm$0.77	2.64 $\pm$ 0.77	1.26 $\pm$ 0.48
	Bio-VTAS	3.16 $\pm$ 2.83	3.16 $\pm$ 1.71	2.69 $\pm$ 0.84	1.41 $\pm$ 0.51
	Hybrid-VTAS	3.16 $\pm$ 3.17	3.16 $\pm$ 2.26	3.03 $\pm$ 0.76	1.58 $\pm$ 0.62
Slot 2	VTAS	3.13 $\pm$ 3.27	3.13 $\pm$ 2.33	2.99$\pm$0.89	1.54 $\pm$ 0.58
	LNN	3.13 $\pm$ 2.44	3.09 $\pm$ 1.11	2.30 $\pm$ 0.73	1.24 $\pm$ 0.47
	NN	4.92$\pm$1.20	3.82$\pm$0.96	2.58 $\pm$ 0.73	1.29 $\pm$ 0.57
	Bio-VTAS	3.13 $\pm$ 2.46	3.13 $\pm$ 1.58	2.55 $\pm$ 0.84	1.44 $\pm$ 0.56
	Hybrid-VTAS	3.13 $\pm$ 2.97	3.13 $\pm$ 2.04	2.82 $\pm$ 0.79	1.55$\pm$0.56

Table 6.7: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different users counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each user count.

the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms. The results represent the mean of five executions. The first notable observation is the exponential behaviour of the NN algorithm across all campaigns. As noted in the previous chapter, the LFI of VTAS and Hybrid-VTAS remains minimal. Meanwhile, the LFI of Bio-VTAS and LNN exhibits similar behaviour, positioned between the LFI of VTAS and that of NN. This indicates that VTAS and Hybrid-VTAS maintain the lowest LFI, followed by Bio-VTAS, and then LNN.

The MEAN $\pm$ SD of this metric is presented in Table 6.8. The highlighted results indicate the algorithms with the lowest mean for each user count and slot. These findings reinforce previous observations: the NN algorithm consistently exhibits significantly higher mean LFI values due to its exponential behaviour, whereas the VTAS and Hybrid-VTAS algorithms maintain lower LFI values across all cases. In the previous chapter, an increase in the LFI of Bio-VTAS was observed in the second slot. A similar pattern emerges in this experiment with Hybrid-VTAS; however, in this case, the increase is less pronounced compared to Bio-VTAS. Moreover, the MEAN $\pm$ SD values associ-

		MEAN $\pm$ SD (LFI)			
		Number of Users			
		100 Users	75 Users	50 Users	25 Users
Slot 1	VTAS	0.22$\pm$0.22	0.24$\pm$0.17	0.32$\pm$0.06	0.26$\pm$0.05
	LNN	2.23 $\pm$ 1.99	3.83 $\pm$ 2.21	4.70 $\pm$ 2.27	3.20 $\pm$ 2.10
	NN	32.55 $\pm$ 26.57	16.23 $\pm$ 13.32	8.75 $\pm$ 7.05	3.64 $\pm$ 2.81
	Bio-VTAS	1.04 $\pm$ 0.95	1.52 $\pm$ 1.12	2.90 $\pm$ 1.55	3.29 $\pm$ 2.23
	Hybrid-VTAS	0.24 $\pm$ 0.24	0.28 $\pm$ 0.19	0.38 $\pm$ 0.13	0.33 $\pm$ 0.17
Slot 2	VTAS	0.22$\pm$0.23	0.25$\pm$0.18	0.36$\pm$0.09	0.28$\pm$0.07
	LNN	1.96 $\pm$ 1.87	3.54 $\pm$ 2.22	4.35 $\pm$ 2.22	3.08 $\pm$ 2.04
	NN	28.91 $\pm$ 25.09	15.12 $\pm$ 12.82	8.99 $\pm$ 7.40	3.64 $\pm$ 3.01
	Bio-VTAS	1.80 $\pm$ 1.71	2.74 $\pm$ 2.04	4.50 $\pm$ 2.50	4.13 $\pm$ 2.93
	Hybrid-VTAS	0.34 $\pm$ 0.33	0.69 $\pm$ 0.56	2.56 $\pm$ 1.38	2.96 $\pm$ 2.00

Table 6.8: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different user counts (100, 75, 50, and 25 Users). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each user count.

ated with Hybrid-VTAS are considerably lower, suggesting that Hybrid-VTAS achieves a more equitable distribution of tasks across the cells than Bio-VTAS. Additionally, this increase is only observed in scenarios where the workload limitation is not reached, representing an improvement over the results obtained with Bio-VTAS. Nonetheless, while the task distribution achieved by Hybrid-VTAS does not match the efficiency of VTAS, it is significantly closer to VTAS's performance than to that of Bio-VTAS.

Figure 6.5 illustrates the distribution of tasks among the cells at the end of the first slot of the campaign for Hybrid-VTAS with 25 users. The results display the number of realised and accepted tasks for each cell, with the completion colour scale developed to visualise task distribution. A significant dispersion of colours can be observed, indicating that Hybrid-VTAS does not achieve an entirely equitable distribution of tasks. However, the colour dispersion is less pronounced compared to that of Bio-VTAS (Figure 5.8), yet not as balanced as the distribution achieved by VTAS (Figure 4.8). While this chapter presents the visualisation for 25 users, the corresponding images for the other scenarios in this experiment can be found in Appendix E, Section

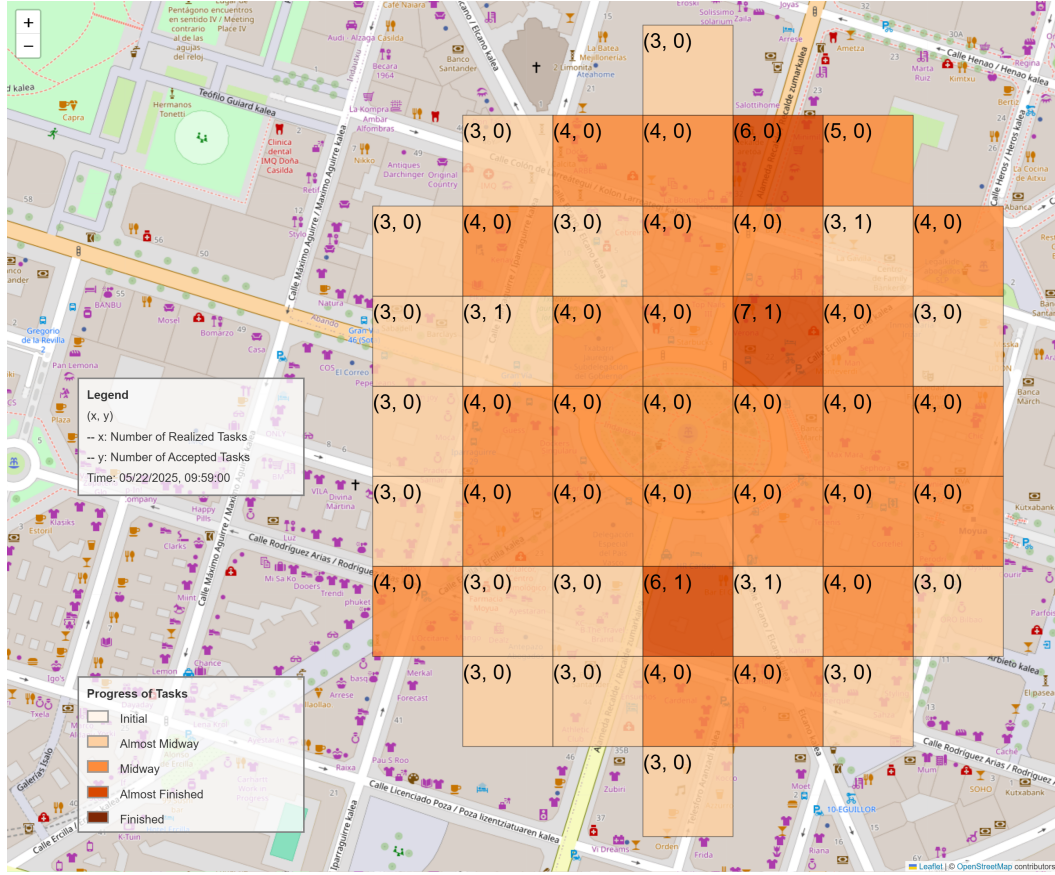


Figure 6.5: The task completion campaign involved 25 users using the Hybrid-VTAS algorithm. This image illustrates the distribution of tasks among the cells at the end of the first slot of the campaign.

E.1.

This section evaluates Hybrid-VTAS with different numbers of users and compares its performance to the Bio-VTAS, VTAS, LNN, and NN algorithms. Hybrid-VTAS achieves significantly better and more noteworthy results, with its performance closely resembling that of VTAS in terms of TCR and RT, in some cases even surpassing it. This confirms the effectiveness of Hybrid-VTAS in achieving Goal 2 across different user counts. A similar trend is observed in the LFI metric. While Hybrid-VTAS still experiences an increase in this metric during the second slot in certain cases, this increase only occurs when the workload limitation is not reached and is less pronounced than with Bio-VTAS. This collaborative solution therefore improves upon the results ob-

tained with Bio-VTAS while maintaining user inclusion within the algorithm. Additionally, as the LFI of Hybrid-VTAS achieves better results than both LNN and NN, we can conclude that Hybrid-VTAS successfully fulfils Goal 3.

6.3.2 Experiment 2: User Availability probability

This experiment evaluates the performance of the Hybrid-VTAS algorithm under different levels of users' motivation, defined by User Availability Probabilities of 75 % (high motivation), 50%, and 25% (low motivation). The study assesses the algorithm's effectiveness in achieving the task-completion goal (Goal 2) and its ability to ensure an equitable task distribution when full completion is not possible (Goal 3). The campaign parameters remained consistent with the previous section, with a fixed User Task Completion Probability of 75% and a total of 50 users. The Hybrid-VTAS was compared against VTAS, Bio-VTAS, LNN, and NN strategies. Each scenario was simulated five times per algorithm, and the results presented reflect the average outcomes.

		TCR		
		User available rate		
		75%	50%	25%
Slot 1	VTAS	100%	94.52%	70.27%
	LNN	91.7%	74.95%	48.62%
	NN	101.9%	83.40%	49.04%
	Bio-VTAS	97.50%	85%	59.15%
	Hybrid-VTAS	100%	95.96%	65.48%
Slot 2	VTAS	100%	95.43%	61.38%
	LNN	90.7%	73.46%	48.62%
	NN	102.2%	82.29%	46.33%
	Bio-VTAS	96.81%	81.38%	58.46%
	Hybrid-VTAS	99.68%	89.89%	61.49%

Table 6.9: TCR evaluation over the different user availability probabilities using the VTAS, Bio-VTAS, Hybrid-VTAS, LNN, and NN task allocation algorithms. The highlighted results correspond to the strategies with the highest TCR in each slot for each tested User Availability Probability.

Table 6.9 presents the TCR metric for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Availability Probab-

ilities. The highlighted results correspond to the algorithms with the highest TCR in each slot for each tested User Availability Probability. It was observed that VTAS, Bio-VTAS, Hybrid-VTAS, and LNN reach the workload limitation in scenarios with a 75% of User Availability Probability. Specifically, both Hybrid-VTAS and VTAS achieve 100% in the TCR metric. The NN algorithm, lacking a workload limitation, exceeds 100% TCR in the 75% probability scenario. This is the only case where NN outperforms all other algorithms, and, as previously mentioned, this is due to the workload limitations imposed on the other algorithms. Furthermore, given that this study is conducted within a citizen science context, Hybrid-VTAS's performance in the 75% probability case is considered as superior to that of LNN and NN. In the remaining two scenarios, VTAS or Hybrid-VTAS outperform all other algorithms. In all the cases, the Hybrid-VTAS' performance is very close to that of VTAS. This demonstrates the effectiveness of the Hybrid-VTAS algorithm in managing varying levels of User Availability Probability while also nearly matching the efficiency of VTAS in all scenarios.

The right-hand side of Figure 6.6 illustrates the Realised Tasks (RT- Eq. 4.3) throughout the first slot of the campaign using the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms. This graph demonstrates that, at a 75% probability, both Hybrid-VTAS and VTAS exhibit a decline towards the end of the slot, which can be attributed to reaching the workload limitation. This behaviour is confirmed by the TCR values presented in Table 6.9. Furthermore, given that this study is conducted within a citizen science context, Hybrid-VTAS's performance in the 75% probability case is considered as superior to that of LNN and NN. Moreover, in all other scenarios, it is evident that Hybrid-VTAS and VTAS consistently achieve superior performance compared to the other algorithms. Additionally, both algorithms perform at a very similar level.

Table 6.10 presents the $\text{MEAN} \pm \text{SD}$ of the Realised Tasks for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Availability Probabilities. The highlighted results correspond to the algorithms with the highest mean in each slot for each tested User Availability Probability. These values support the previously mentioned findings. In the case of

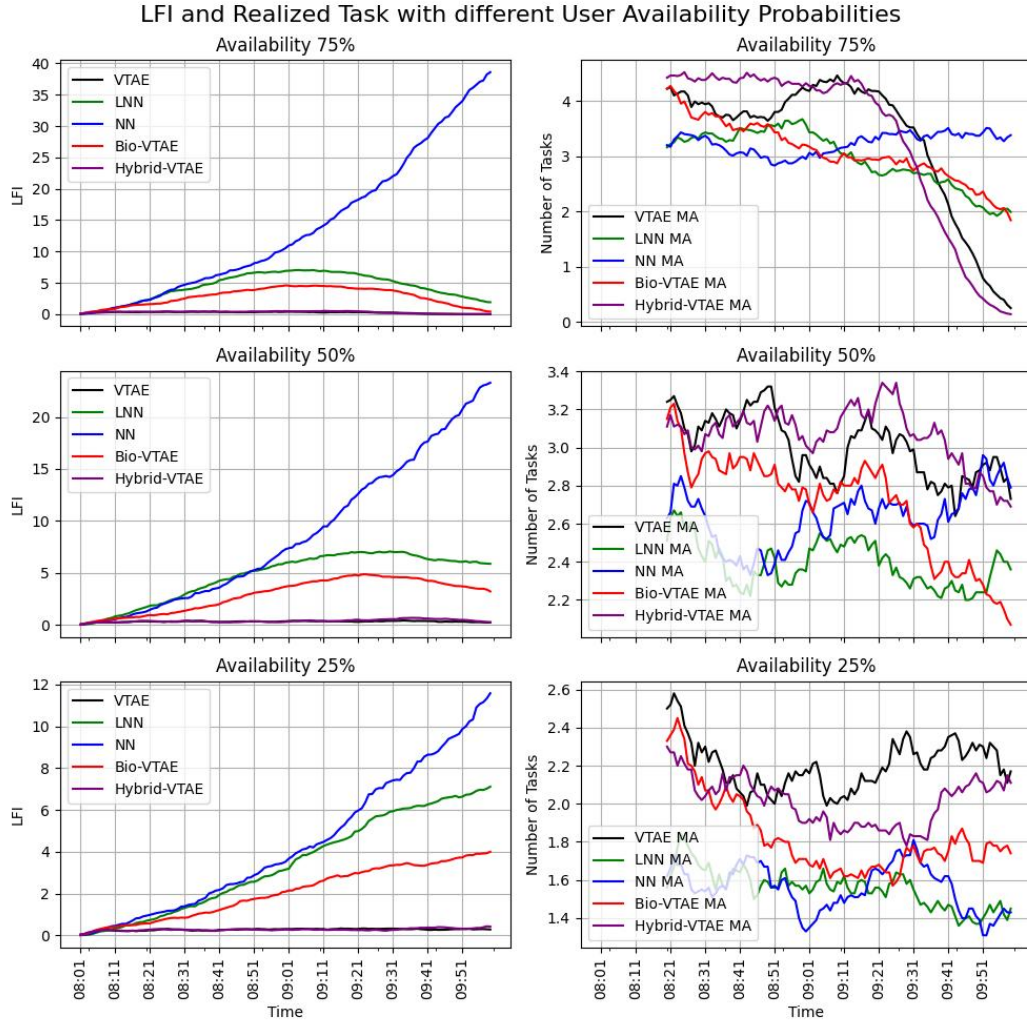


Figure 6.6: The LFI (Eq. 4.4 left figures) and the realised task (Eq. 4.3 right figures) over the first slot of the campaign across different User Availability Probabilities. The task allocation strategies used are the VTAS, Bio-VTAS, Hybrid-VTAS, LNN, and NN algorithms. To illustrate the evolution of the realised tasks, the moving average (MA) of 20 values was used. The results show the mean of five executions.

a 75% probability, NN outperforms all other algorithms in this metric. In all other scenarios, either Hybrid-VTAS or VTAS achieve superior performance compared to the other algorithms.

The left-hand side of Figure 6.6 shows the LFI (Eq. 4.4) metric for the first slot of the campaign, using the Hybrid-VTAS, Bio-VTAS, VTAS, LNN,

		MEAN $\pm$ SD (RT)		
		User available rate		
		75%	50%	25%
Slot 1	VTAS	3.16 $\pm$ 1.71	2.99 $\pm$ 0.81	2.22$\pm$0.81
	LNN	2.90 $\pm$ 0.91	2.37 $\pm$ 0.70	1.54 $\pm$ 0.57
	NN	3.22$\pm$0.83	2.64 $\pm$ 0.77	1.55 $\pm$ 0.56
	Bio-VTAS	3.08 $\pm$ 1.16	2.69 $\pm$ 0.84	1.87 $\pm$ 0.70
	Hybrid-VTAS	3.16 $\pm$ 1.88	3.03$\pm$0.76	2.07 $\pm$ 0.72
Slot 2	VTAS	3.13 $\pm$ 1.78	2.99$\pm$0.89	1.92 $\pm$ 0.67
	LNN	2.84 $\pm$ 0.96	2.30 $\pm$ 0.73	1.52 $\pm$ 0.52
	NN	3.20$\pm$0.85	2.58 $\pm$ 0.73	1.45 $\pm$ 0.51
	Bio-VTAS	3.03 $\pm$ 1.07	2.55 $\pm$ 0.84	1.83 $\pm$ 0.65
	Hybrid-VTAS	3.12 $\pm$ 1.46	2.82 $\pm$ 0.79	1.93$\pm$0.60

Table 6.10: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Availability Probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Availability Probability.

and NN algorithms. The first observation is that NN continues to exhibit exponential behaviour. The Hybrid-VTAS algorithm demonstrates a behaviour very similar to that of VTAS, as its LFI remains minimal.

Table 6.11 presents the MEAN $\pm$ SD of the LFI metric for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms under different User Availability Probabilities. The highlighted results correspond to the algorithms with the lowest mean in each slot for each User Availability Probability. These values support the previously mentioned findings. VTAS achieves the best LFI in all scenarios, as observed in previous experiments. In contrast, NN exhibits a very high LFI due to its exponential behaviour. In the first slot, the Hybrid-VTAS values closely align with those of VTAS. However, in the second slot, the Hybrid-VTAS values are slightly higher. This trend was also observed in the previous experiment. Despite this increase, Hybrid-VTAS remains superior to Bio-VTAS and LNN in all scenarios.

The visual representation of the LFI for all tested probabilities and algorithms is provided in Appendix E, Section E.2. These examples have been excluded from this section, as the results are consistent with those obtained

		MEAN $\pm$ SD (LFI)		
		User available rate		
		75%	50%	25%
Slot 1	VTAS	0.26$\pm$0.13	0.32$\pm$0.06	0.28$\pm$0.05
	LNN	4.38 $\pm$ 2.13	4.70 $\pm$ 2.27	3.44 $\pm$ 2.33
	NN	13.80 $\pm$ 11.34	8.75 $\pm$ 7.05	4.37 $\pm$ 3.34
	Bio-VTAS	2.77 $\pm$ 1.41	2.90 $\pm$ 1.55	2.07 $\pm$ 1.23
	Hybrid-VTAS	0.31 $\pm$ 0.17	0.38 $\pm$ 0.13	0.28$\pm$0.06
Slot 2	VTAS	0.26$\pm$0.14	0.36$\pm$0.09	0.33$\pm$0.07
	LNN	4.51 $\pm$ 2.23	4.35 $\pm$ 2.22	3.33 $\pm$ 2.24
	NN	12.31 $\pm$ 10.25	8.99 $\pm$ 7.40	3.88 $\pm$ 3.11
	Bio-VTAS	4.27 $\pm$ 2.34	4.50 $\pm$ 2.50	3.84 $\pm$ 2.24
	Hybrid-VTAS	2.33 $\pm$ 1.65	2.56 $\pm$ 1.38	1.01 $\pm$ 0.63

Table 6.11: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Availability Probabilities (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the lowest mean in each slot for each tested User Availability Probability.

in the previous experiment, and the visual representation of the LFI is considered unnecessary. Nevertheless, the reader is encouraged to consult the appendix for a more thorough and complete overview of the results obtained in this experiment.

The results of Experiment 2 demonstrate the efficacy of the Hybrid-VTAS algorithm in managing varying levels of User Availability Probability. Hybrid-VTAS consistently outperforms both the LNN and NN algorithms in terms of TCR and Realised Tasks (RT), except in the scenario where the User Availability Probability is 75%. In this case, the NN algorithm surpasses all other results due to the absence of workload constraints. However, adhering to workload restrictions is considered more beneficial than achieving higher TCR and RT metrics in a Citizen Science context. Therefore, Hybrid-VTAS demonstrates superior performance in terms of compliance with workload limitations. These findings confirm that Hybrid-VTAS successfully achieves Goal 2 across varying User Availability Probabilities. Furthermore, the performance of these algorithms in both metrics is very similar to that of VTAS, confirming their overall effectiveness. Additionally, Hybrid-VTAS maintains a low Load

Fairness Index (LFI) across all tested probabilities, indicating a equitable distribution of tasks. In contrast, the NN algorithm exhibits a significantly higher LFI. This further affirms the effectiveness of VTAS with respect to Goal 3. Although Hybrid-VTAS experiences a slight increase in the LFI metric in the second slot, this increase is not as substantial as that observed with Bio-VTAS. Thus, it is clear that Hybrid-VTAS not only maintains this metric but also mitigates the negative effects seen in Bio-VTAS. In conclusion, these findings underscore the robustness of the Hybrid-VTAS algorithm in ensuring equitable task allocation and maintaining high performance, even under challenging conditions of user availability.

6.3.3 Experiment 3: User Task Completion Probability

This experiment evaluates the performance of the Hybrid-VTAS algorithm under different levels of user commitment, defined by User Task Completion Probabilities of 75% (high commitment), 50%, and 25% (low commitment). The aim is to assess how effectively Hybrid-VTAS achieves the task-completion goal (Goal 2) at varying commitment levels and its ability to ensure fair task distribution when the goal is not fully achievable (Goal 3). The campaign configuration and parameters remained consistent, with a User Availability Probability of 50% and 50 users. The Hybrid-VTAS algorithm was compared against baseline models (LNN, NN, VTAS, and Bio-VTAS). Each scenario was simulated five times, with the results representing the average outcomes.

The Task Completion Rate (TCR) metric is presented in Table 6.12. The first observation is that none of the algorithms reach the workload limit, although Hybrid-VTAS and VTAS come very close. Additionally, in all scenarios, Hybrid-VTAS and VTAS achieve the highest TCR across the three cases. The results demonstrate the effectiveness of Hybrid-VTAS in task completion when compared to the other algorithms.

The right-hand side of Figure 6.7 shows the Realised Tasks for the first slot of the campaign using the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms. As in the other experiments, Hybrid-VTAS and VTAS demonstrate superior performance in task completion compared to the other algorithms.

		TCR		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	94.52%	66.22%	34.41%
	LNN	74.95%	51.33%	25.05%
	NN	83.40%	49.10%	23.62%
	Bio-VTAS	85%	61.6%	27.29%
	Hybrid-VTAS	95.96%	64.63%	32.29%
Slot 2	VTAS	95.43%	62.45%	31.33%
	LNN	73.46%	52.45%	24.63%
	NN	82.29%	50.43%	26.22%
	Bio-VTAS	81.38%	56.6%	26.17%
	Hybrid-VTAS	89.89%	60.05%	32.13%

Table 6.12: TCR evaluation over the different User Task Completion Probability using the VTAS, Bio-VTAS, Hybrid-VTAS LNN, and NN.

Table 6.13 presents the MEAN $\pm$ SD of the Realised Tasks for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probabilities. The highlighted results correspond to the algorithms with the highest mean in each slot for each User Task Completion Probability. This table demonstrates that Hybrid-VTAS and VTAS exhibit better performance in realised tasks compared to the other algorithms. The results obtained by VTAS are very comparable to those of Hybrid-VTAS, as shown in the figure.

The left-hand side of Figure 6.7 illustrates the LFI metric (Eq. 4.4) throughout the first slot of the campaign for various probabilities tested with the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms. The NN exhibits exponential behaviour, as seen in previous cases. In contrast, the LFI of both VTAS and Hybrid-VTAS remains very low across all scenarios, with the performance of both algorithms being indistinguishable.

Table 6.14 presents the MEAN $\pm$ SD of the LFI metric for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms under different User Task Completion Probabilities. The highlighted results correspond to the algorithms with the lowest mean LFI in each slot for each probability. These values further support the previous observations. Indeed, the results confirm

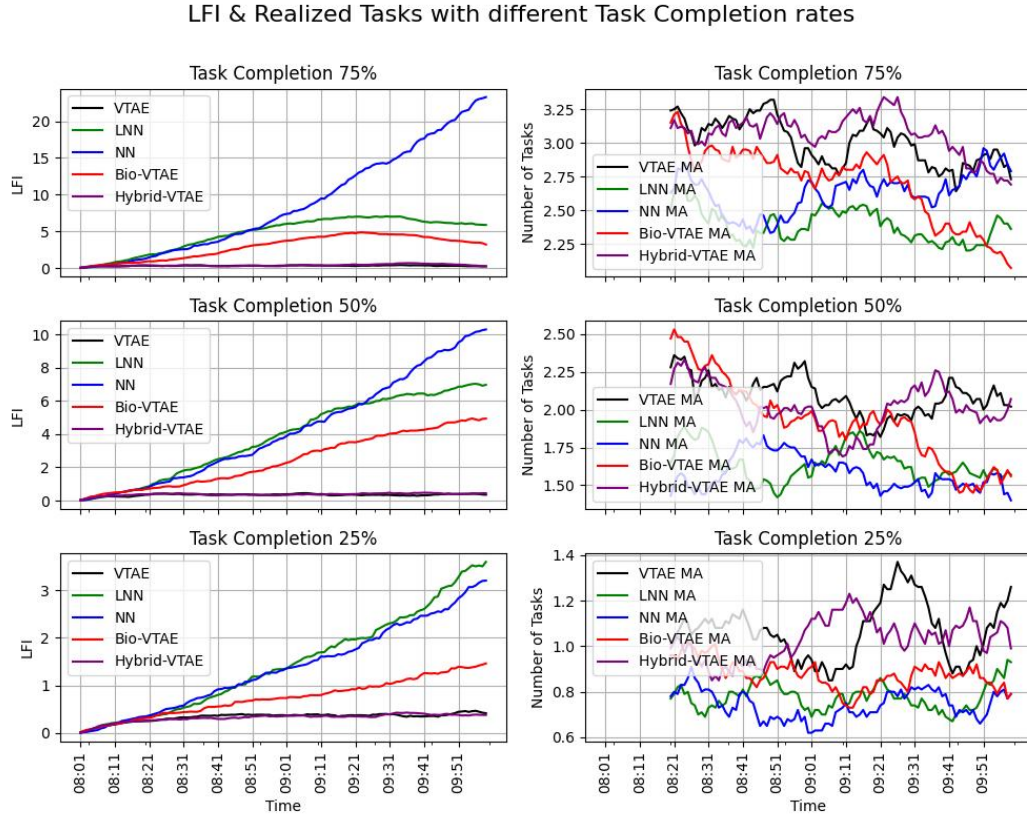


Figure 6.7: The LFI (Eq. 4.4 - left figures) and the Realised Task (Eq. 4.3 - right figures) over the first slot of the campaign across different task completion probabilities. The task allocation strategies used are the VTAS, Bio-VTAS, Hybrid-VTAS, the LNN, and the NN algorithms. To illustrate the evolution of the Realised Task, the moving average of 20 values was used. The results show the mean of five executions.

that VTAS and Hybrid-VTAS are the algorithms with the most favourable outcomes in the LFI metric. Similarly, it is observed that in the second slot, the values for Hybrid-VTAS increase, although to a lesser extent compared to Bio-VTAS.

The visual representation of the LFI for all tested probabilities and algorithms is provided in Appendix E, Section E.3. These examples have been excluded from this section, as the results are consistent with those obtained in the previous experiments, and the visual representation of the LFI is considered unnecessary. Nevertheless, the reader is encouraged to consult the

		MEAN $\pm$ SD (RT)		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	2.99 $\pm$ 0.81	2.09$\pm$0.75	1.09$\pm$0.48
	LNN	2.37 $\pm$ 0.70	1.62 $\pm$ 0.56	0.79 $\pm$ 0.37
	NN	2.64 $\pm$ 0.77	1.55 $\pm$ 0.57	0.75 $\pm$ 0.39
	Bio-VTAS	2.69 $\pm$ 0.84	1.95 $\pm$ 0.73	0.88 $\pm$ 0.42
	Hybrid-VTAS	3.03$\pm$0.76	2.04 $\pm$ 0.64	1.02 $\pm$ 0.55
Slot 2	VTAS	2.99$\pm$0.89	1.96$\pm$0.69	0.98 $\pm$ 0.50
	LNN	2.30 $\pm$ 0.73	1.64 $\pm$ 0.58	0.77 $\pm$ 0.39
	NN	2.58 $\pm$ 0.73	1.58 $\pm$ 0.61	0.82 $\pm$ 0.36
	Bio-VTAS	2.55 $\pm$ 0.84	1.77 $\pm$ 0.64	0.83 $\pm$ 0.40
	Hybrid-VTAS	2.82 $\pm$ 0.79	1.88 $\pm$ 0.62	1.01$\pm$0.57

Table 6.13: The MEAN $\pm$ SD of the Realised tasks (Eq. 4.3) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted results correspond to the algorithms with the highest mean in each slot for each Task Completion Probability.

		MEAN $\pm$ SD (LFI)		
		Task completion probability		
		75%	50%	25%
Slot 1	VTAS	0.32$\pm$0.06	0.34$\pm$0.07	0.33 $\pm$ 0.09
	LNN	4.70 $\pm$ 2.27	3.84 $\pm$ 2.35	1.48 $\pm$ 1.04
	NN	8.75 $\pm$ 7.05	4.26 $\pm$ 3.15	1.38 $\pm$ 0.91
	Bio-VTAS	2.90 $\pm$ 1.55	2.43 $\pm$ 1.57	0.74 $\pm$ 0.39
	Hybrid-VTAS	0.38 $\pm$ 0.13	0.37 $\pm$ 0.08	0.32$\pm$0.08
Slot 2	VTAS	0.36$\pm$0.09	0.38$\pm$0.10	0.38$\pm$0.12
	LNN	4.35 $\pm$ 2.22	3.65 $\pm$ 2.35	1.36 $\pm$ 0.93
	NN	8.99 $\pm$ 7.40	4.62 $\pm$ 3.51	1.69 $\pm$ 1.20
	Bio-VTAS	4.50 $\pm$ 2.50	3.83 $\pm$ 2.48	1.54 $\pm$ 1.03
	Hybrid-VTAS	2.56 $\pm$ 1.38	1.24 $\pm$ 0.86	0.41 $\pm$ 0.16

Table 6.14: The MEAN $\pm$ SD of the LFI metric (Eq. 4.4) for the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms across different User Task Completion Probability (75%, 50% and 25%). The results are the mean of five executions. The highlighted values correspond to the algorithms with the lowest mean in each slot for each User Task Completion Probability.

appendix for a more thorough and complete overview of the results obtained in this experiment.

In summary, in this experiment, the User Task Completion Probability variable was tested to understand how user commitment to their accepted tasks affects the effectiveness of the Hybrid-VTAS. A high User Task Completion Probability indicates strong user commitment, leading to more realised tasks. The study evaluated probabilities of 75%, 50%, and 25% using the Hybrid-VTAS, Bio-VTAS, VTAS, LNN, and NN algorithms, with 50 users and a User Availability Probability of 50%. This experiment confirms that the Hybrid-VTAS exhibits behaviour very similar to that of the VTAS. In this case, both VTAS and Hybrid-VTAS are the algorithms with the best results in the RT and TCR metrics (Goal 2). It can also be observed that the performance of the Hybrid-VTAS in the LFI metric is very similar to that of the VTAS. Although it does not surpass the VTAS, the results are very promising and further confirm that the Hybrid-VTAS achieves Goal 3. It is noted that in the second slot, the values for Hybrid-VTAS increase, although to a lesser extent compared to the Bio-VTAS. This demonstrates that the ensemble algorithm enables the VTAS to regularise the behaviour of the Bio-VTAS, still maintaining its advantages

6.3.4 Conclusions of the evaluation of the Hybrid-VTAS Algorithm in the Simulation Use Case

In this chapter, the Hybrid-VTAS strategy is evaluated through three different experiments. In the first experiment, the strategy is tested with different population sizes, assessing its performance under varying user efforts. The second experiment evaluates the algorithm with different User Availability Probabilities, testing the Hybrid-VTAS against different levels of user motivation. In the third experiment, the algorithm is analysed with different User Task Completion Probabilities, challenging the Hybrid-VTAS under varying levels of user commitment.

The results of the three experiments were consistent. The Hybrid-VTAS outperformed both the NN and LNN in all metrics when the workload lim-

itation was not present. In scenarios where this limitation prevented the completion of additional tasks in the Hybrid-VTAS, the NN achieved better performance in terms of TCR and realised tasks (RT). However, it was demonstrated that in the absence of this limitation, the Hybrid-VTAS would outperform the NN. As mentioned in the experiments, it is preferable to respect the workload limitation, even if it results in lower performance in the TCR and Realised Tasks (RT) metrics. In our context of Citizen Science, it is vital to ensure that users do not feel overwhelmed, making the limitation of workload a key consideration. This is also a fundamental aspect of achieving Goal 2. Therefore, even in cases where the Hybrid-VTAS does not achieve the best values in TCR and realised tasks (RT)-i.e., when the workload limitation is involved-its behaviour remains superior to the baselines. In the remaining cases, the Hybrid-VTAS showed the better performance in these metrics respecting the TCR and RT metric versus the LNN and the NN. Thus, the effectiveness of the Hybrid-VTAS in all the experiments regarding Goal 2 is demonstrated. Respecting the effectiveness of the Hybrid-VTAS regarding Goal 2, the Hybrid-VTAS showed the best results alongside the VTAS, outperforming the Bio-VTAS in all cases.

With respect to Goal 3, this is measured using the LFI metric. In all the evaluated cases, the LFI of the Hybrid-VTAS was lower than that of the baselines LNN and NN, demonstrating that the Hybrid-VTAS maintains an equitable distribution of tasks across cells. In contrast, both the NN and the LNN exhibited exponential behaviour in LFI, indicating an inequitable distribution of tasks. Consequently, the Hybrid-VTAS meets Goal 3 in all the evaluated cases. Hence, the Hybrid-VTAS showed the best results alongside the VTAS, outperforming the Bio-VTAS in all cases. In the second slot, the Hybrid-VTAS achieves higher results than those obtained in the first slot. However, this increase is not as significant as that observed with the Bio-VTAS. Additionally, since the VTAS achieves the lowest results in the LFI metric, the second-best algorithm is observed to be the Hybrid-VTAS.

Overall, the Hybrid-VTAS algorithm successfully combined the best characteristics of both models, VTAS and Bio-VTAS. Based on the conducted

experiments, it demonstrated superior effectiveness in ensuring high performance and equitable task allocation across different scenarios, achieving in all of them the Goal 2 and Goal 3. It consistently met task-completion goals while maintaining an equitable workload distribution, proving to be a reliable solution for Spatial Crowdsourcing applications in a Citizen Science context. The algorithm's adaptability to varying crowd sizes, user availability, and task-completion probabilities highlights its versatility and robustness, making it well-suited to address real-world challenges.

6.4 Summary and Conclusions

This chapter presented the Hybrid Volunteer Task Allocation Strategy (Hybrid-VTAS), a novel solution designed to address the misalignment between Spatial Crowdsourcing (SC) and Citizen Science (CS). The Hybrid-VTAS algorithm integrates the strengths of the VTAS and Bio-VTAS algorithms to achieve three primary goals: Goal 1 enhancing user perception, Goal 2 optimizing task allocation, and Goal 3 creating an equitable distribution of tasks across all cells. Hence, it balances the performance of both VTAS (more campaign-oriented approach) and the Bio-VTAS (more user-oriented approach).

The chapter began with an analysis of face-to-face survey interviews conducted with participants from the Deusto Use Case. The insights gained from analysing the testimonies gathered during the interviews were used to distill the user requirements for the Hybrid-VTAS strategy, ensuring that it aligns with users' preferences and interaction habits. The Hybrid-VTAS algorithm was then evaluated through simulations, as real-world experimentation was not feasible due to time constraints.

The evaluation consisted of three experiments. The first experiment assessed the algorithm's performance with different crowd sizes, demonstrating that Hybrid-VTAS maintains high Task Completion Rates (TCR) and Realised Tasks (RT) metrics, even with fewer users. The second experiment evaluated the algorithm under varying User Availability Probabilities, showing that Hybrid-VTAS adapts effectively to different levels of user availability while maintaining a low Load Fairness Index (LFI). The third experiment tested

the algorithm with different User Task Completion Probabilities, confirming its robustness in handling varying levels of user commitment.

The results consistently show that the Hybrid-VTAS outperforms NN and LNN across all metrics when workload limitations are not present. While NN achieves higher task completion rates (TCR) and realised tasks (RT) in constrained scenarios, adhering to workload limits is crucial in Citizen Science to prevent user overload. Thus, even when limited, the Hybrid-VTAS maintains superior overall performance, proving its effectiveness in achieving Goal 2. Compared to VTAS and Bio-VTAS, the Hybrid-VTAS performs on par with VTAS, consistently surpassing Bio-VTAS. Regarding Goal 3, measured by the Load Fairness Index (LFI), the Hybrid-VTAS consistently achieves lower LFI than NN and LNN, ensuring equitable task distribution. While VTAS exhibits the lowest LFI, Hybrid-VTAS emerges as the second-best performer, significantly outperforming Bio-VTAS. Overall, the Hybrid-VTAS proves to be a robust and adaptable solution for Spatial Crowdsourcing in Citizen Science, excelling in both high performance and fair task allocation.

The effectiveness of this solution in terms of user perception has not yet been tested in a real-world scenario. However, as it is an ensemble of two algorithms that have been validated and shown to achieve Goal 1, the Hybrid-VTAS is expected to maintain high levels of user acceptance and adoption. While this approach does not exceed the VTAS in terms of performance and task distribution, its results remain highly comparable. Moreover, the Hybrid-VTAS incorporates user participation by accounting for their decision-making processes and routines, enabling personalised task recommendations. Given that it also considers user behaviour patterns it is expected to achieve a higher acceptance and adoption than the VTAS strategy. Therefore, it can be reasonably concluded that this algorithm effectively and efficiently fulfils Goal 1. This results in an algorithm that combines the strengths of the versions presented in the previous two chapters.

In conclusion, the Hybrid-VTAS algorithm demonstrates significant improvements over the baseline models in terms of task completion and equitable distribution. It effectively handles varying crowd sizes, user availability,

and task completion probabilities, making it a robust solution for task allocation in Spatial Crowdsourcing scenarios. The results confirm the algorithm's ability to achieve the defined goals and highlight its potential for practical applications in real-world campaigns.

The main takeaway is that Hybrid-VTAS can be considered as the best strategy of the 3 described ones, since it addresses Goal 1, Goal 2 and Goal 3 in a more holistic manner. It might not be the best in any of the goals in particular, but it is the ones that achieves the best balance between the campaign's needs and the users' acceptance.

*Discovery consists of seeing what everybody has seen
and thinking what nobody has thought.*

Albert Szent-Györgyi

CHAPTER

7

Conclusions and Future Work

This chapter seeks to provide an overview of the described work and the contributions this dissertation has generated. Therefore, the hypothesis and objectives introduced in Chapter 1 are analysed to evaluate to what extent they have been achieved. Some of the results of this research have been disseminated to validate their contribution with the research community. That is, the accomplishment of those objectives are objectively validated by the publications reviewed in this chapter. To conclude with this dissertation, future lines of work and ideas are proposed to continue the presented research line, and some final remarks in the form of discussion are presented.

The rest of the chapter is structured as follows: Section 7.1 summarizes the work done in this dissertation and the conclusions obtained throughout it, Section 7.2 lists the contributions from this work, Section 7.3 validates the hypothesis and the objectives that were set in Chapter 1, and Section 7.4 details the scientific publications obtained during the development of this dissertation. To finish with, Section 7.5 introduces future work, while Section 7.6 draws some final remarks.

7.1 Summary of work and conclusions

This thesis has addressed the problem of Spatial Crowdsourcing within the context of Citizen Science, proposing three innovative solutions to optimise task distribution and enhance user participation in such initiatives. The inherent challenges in aligning these two domains, coupled with the voluntary and unpredictable nature of participation in Citizen Science, have led to the quest for solutions that not only manage task allocation efficiently but also respect user limitations and preferences for promoting higher participation without relying on financial incentives. This research has proposed three iterative solutions, developed in a continuous improvement cycle: VTAS, Bio-VTAS, and Hybrid-VTAS.

To test the hypothesis of this work—*Human-AI collaboration can improve user acceptance and adoption of Spatial Crowdsourcing task allocation recommendations within Citizen Science by balancing efficient task completion procedures with a positive user perception, without relying on economic incentives*.—three key goals were established for the proposed solutions: Goal 1 enhancing user perception and engagement with the system, Goal 2 optimising task allocation, and Goal 3 ensuring equitable task distribution across spatial cells. These goals formed the foundation for evaluating the proposed solutions and verify the hypothesis of this dissertation. The assessment was then conducted using two distinct use cases: a simulated use case and a real-world experiment, referred to as the Deusto use case. The simulated use case enabled a controlled evaluation of system performance under predefined conditions, ensuring a rigorous analysis of the proposed spatial task allocation strategies. In contrast, the Deusto use case provided valuable insights into the real-world applicability of these solutions, offering a practical perspective on their effectiveness within a Citizen Science context. To evaluate the proposed solutions, the results from the three defined goals were compared against baseline strategies, namely Limited Nearest Neighbours (LNN) and Nearest Neighbour (NN), within the defined use cases.

The first spatial task allocation strategy, named VTAS, is effective in achieving equitable task distribution and maximising task completion. The

results demonstrate that users exhibit high acceptance of the system (Quantitative results are available in Table 4.19); however, it has been identified that its reliance on proximity as the primary parameter for personalisation limits the ability of the system to adapt to individual user routines. This limitation respecting the Goal 1 highlights the demand for integrating more specific user considerations into the spatial task allocation process.

The second solution, Bio-VTAS, improves personalisation by adapting to users' participation routines, thereby increasing their positive perception and engagement with the system. This improvement is reflected in the fact that users, according to the interviews conducted, no longer demand as many improvements as happen with the first strategy. Such a shift suggests that users are becoming more comfortable and familiar with the spatial task allocation system, even though the acceptance and adoption quantitative values remain quite similar compared to the results obtained with the VTAS (comparable quantitative results between the two strategies are available in Table 5.21). It is a positive indicator of the effectiveness of the proposed solutions in enhancing user experience and engagement. However, despite improving user perception Goal 1, Bio-VTAS does not surpass VTAS in relation to the objectives concerning task allocation Goal 2 and equitable distribution (Goal 3) (see Section 5.2). While Bio-VTAS offers improvements in terms of personalisation (Goal 1), its performance in efficiency (references to Goal 2 and Goal 3) is not superior to VTAS.

The final solution, Hybrid-VTAS, combines the strengths of both VTAS and Bio-VTAS. This integrated approach aims not only to enhance efficiency and ensure equitable task distribution in cells but also to foster more active user participation by adapting to their preferences and participation habits. This solution devised as result of the analysis of the one-to-one interviews conducted with users to understand how they wished to interact with this type of technology (see the results of the interviews at Section 6.1). The insights gained from their responses were instrumental in designing spatial task allocation strategy that aligns with user expectations (see Section 6.2). This strategy has been evaluated using the simulation use case, where Hybrid-VTAS demonstrated results very similar to VTAS concerning Goal 2 and

Goal 3, achieving an equitable task distribution and a high task completion rate. This confirms that the solution remains user-centric while nearly matching VTAS's efficiency. While further validation under real-world conditions would strengthen these findings, as an ensemble strategy, it is expected to perform similarly to VTAS, with additional improvements in user perception. In summary, this strategy achieve the best balance between the three goals, offering a well-rounded approach to spatial task allocation in Citizen Science campaigns.

Additionally, the first two strategies, VTAS and Bio-VTAS, were evaluated within the framework of the European SOCIO-BEE project. Although these evaluations are not formally integrated into this dissertation due to project's and dissertation's differences in methodologies and objectives, they serve as a further validation of the proposed approaches in real-world Citizen Science initiatives. The long-term nature of these external evaluations also highlights the potential sustainability and adaptability of the solutions beyond controlled testing environments.

The results presented throughout this dissertation indicate that all three proposed solutions successfully achieve the defined goals, thereby supporting the central hypothesis of this research. In the simulation, VTAS consistently delivers the best results for Goal 2 and Goal 3. However, its performance concerning Goal 1 is not as strong. Moreover, this solution is not truly user-centric, as it relies solely on user mechanics without offering personalization beyond this parameter. Interviews reveal that users have a strong need for system improvements. On the other hand, Bio-VTAS, while demonstrating a more refined and consistent response to Goal 1-enhancing user perception and engagement-does not surpass VTAS in terms of overall efficiency. Although Bio-VTAS offers stronger personalisation, its task allocation performance remains slightly inferior to that of VTAS. Hybrid-VTAS presents the best balanced approach, integrating the strengths of both VTAS and Bio-VTAS. While it does not fully surpass VTAS in efficiency, it comes close to achieving Goal 2 and Goal 3. Preliminary findings indicate Hybrid-VTAS may match or even exceed VTAS's performance in certain areas—particularly

user engagement and adaptability (Goal 1). Further validation through extended real-world trials will clarify its long-term effectiveness in dynamic Citizen Science contexts. Therefore, the third and final strategy presented in this dissertation not only validates the proposed hypothesis but also achieves the optimal balance between the defined objectives, offering the best compromise between efficiency and user perception.

Despite the progress made, several limitations remain that should be addressed in future research. The long term adoption of the presented solutions should be evaluated in real-world environments with users over extended periods of time to validate their effectiveness in real-world Citizen Science campaigns, where spatial crowdsourcing is a core demand. Additionally, further exploration is needed to integrate better user routines and preferences, considering factors such as dynamic change of location, time availability, and individual schedules. Furthermore, the scalability of these solutions must be assessed in larger campaigns, evaluating their performance in broader contexts with a larger number of participants. Finally, there is a need to investigate more effective mechanisms for user engagement and motivation to maintain user commitment in Citizen Science campaigns in the long run.

In conclusion, this thesis presents a robust framework and practical solutions to address the challenges of the Spatial Crowdsourcing core phase within Citizen Science campaigns. The proposed solutions, which prioritise active user participation, offer an innovative approach to balancing task allocation efficiency with respect for user preferences and desires. These contributions lay the foundation for the development of future sustainable and efficient Citizen Science initiatives, while opening new avenues for research to address current limitations and expand the applicability of these solutions to other contexts. Future research directions, based on the findings of this thesis, include the deployment of these spatial task allocation strategies in larger and more diverse environments, involving real users over extended periods to validate their long-term effectiveness. Moreover, improving the adaptability of the solutions to the variability of user behaviour, particularly in campaigns with less consistent (more erratic) participants, will be crucial. Additionally, exploring how these solutions can be applied to large-scale Citizen Science

projects with thousands of geographically distributed users, covering wide geographical areas, represents another important avenue for future work. Ultimately, this thesis sets the groundwork and provides the instruments for creating more human-centric Citizen Science campaigns, encountering a good balance between user engagement and responsiveness and the spatial crowdsourcing problem objectives and restrictions.

7.2 Contributions

As illustrated in previous chapters of this dissertation, the knowledge and findings gained from this thesis are diverse. The contributions achieved are categorised into two types: scientific contributions and technical contributions, as outlined below.

7.2.1 Scientific Contributions

- (C1) *Examination of existing challenges and advancements in Spatial Crowdsourcing within the context of Citizen Science, highlighting the limitations of current approaches.* This contribution is related to **O1** and corresponds to Sections 2.1 and 2.2 of Chapter 2. It provides an overview of the current state of Spatial Crowdsourcing in Citizen Science, focusing on the task allocation problem and the strategies employed. The analysis identifies key limitations of existing methods and explores the potential of Human-AI Collaboration in addressing these challenges more effectively.
- (C2) *Analysis of Human-AI Collaboration workflows in the literature.* This contribution is related to **O2** and corresponds to Section 2.3 of Chapter 2. It examines the workflows of existing Human-AI Collaboration systems, identifying key components and their interactions. The analysis facilitates the categorisation of different collaborative system architectures and provides a framework for understanding the essential elements and dynamics of such systems. These insights lay the foundation for

designing a collaborative system that effectively integrates users into the decision-making process.

- (C3) *Analysis of Human-AI Collaboration design principles.* This contribution is related to **O3** and corresponds to Section 2.3 of Chapter 2. It presents an in-depth investigation of the design principles required to develop an effective Human-AI collaborative system. The study emphasises the role of user adaptation as a critical factor in integrating individuals into the decision-making loop.
- (C4) *Definition of a Spatial Crowdsourcing model.* This contribution is related to **O4** and corresponds to Chapter 3. The model defines spatial and temporal constraints, task completion goals, and workload limitations within Citizen Science experiments. It integrates user limitations, as required by the Citizen Science field, into the Spatial Crowdsourcing problem, facilitating a collaborative framework that merges both domains.
- (C5) *Design and implementation of a collaborative spatial task allocation recommender system.* This contribution is related to **O5**. A Human-AI collaborative task allocation system is developed to balance user engagement and acceptance while optimising task completion across spatial areas divided into cells. Three primary spatial crowdsourcing strategies are proposed to address the spatial task allocation problem in Citizen Science through collaboration:
 - (C5.1) *Volunteers Task Allocation Strategy (VTAS) Algorithm.* This contribution is related to **O5** and corresponds to Chapter 4. It introduces an initial solution that balances task completion requirements with user experience by distributing tasks evenly across spatial areas. The VTAS allows participants to manage incomplete tasks, customise assignments, and reject tasks.
 - (C5.2) *Bio-Inspired Volunteers Task Allocation Strategy (Bio-VTAS) Algorithm.* This contribution is related to **O5** and corresponds to

Chapter 5. It extends the VTAS algorithm by incorporating user behaviour trajectories to generate adaptive task recommendations. This user-centre refinement enhances alignment between task assignments and user crowdsourcing patterns, improving engagement.

(C5.3) *Hybrid Volunteers Task Allocation Strategy (Hybrid-VTAS) Algorithm.* This contribution is related to **O5** and corresponds to Chapter 6. The Hybrid-VTAS integrates insights from previous versions to optimise task completion goals while considering user preferences. This solution brings together the best features of the previous two spatial task allocation strategies.

(C6) *Evaluation of the proposed solutions.* This contribution is related to **O6** and concerns the assessment of the three proposed solutions. Based on the hypothesis, three goals were established to determine their effectiveness in the context of Citizen Science. Several metrics (TCR - Eq. 4.2, RT - Eq. 4.3 and LFI - Eq. 4.4) and baselines (LNN and NN explained at Subsection 4.2.1) were selected to evaluate these goals and, consequently, the hypothesis itself. The goals are outlined in Section 4.1 of Chapter 4, while the explanation of the metrics, baselines, and use cases is provided in Section 4.2 of the same chapter. The evaluation includes:

- The VTAS (Chapter 4) was assessed using both a simulated use case (Section 4.3) and a real-world use case (Section 4.4).
- The Bio-VTAS (Chapter 5) was evaluated similarly, with a simulated use case (Section 5.2) and a real-world use case (Section 5.3).
- The Hybrid-VTAS (Chapter 6) was evaluated in a simulated use case (Section 6.3).

7.2.2 Technical Contributions

- (T1.) *Design and implementation of a collaborative spatial task allocation recommender system.* The three spatial task allocation strategies' implementations are publicly available in a GitHub repository¹.
- (T2.) *Simulation-based task allocation engine.* This contribution is related to **O6** and corresponds to Sections 4.3, 5.2, and 6.3 of Chapters 4, 5, and 6, respectively. A simulation engine was developed to assess the proposed solutions under various scenarios. The engine models different user behaviours and preferences, providing insights into the effectiveness of the proposed Human-AI collaborative solutions. The three spatial allocation strategies were tested under different levels of individual effort, availability, and commitment to determine whether they achieved the sought balance under varying conditions. It is publicly available in a GitHub repository².
- (T3.) *Development of a Telegram bot for Citizen Science.* This contribution is related to **O6** and corresponds to Sections 4.4 and 5.3 of Chapters 4 and 5, respectively. A Telegram bot was developed to facilitate user interaction with the proposed solutions. The bot enables users to receive recommendations, accept tasks, and manage assignments, enhancing overall engagement. The VTAS and Bio-VTAS solutions were tested in real-world scenarios using this bot to facilitate communication between users and the system. It is publicly available in a GitHub repository³.

7.3 Hypothesis and objective validation

The research gaps identified which have motivated this research are described next:

¹<https://github.com/mpuerta004/RecommenderSystem>

²<https://github.com/mpuerta004/RecommenderSystem>

³<https://github.com/mpuerta004/RecommenderSystem>

G.1 *Misalignment between Citizen Science and Spatial Crowdsourcing:* Traditionally, Spatial Crowdsourcing strategies have not been well-suited to the specific requirements of Citizen Science campaigns. They have primarily focused on achieving measurement objectives while neglecting the most crucial element of Citizen Science: the volunteers—the human participants who drive these initiatives. This gap seeks to analyze the misalignment between traditional Spatial Crowdsourcing strategies and their application in Citizen Science. Specifically, it aims to develop task allocation approaches that adapt to users’ preferences and behaviours, fostering acceptance and participation without relying on monetary incentives.

G.2 *Exploration of Human-AI Collaboration:* Investigate methods to integrate users into the decision-making loop while respecting their preferences, needs and demands. This field aims to address the specific needs of participants in Citizen Science projects within the context of Spatial Crowdsourcing.

G.3 *Development of an Integrated Model:* Design a model that incorporates the design criteria for more human-centred Citizen Science and Spatial Crowdsourcing projects. This model will serve as the foundation for the development of task allocation strategies and must be evaluated to ensure a balance between enhancing user engagement and maintaining task completion efficiency. By fostering sustained and effective participation without financial rewards, the central design focus shifts from purely optimal task completion to achieving a balance between a crowdsourcing campaign’s objectives and the participant users’ perspective.

At the beginning of the dissertation, in Section 1.2, the following hypothesis was formulated:

Hypothesis. Human-AI Collaboration can improve user acceptance and adoption of Spatial Crowdsourcing task allocation recommendations

within Citizen Science by balancing efficient task completion procedures with a positive user perception without relying on economic incentives.

Then, in order to validate it, the following aim was proposed:

Aim. The aim of this thesis is to develop a task allocation strategy based on a Spatial Crowdsourcing model, tailored to the context of Citizen Science, through a collaborative approach between humans and artificial intelligence. The solution aims to ensure effective task completion whilst enhancing user acceptance and adoption, without relying on economic incentives.

To achieve this aim, more specific objectives were defined. Although these objectives have been introduced in the review of the contributions of this dissertation, the following is a more specific description of how each sub-objective has been addressed in this dissertation:

- (O1) *Analysis of the Current State of Spatial Crowdsourcing in Citizen Science contexts:* This objective was achieved by **C1** in Sections 2.1 and 2.2 of Chapter 2, where the current state of Spatial Crowdsourcing in Citizen Science, focusing on the task allocation problem and the strategies is provided. The generated analysis identifies key limitations of existing methods and explores the potential of Human-AI Collaboration in addressing these challenges more effectively.
- (O2) *Analysis of the Current State of Human-AI Collaboration Workflow:* This objective was achieved by **C2** in Section 2.3 of Chapter 2. In this section, the workflows of existing Human-AI collaboration systems are examined, identifying key components and their interactions. The provided analysis facilitates the categorisation of different collaborative system architectures and provides a framework for understanding the essential elements and dynamics of such systems. These insights lay the foundation for designing a collaborative system that effectively integrates users into the decision-making process.

- (O3) *Analysis of Human-AI Collaboration Design Principles:* This objective was achieved by **C3** in Section 2.3 of Chapter 2. Such section presents an in-depth investigation of the design principles required to develop an effective Human-AI collaborative system. The study emphasises the role of user adaptation as a critical factor in integrating individuals into the decision-making loop.
- (O4) *Design of a Spatial Crowdsourcing Model:* This objective was achieved by **C4** in Chapter 3. The model described in Chapter 3 defines spatial and temporal constraints, task completion goals, and workload limitations within Citizen Science experiments. It integrates user constraints and requirements, as required by the Citizen Science field, into the Spatial Crowdsourcing problem, facilitating a collaborative framework that merges both domains.
- (O5) *Design and Implementation of a Task Allocation Recommender System:* This objective was achieved by **C5**, **C5.1**, **C5.2**, and **C5.3** in Chapter 4, Chapter 5, and Chapter 6. As a result, a Human-AI collaborative spatial task allocation suite of strategies are developed which balance user engagement and acceptance while optimising task completion efficiency. Three incremental strategies are proposed to address the task allocation problem in Citizen Science through collaboration.
- (O6) *Establish an Evaluation Methodology:* This contribution was achieved by **C6**. Based on the hypothesis, three design goals were established to determine their effectiveness in the context of Citizen Science. Several metrics and baselines were selected to evaluate these goals and, consequently, the hypothesis itself. The goals are outlined in Section 4.1 of Chapter 4, while the explanation of the metrics, baselines, and use cases is provided in Section 4.2 of the same chapter. The evaluation comprises:
- The VTAS (Chapter 4) was assessed using both a simulated use case (Section 4.3) and a real-world use case (Section 4.4).

- The Bio-VTAS (Chapter 5) was evaluated similarly, with a simulated use case (Section 5.2) and a real-world use case (Section 5.3).
- The Hybrid-VTAS (Chapter 6) was evaluated exclusively in a simulated use case (Section 6.3).

The correlation of the gaps, objectives, and contributions of this thesis is summarised in Table 7.1.

Gaps/Motivation	Objectives	Contributions
G1: Misalignment between CS and SC	O1: Analysis of the Current State of SC in CS contexts	C1: Examination of existing challenges in SC within the context of CS
G2: Exploration of Human-AI Collaboration	O2: Analysis of the Current State of Human-AI Collaboration Workflow	C2: Analysis of Human-AI Collaboration workflows in the literature.
G2: Exploration of Human-AI Collaboration	O3: Analysis of Human-AI Collaboration Design Principles	C3: Analysis of Human-AI Collaboration design principles.
G1: Misalignment between CS and SC	O4: Design of a SC Model	C4: Definition of a SC model.
G2: Exploration of Human-AI Collaboration		
G3: Development of an Integrated Model		
G3: Development of an Integrated Model	O5: Design and Implementation of a Task Allocation Recommender System	C5: Design and implementation of a collaborative task allocation recommender system: C5.1: VTAS, C5.2: Bio-VTAS, C5.3: Hybrid-VTAS
G1: Misalignment between CS and SC	O6: Establish an Evaluation Methodology	C6: Evaluation of the proposed solutions
G2: Exploration of Human-AI Collaboration		
G3: Development of an Integrated Mode		

Table 7.1: Correlation between gaps, objectives and contributions of this thesis.

7.4 Relevant publications

Through this dissertation, several scientific publications have been produced in order to validate with the research community the advances reflected in this work. The included publications fall into two categories: (i) the ones that are directly related to the specific context of this work and its contributions, and (ii) those complementary works that extend this context to study the role of Citizen Science in the context of SOCIO-BEE project. We also include in this section those supplementary materials that have been released and that contribute to the body of knowledge.

7.4.1 International JCR Journals and Book Chapters

The study of the state of the art in Human-AI collaboration, based on different workflows of collaboration and design principles (**C2** and **C3**), is presented in the following journal publication:

- Puerta-Beldarrain, M., Gómez-Carmona, O., Sánchez-Corcuera, R., Casado-Mansilla, D., López-de-Ipiña, D., & Chen, L. (2025). A multifaceted vision of the Human-AI collaboration: a comprehensive review. IEEE Access.

An explanation of the first proposed solution **C5.1**, the VTAS, has been published in the following journal:

- Puerta-Beldarrain, M.; Gómez-Carmona, O.; Chen, L.; López-de-Ipiña, D.; Casado-Mansilla, D.; Vergara-Borge, F. A Spatial Crowdsourcing Engine for Harmonizing Volunteers' Needs and Tasks' Completion Goals. *Sensors* 2024, 24, 8117. <https://doi.org/10.3390/s24248117>

An article on creating engaging and accessible tools for Citizen Science activities aimed at improving air quality in cities has been published in the following journal:

- Atutxa, E., Garcia-Torres, S., Kyfonidis, C., Karanassos, D., Kopsacheilis, E., Tsita, C., Casado-Mansilla, C., Emvolidis, A., Angelis,

G., López-de-Ipiña, G., Puerta, M., Drosou, A., & Tzovaras, D. (2024). Engagement and accessibility tools for pro-environmental action on air quality: the SOCIO-BEE paradigm. *Universal Access in the Information Society*, 1-17.

7.4.2 International Conferences

A description of the methodology carried out in the SOCIO-BEE project for monitoring air pollution levels, promoting awareness, and contributing to data-driven urban planning through citizen science initiatives is presented in the following conference publication:

- López-de-Ipiña, D., Puerta, M., Barco, A., Casado-Mansilla, D., Gómez-Carmona, O., Morresi, N., Kotzagianni, M., López-Gomollón, C. & Urdina, S. (2024, November). Towards the Democratization of Citizen Science Campaigns for Air Quality Assessment. In *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 729-740). Cham: Springer Nature Switzerland.

A study on the design of a collaborative environment for co-delivering citizen science campaigns is presented in the following conference conference:

- López-de-Ipiña, D., Silva, D. A., Sánchez, R., Not, E., Leonardi, C., Puerta, M., Oihane Gómez-Carmona, O., Casado-Mansilla, D. & Vergara, F. (2023, November). A collaborative environment for co-delivering citizen science campaigns. In *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 322-333). Cham: Springer Nature Switzerland.

The analysis of the Micro-Volunteer engine's work in the context of air quality monitoring is presented in the following conference:

- Puerta-Beldarrain, M., Gómez-Carmona, O., López-de-Ipiña, D., Casado-Mansilla, D., Barco, A., Jayo, U. H., & García-Zubia, J. (2023, June). A Micro-volunteering Engine to drive crowd-measuring of Air Quality in Citizen Science. In *2023 8th International Conference on Smart and Sustainable Technologies (SpliTech)* (pp. 1-6). IEEE.

An analysis of the most important human factors in Human-AI collaboration to promote trust, engagement, and adaptation in the process of pro-environmental and health behaviour change is presented in the following publication:

- Puerta-Beldarrain, M., Gómez-Carmona, O., Casado-Mansilla, D., & López-de-Ipiña, D. (2022, November). Human-AI collaboration to promote trust, engagement and adaptation in the process of pro-environmental and health behaviour change. In *International Conference on Ubiquitous Computing and Ambient Intelligence* (pp. 381-392). Cham: Springer International Publishing.

An example of how to carry out co-creation and assess the impact of the campaign related to the SOCIO-BEE project is presented in the following conference:

- Hernandez-Jayo, U., Garcia-Zubia, J., López-De-Ipiña, D., Casado-Mansilla, D., Puerta-Beldarrain, M., & Martelo, A. B. (2023, June). SOCIO-BEE: Example of a Citizen Science Community for the Co-Creation of Measures to Reduce the Impact of Environmental Pollution in the Framework of Smart Cities. In *2023 6th Experiment International Conference (exp. at'23)* (pp. 101-105). IEEE.

7.4.3 Software

- *Github Repository*: The technical contributions: **T1**, **T2** and **T3** can be found in this public Github repository¹.

7.5 Future Work

The following future research aspects have been identified based on the observations and limitations encountered throughout this dissertation:

¹<https://github.com/mpuerta004/RecommenderSystem>

Several potential avenues for further investigation involve considering new additional parameters to the default configurations established within the Spatial Crowdsourcing (SC) model defined in this work. These different and additional parameters' configurations could be tested to determine whether they lead to improved results or not, or whether the new configurations enable the SC model to maintain or achieve equilibrium or not. The areas identified for further analysis include:

- A key area for future research is investigating the impact of a geographical area's internal cell shapes on the performance of task allocation solutions. Specifically, this research would focus on how different cell configurations can improve the alignment between user perception and task completion efficiency.
- A promising future research direction involves exploring the impact of dynamic Spatial Crowdsourcing, where cells are no longer static, on the performance of task allocation solutions. Unlike the static cell-based approach, dynamic solutions would adapt to evolving spatial conditions, potentially enhancing both efficiency and user engagement.
- A key avenue for future research is investigating the effect of choice overload on the acceptance and adoption of the system. This could involve testing the optimal number of cells to recommend per request, striking a balance between usability and the diversity of available options.
- Another potential research direction is to replicate the experiment within a European project, scaling the proposed Human–AI collaboration framework to a much broader population to compare the results.

Modifying the SC model could lead to significant changes in the users' perspectives. Furthermore, exploring how collaboration can be integrated within these models could be a crucial step in future research, as it would expand the types of SC models applicable to CS, potentially enhancing the overall effectiveness. As a matter of fact, the Spatial Crowdsourcing model contributed in this dissertation should not be considered as the final one. It

is envisaged that several varieties of this seed model could be devised to best suit Spatial Crowdsourcing problems, different to the ones which have inspired this work, i.e. air quality data crowdsourcing in diverse geographical areas and periods of time, as was demanded by SOCIO-BEE project.

Additionally, further research could be conducted to investigate how various parameters, as defined in this thesis, influence the results obtained:

- A future line of work involves integrating motivation and engagement mechanisms (e.g. gamification, incentivization with rewards (social coin) and so on) into the solutions presented in this thesis to evaluate how these new mechanisms influence user perception while maintaining task completion efficiency. The incorporation of such mechanisms could enhance user perception, and testing how collaboration, supported by these tools, contributes to achieving the desired balance between user perception and task completion efficiency, would be a promising direction for future research.
- Analysing the impact of the number of recommended cells per request on the performance of the proposed solutions, focusing on balancing usability with the diversity of available options. Increasing personalisation by allowing users to select the number of recommended cells per request, potentially through an initial test that lets users choose the algorithm that best fits their needs, could improve user perception. Along with collaboration, this personalisation could significantly enhance the user experience. Investigating this new research line could represent a major step in future research.
- Creating a collaborative environment that facilitates the integration of user feedback during their engagement in a campaign. While these aspects were deliberately excluded from this thesis, incorporating them could substantially improve the effectiveness of the proposed solutions. Integrating these insights into the collaboration mechanisms defined in this dissertation would be an important step forward in future research.

- Developing an Active Learning or Reinforcement Learning model that can learn from user feedback and adapt recommendations based on individual preferences. This AI-based approach represents a promising direction for future research.

All of these proposed lines of future research could be explored within the context of the central idea addressed in this dissertation: integrating the fields of Crowdsourcing (CS) and Spatial Crowdsourcing (SC) to enhance user perception and maintain task completion efficiency.

7.6 Final remarks

This dissertation aimed to make significant contributions to the research field of Citizen Science and Spatial crowdsourcing by proposing a novel approach to spatial task allocation that integrates human and artificial intelligence collaboration. We firmly believe that Human-AI Collaboration is the key to improving user acceptance and adoption in spatial crowdsourcing task allocation problems within Citizen Science. By balancing efficient task completion procedures with a positive user perception without relying on economic incentives, we can still foster sustained and effective participation. Hence, this dissertation opens the door toward a future where intelligent collaboration with smart AI-based agents and/or recommender places citizens at the heart of innovation, transforming spatial crowdsourcing, and potentially other disciplines requiring Human-AI collaboration, into an engaging, sustainable, and truly human-centred experience.

Bibliography

- Abdelghani, W., Zayani, C. A., Amous, I., and Sèdes, F. (2019). User-centric iot: challenges and perspectives. In *UBICOMM 2018: The Twelfth International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies*, pages 27–34. 59
- Adetayo, A. J. (2024). Reimagining learning through ai art: the promise of dall-e and midjourney for education and libraries. *Library Hi Tech News*, (ahead-of-print). 43
- Aggarwal, A., Tam, C. C., Wu, D., Li, X., and Qiao, S. (2023). Artificial intelligence-based chatbots for promoting health behavioral changes: Systematic review. *Journal of Medical Internet Research*, 25:e40789. 62
- Ait Baha, T., El Hajji, M., Es-Saady, Y., and Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, 29(8):10153–10176. 43
- Akata, Z., Balliet, D., De Rijke, M., Dignum, F., Dignum, V., Eiben, G., Fokkens, A., Grossi, D., Hindriks, K., Hoos, H., et al. (2020). A research agenda for hybrid intelligence: augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(08):18–28. 28
- Ala’a, A.-M. and Ramayah, T. (2023). A review of the technology acceptance model in electronic health records. *International Journal of Business and Technology Management*, 5(2):8–19. 5

- Allen, J., Moussa, A., and Liu, X. (2019). Human-in-the-loop learning of qualitative preference models. In *The Thirty-Second International Flairs Conference*. 32, 33
- Alt, F., Shirazi, A. S., Schmidt, A., Kramer, U., and Nawaz, Z. (2010). Location-based crowdsourcing: extending crowdsourcing to the real world. In *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries*, pages 13–22. 23
- Amershi, S., Cakmak, M., Knox, W. B., and Kulesza, T. (2014). Power to the people: The role of humans in interactive machine learning. *Ai Magazine*, 35(4):105–120. 58, 59
- Amershi, S., Chickering, M., Drucker, S. M., Lee, B., Simard, P., and Suh, J. (2015). Modeltracker: Redesigning performance analysis tools for machine learning. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 337–346. 51
- Beierle, F., Aizawa, A., and Beel, J. (2017). Exploring choice overload in related-article recommendations in digital libraries. *arXiv preprint arXiv:1704.00393*. 83, 88, 149
- Berger, B., Adam, M., Rühr, A., and Benlian, A. (2021). Watch me improve—algorithm aversion and demonstrating the ability to learn. *Business & Information Systems Engineering*, 63(1):55–68. 39, 52
- Bollen, D., Knijnenburg, B. P., Willemsen, M. C., and Graus, M. (2010). Understanding choice overload in recommender systems. In *Proceedings of the fourth ACM conference on Recommender systems*, pages 63–70. 83, 88, 149
- Bonabeau, E., Sobkowski, A., Theraulaz, G., Deneubourg, J.-L., et al. (1997). Adaptive task allocation inspired by a model of division of labor in social insects. In *BCEC*, pages 36–45. 144

- Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., and Wilderman, C. C. (2009a). Public participation in scientific research: Defining the field and assessing its potential for informal science education. a caise inquiry group report. *Online submission*. 19
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V., and Shirk, J. (2009b). Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59(11):977–984. 4, 18, 19
- Bonney, R., Phillips, T. B., Ballard, H. L., and Enck, J. W. (2016). Can citizen science enhance public understanding of science? *Public understanding of science*, 25(1):2–16. 20
- Bosch, K. v. d., Schoonderwoerd, T., Blankendaal, R., and Neerincx, M. (2019). Six challenges for human-ai co-learning. In *International Conference on Human-Computer Interaction*, pages 572–589. Springer. 6, 28
- Breckner, K., Neumayr, T., Streit, M., and Augstein, M. (2024). Personalized complementarity in human-ai collaboration. In *Mensch und Computer 2024-Workshopband*, pages 10–18420. Gesellschaft für Informatik eV. 61
- Buehler, M. C. and Weisswange, T. H. (2020). Theory of mind based communication for human agent cooperation. In *2020 IEEE International Conference on Human-Machine Systems (ICHMS)*, pages 1–6. IEEE. 53
- Cai, Q., Liu, S., Wang, X., Zuo, T., Xie, W., Yang, B., Zheng, D., Jiang, P., and Gai, K. (2023). Reinforcing user retention in a billion scale short video recommender system. In *Companion Proceedings of the ACM Web Conference 2023*, pages 421–426. 31, 32
- Cai, Z., Duan, Z., and Li, W. (2020). Exploiting multi-dimensional task diversity in distributed auctions for mobile crowdsensing. *IEEE Transactions on Mobile Computing*, 20(8):2576–2591. 23, 25

- Cao, C., Fu, Y., Xu, S., Zhang, R., and Li, S. (2024). Enhancing human-ai collaboration through logic-guided reasoning. In *The Twelfth International Conference on Learning Representations*. 46, 47
- Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., and Sun, L. (2023). A comprehensive survey of ai-generated content (aigc): A history of generative ai from gan to chatgpt. *arXiv preprint arXiv:2303.04226*. 6, 28
- Carney, R. M., Long, A., Low, R. D., Zohdy, S., Palmer, J. R., Elias, P., Bartumeus, F., Njoroge, L., Muniafu, M., Uelmen, J. A., et al. (2023). Citizen science as an approach for responding to the threat of *Anopheles stephensi* in africa. *Citizen science: theory and practice*, 8(1). 19, 20
- Carvalho, I. and Ivanov, S. (2024). Chatgpt for tourism: applications, benefits and risks. *Tourism Review*, 79(2):290–303. 41
- Cen, C., Cheng, S.-F., Lau, H. C., and Misra, A. (2015). Towards city-scale mobile crowdsourcing: Task recommendations under trajectory uncertainties. 23, 26
- Chander, B., John, C., Warriar, L., and Gopalakrishnan, K. (2024). Toward trustworthy artificial intelligence (tai) in the context of explainability and robustness. *ACM Computing Surveys*. 57
- Chen, X., Wang, S., McAuley, J., Jannach, D., and Yao, L. (2024a). On the opportunities and challenges of offline reinforcement learning for recommender systems. *ACM Transactions on Information Systems*, 42(6):1–26. 31
- Chen, Y. and Wang, H. (2020). Intelligentcrowd: Mobile crowdsensing via multi-agent reinforcement learning. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 5(5):840–845. 23, 26
- Chen, Z., Xu, M., and Su, C. (2024b). Online quality-based privacy-preserving task allocation in mobile crowdsensing. *COMPUTER NETWORKS*, 251. 23, 26

- Cooke, N. J., Gorman, J. C., Myers, C. W., and Duran, J. L. (2013). Interactive team cognition. *Cognitive science*, 37(2):255–285. 61
- Cunha, D. G., Marques, J. F., Resende, J. C., Falco, P. B., Souza, C. M., and Loisel, S. A. (2017). Citizen science participation in research in the environmental sciences: key factors related to projects’ success and longevity. *Anais da Academia Brasileira de Ciências*, 89:2229–2245. 19
- Davis, F. D. et al. (1989). Technology acceptance model: Tam. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption*, 205:219. 93, 123, 137, 291
- Dawar, S., Chatterjee, S., Hossain, M. F., and Malarvizhi, S. (2023). Music recommendation system using real time parameters. pages 1–6. 31, 32
- Deng, W., Xu, J., Song, Y., and Zhao, H. (2020). An effective improved co-evolution ant colony optimisation algorithm with multi-strategies and its application. *International Journal of Bio-Inspired Computation*, 16(3):158–170. 45
- Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., and Robert, L. P. (2024). Shaping human-ai collaboration: Varied scaffolding levels in co-writing with language models. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18. 42
- Difallah, D. E., Catasta, M., Demartini, G., Ipeirotis, P. G., and Cudré-Mauroux, P. (2015). The dynamics of micro-task crowdsourcing: The case of amazon mturk. In *Proceedings of the 24th international conference on world wide web*, pages 238–247. 22
- Ding, X., Lv, R., Pang, X., Hu, J., Wang, Z., Yang, X., and Li, X. (2022). Privacy-preserving task allocation for edge computing-based mobile crowdsensing. *Computers & Electrical Engineering*, 97:107528. 23, 26
- Duan, J., Lu, J., Jiang, W., and Yang, S. (2020). Incentivizing fairness-aware task allocation in mobile crowdsensing with sweep coverage and stability control. *APPLIED SOFT COMPUTING*, 97(A). 26

- Duan, W., McNeese, N., and Zhang, R. (2023). Communication in human-ai teaming. *Group Communication*, pages 340–352. 61
- Dudley, J. J. and Kristensson, P. O. (2018). A review of user interface design for interactive machine learning. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 8(2):1–37. 60
- Earp, H. S. and Liconti, A. (2020). *Science for the Future: The Use of Citizen Science in Marine Research and Conservation*, pages 1–19. Springer International Publishing, Cham. 19
- Eiband, M., Buschek, D., and Hussmann, H. (2021). How to support users in understanding intelligent systems? structuring the discussion. In *26th International Conference on Intelligent User Interfaces*, pages 120–132. 57
- Endsley, M. R. (2023). Supporting human-ai teams: Transparency, explainability, and situation awareness. *Computers in Human Behavior*, 140:107574. 36, 57
- English, P. B., Richardson, M. J., and Garzón-Galvis, C. (2018). From crowdsourcing to extreme citizen science: participatory research for environmental health. *Annual review of public health*, 39:335–350. 21
- Finke, P. (2014). Citizen science. *Das unterschätzte Wissen der Laien. München: oekom.* xxx, 1
- Froeling, F., Gignac, F., Toran, R., Ortiz, R., Ficorilli, A., De Marchi, B., Biggeri, A., Kocman, D., Ftičar, J., Tratnik, J. S., et al. (2024). Implementing co-created citizen science in five environmental epidemiological studies in the cities-health project. *Environmental research*, 240:117469. 19, 21
- Gao, H. and Zhao, H. (2022). A personalized task allocation strategy in mobile crowdsensing for minimizing total cost. *SENSORS*, 22(7). 23, 25
- Gillotte, J. L. (2019). Copyright infringement in ai-generated artworks. *UC Davis L. Rev.*, 53:2655. 38, 43

- Glas, N. and Pelachaud, C. (2015). Definitions of engagement in human-agent interaction. In *2015 International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 944–949. IEEE. 59
- Gómez-Carmona, O., Casado-Mansilla, D., López-de Ipiña, D., and García-Zubia, J. (2024). Human-in-the-loop machine learning: Reconceptualizing the role of the user in interactive approaches. *Internet of Things*, 25:101048. 4, 21
- Grier, R. A., Bangor, A., Kortum, P., and Peres, S. C. (2013). The system usability scale: Beyond standard usability testing. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 57, pages 187–191. SAGE Publications Sage CA: Los Angeles, CA. 93, 123, 137, 291
- Guimarães, D., Paulino, D., Correia, A., Trigo, L., Brazdil, P., and Paredes, H. (2021). Towards a human-ai hybrid framework for inter-researcher similarity detection. In *2021 IEEE 2nd International Conference on Human-Machine Systems (ICHMS)*, pages 1–4. IEEE. 38, 39
- Gummidi, S. R. B., Xie, X., and Pedersen, T. B. (2019). A survey of spatial crowdsourcing. *ACM Transactions on Database Systems (TODS)*, 44(2):1–46. xxxii, 1, 2, 22, 23, 24
- Guo, B., Liu, Y., Wang, L., Li, V. O., Lam, J. C., and Yu, Z. (2018). Task allocation in spatial crowdsourcing: Current state and future directions. *IEEE Internet of Things Journal*, 5(3):1749–1764. 2, 22, 26
- Gómez-Carmona, O., García-Zubia, J., and Diego Casado-Mansilla, D. (2021). *Promoting the Perception of Emerging Technologies in Work Environments through Edge Computing and Hybrid Intelligence*. PhD thesis, University of Deusto. 45, 46
- Haklay, M. (2012). Citizen science and volunteered geographic information: Overview and typology of participation. *Crowdsourcing geographic knowledge: Volunteered geographic information (VGI) in theory and practice*, pages 105–122. 4, 18

- Haklay, M. (2015). Citizen science and policy: A european perspective. 20
- Haklay, M., Dörler, D., Heigl, F., Manzoni, M., Hecker, S., Vohland, K., et al. (2021). What is citizen science? the challenges of definition. *The science of citizen science*, 13. xxx, 1, 18
- Halder, R. K., Uddin, M. N., Uddin, M. A., Aryal, S., and Khraisat, A. (2024). Enhancing k-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, 11(1):113. 90
- Han, W., Tian, Z., Zhu, C., Huang, Z., Jia, Y., and Guizani, M. (2019). A topic representation model for online social networks based on hybrid human–artificial intelligence. *IEEE Transactions on Computational Social Systems*, 8(1):191–200. 32
- Han, X., Zhang, Z., Ding, N., Gu, Y., Liu, X., Huo, Y., Qiu, J., Yao, Y., Zhang, A., Zhang, L., et al. (2021). Pre-trained models: Past, present and future. *AI Open*, 2:225–250. 6, 28
- Hanika, T., Herde, M., Kuhn, J., Leimeister, J. M., Lukowicz, P., Oeste-Reiß, S., Schmidt, A., Sick, B., Stumme, G., Tomforde, S., et al. (2019). Collaborative interactive learning—a clarification of terms and a differentiation from other research fields. 28
- Hanna, N. and Richards, D. (2019). Speech act theory as an evaluation tool for human–agent communication. *Algorithms*, 12(4):79. 62
- Hasegawa-Johnson, M., Black, A., Ondel, L., Scharenborg, O., and Ciannella, F. (2017). Image2speech: Automatically generating audio descriptions of images. *Casablanca 2017*, page 65. 38, 44
- Heinrich, J., Lanctot, M., and Silver, D. (2015). Fictitious self-play in extensive-form games. In *International conference on machine learning*, pages 805–813. PMLR. 48

- Hekler, A., Utikal, J. S., Enk, A. H., Hauschild, A., Weichenthal, M., Maron, R. C., Berking, C., Haferkamp, S., Klode, J., Schadendorf, D., et al. (2019). Superior skin cancer classification by the combination of human and artificial intelligence. *European Journal of Cancer*, 120:114–121. 38, 39
- Hemmer, P., Schemmer, M., Riefle, L., Rosellen, N., Vössing, M., and Kühl, N. (2022). Factors that influence the adoption of human-ai collaboration in clinical decision-making. *arXiv preprint arXiv:2204.09082*. 39
- Hettiachchi, D., Kostakos, V., and Goncalves, J. (2022). A survey on task assignment in crowdsourcing. *ACM Computing Surveys (CSUR)*, 55(3):1–35. 2
- Hofeditz, L., Mirbabaie, M., and Ortmann, M. (2023). Ethical challenges for human-agent interaction in virtual collaboration at work. *International Journal of Human-Computer Interaction*, pages 1–17. 6, 28, 29
- Holder, E., Huang, L., Chiou, E., Jeon, M., and Lyons, J. B. (2021). Designing for bi-directional transparency in human-ai-robot-teaming. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 65, pages 57–61. SAGE Publications Sage CA: Los Angeles, CA. 61
- Holzinger, A., Plass, M., Holzinger, K., Crisan, G. C., Pintea, C.-M., and Palade, V. (2017). A glass-box interactive machine learning approach for solving np-hard problems with the human-in-the-loop. *arXiv preprint arXiv:1708.01104*. 45, 46
- Holzinger, A., Plass, M., Kickmeier-Rust, M., Holzinger, K., Crişan, G. C., Pintea, C.-M., and Palade, V. (2019). Interactive machine learning: experimental evidence for the human in the algorithmic loop. *Applied Intelligence*, 49(7):2401–2414. 38, 41
- Honeycutt, D., Nourani, M., and Ragan, E. (2020). Soliciting human-in-the-loop user feedback for interactive machine learning reduces user trust and impressions of model accuracy. In *Proceedings of the AAAI Conference on Human Computation and Crowdsourcing*, volume 8, pages 63–72. 32, 34, 58

- Howe, J. (2006a). Crowdsourcing: A definition. *Retrieved March*, 5:2009. 21
- Howe, J. (2006b). The rise of crowdsourcing, wired. <http://www.wired.com/wired/archive/14.06/crowds.html>. 21
- Hua, S., Jin, S., and Jiang, S. (2024). The limitations and ethical considerations of chatgpt. *Data intelligence*, 6(1):201–239. 56
- Huang, R., Li, M., Yang, D., Shi, J., Chang, X., Ye, Z., Wu, Y., Hong, Z., Huang, J., Liu, J., et al. (2024a). Audiogpt: Understanding and generating speech, music, sound, and talking head. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 23802–23804. 38, 44
- Huang, R., Li, M., Yang, D., Shi, J., Chang, X., Ye, Z., Wu, Y., Hong, Z., Huang, J., Liu, J., et al. (2024b). Audiogpt: Understanding and generating speech, music, sound, and talking head. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 23802–23804. 44
- Huang, Y., Chen, H., Ma, G., Lin, K., Ni, Z., Yan, N., and Wang, Z. (2022). Opat: Optimized allocation of time-dependent tasks for mobile crowdsensing. *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, 18(4):2476–2485. 23, 25
- Janhunnen, E., Toivikko, T., Blomqvist, K., and Siemon, D. (2024). Trust in digital human-ai team collaboration: A systematic review. 57
- Jin, W., Xiao, M., Guo, L., Yang, L., and Li, M. (2021). Ulpt: A user-centric location privacy trading framework for mobile crowd sensing. *IEEE Transactions on Mobile Computing*, 21(10):3789–3806. 23, 26
- Jirgl, M., Bradac, Z., and Fiedler, P. (2018). Human-in-the-loop issue in context of the cyber-physical systems. *IFAC-PapersOnLine*, 51(6):225–230. 52
- Kartoun, U. (2017). Text nailing: an efficient human-in-the-loop text-processing method. *Interactions*, 24(6):44–49. 50, 51

- Keikhosrokiani, P. and Fye, G. M. (2024). A hybrid recommender system for health supplement e-commerce based on customer data implicit ratings. *Multimedia Tools and Applications*, 83(15):45315–45344. 31, 32
- Keim, D. A. (2000). Designing pixel-oriented visualization techniques: Theory and applications. *IEEE Transactions on visualization and computer graphics*, 6(1):59–78. 51
- Kelly, J. P. and Bridge, D. (2006). Enhancing the diversity of conversational collaborative recommendations: a comparison. *Artificial Intelligence Review*, 25:79–95. 56
- Knox, W. B., Stone, P., and Breazeal, C. (2013). Training a robot via human feedback: A case study. In *International Conference on Social Robotics*, pages 460–470. Springer. 63
- Köbis, N. and Mossink, L. D. (2021). Artificial intelligence versus maya angelou: Experimental evidence that people cannot differentiate ai-generated from human-written poetry. *Computers in human behavior*, 114:106553. 38, 42
- Kopp, S. and Krämer, N. (2021). Revisiting human-agent communication: The importance of joint co-construction and understanding mental states. *Frontiers in Psychology*, 12:580955. 61
- Kuhail, M. A., Alturki, N., Alramlawi, S., and Alhejori, K. (2023). Interacting with educational chatbots: A systematic review. *Education and Information Technologies*, 28(1):973–1018. 63
- Lane, O. and Klave, K. (2024). Quantifying ethics and trust in human-ai collaboration. 57
- Le, T., Nguyen, T., Ho, N., Bui, H., and Phung, D. (2021). Lamda: Label matching deep domain adaptation. In *International Conference on Machine Learning*, pages 6043–6054. PMLR. 38, 43

- Lee, M. H., Siewiorek, D. P., Smailagic, A., Bernardino, A., and Bermúdez i Badia, S. (2021). A human-ai collaborative approach for clinical decision making on rehabilitation assessment. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14. 32, 33
- Lehmann, J., Castillo, C., Lalmas, M., and Zuckerman, E. (2013a). Finding news curators in twitter. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 863–870. 2, 23
- Lehmann, J., Castillo, C., Lalmas, M., and Zuckerman, E. (2013b). Transient news crowds in social media. 7(1):351–360. 2, 23
- Levontin, L., Gilad, Z., Shuster, B., Chako, S., Land-Zandstra, A., Lavie-Alon, N., and Shwartz, A. (2022). Standardizing the assessment of citizen scientists’ motivations: A motivational goal-based approach. *Citizen Science: Theory and practice*, 7(1):25. 1, 4, 20
- Li, B., Yang, P., Sun, Y., Hu, Z., and Yi, M. (2024a). Advances and challenges in artificial intelligence text generation. *Frontiers of Information Technology & Electronic Engineering*, 25(1):64–83. 42
- Li, H., Ni, T., Agrawal, S., Jia, F., Raja, S., Gui, Y., Hughes, D., Lewis, M., and Sycara, K. (2021a). Individualized mutual adaptation in human-agent teams. *IEEE Transactions on Human-Machine Systems*, 51(6):706–714. 53
- Li, H., Yu, Z., Luo, Y., Cui, H., and Guo, B. (2024b). Continuoussensing: a task allocation algorithm for human-robot collaborative mobile crowd-sensing with task migration. *CCF TRANSACTIONS ON PERVASIVE COMPUTING AND INTERACTION*. 23, 25
- Li, J., Dada, A., Puladi, B., Kleesiek, J., and Egger, J. (2024c). Chatgpt in healthcare: a taxonomy and systematic review. *Computer Methods and Programs in Biomedicine*, page 108013. 43
- Li, J., Wang, Y., Yang, B., Zhou, Q., and Yuan, F. (2022). Research on fuzzy decision-making method of task allocation for ship multiagent collaborative design. *ADVANCES IN MATHEMATICAL PHYSICS*, 2022. 23, 25

- Li, J., Zhang, W., Wang, T., Xiong, G., Lu, A., and Medioni, G. (2023). Gpt4rec: A generative framework for personalized recommendation and user interests interpretation. *arXiv preprint arXiv:2304.03879*. 56
- Li, S., Shen, W., Bilal, M., Xu, X., Dou, W., and Moustafa, N. (2021b). Fair and size-scalable participant selection framework for large-scale mobile crowdsensing. *Journal of Systems Architecture*, 119:102273. 26
- Li, Y., Xu, J., Guo, D., and Liu, H. (2024d). Trust-aware human–robot fusion decision-making for emergency indoor patrolling. *IEEE Transactions on Automation Science and Engineering*. 46, 49
- Lin, J., Tomlin, N., Andreas, J., and Eisner, J. (2024). Decision-oriented dialogue for human-ai collaboration. *Transactions of the Association for Computational Linguistics*, 12:892–911. 35, 37
- Liu, A., Guerra, S., Fung, I., Matute, G., Kamar, E., and Lasecki, W. (2020). Towards hybrid human-ai workflows for unknown unknown detection. In *Proceedings of The Web Conference 2020*, pages 2432–2442. 32, 33
- Liu, H., Yuan, Y., Liu, X., Mei, X., Kong, Q., Tian, Q., Wang, Y., Wang, W., Wang, Y., and Plumbley, M. D. (2024a). Audioldm 2: Learning holistic audio generation with self-supervised pretraining. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. 44
- Liu, J., Wang, Y., Huang, X., Korsós, M. B., Jiang, Y., Wang, Y., and Erdélyi, R. (2021). Reliability of ai-generated magnetograms from only euv images. *Nature Astronomy*, 5(2):108–110. 38, 43
- Liu, X., Zhu, Z., Liu, H., Yuan, Y., Cui, M., Huang, Q., Liang, J., Cao, Y., Kong, Q., Plumbley, M. D., et al. (2023). Wavjourney: Compositional audio creation with large language models. *arXiv preprint arXiv:2307.14335*. 38, 44
- Liu, Y., Caldwell, G., Rittenbruch, M., Belek Fialho Teixeira, M., Burden, A., and Guertler, M. (2024b). What affects human decision making in human–robot collaboration?: A scoping review. *Robotics*, 13(2):30. 48, 53

- Liu, Z., Wang, J., Gong, S., Lu, H., and Tao, D. (2019). Deep reinforcement active learning for human-in-the-loop person re-identification. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6122–6131. 38
- Loeschcke, S., Belongie, S., and Benaim, S. (2022). Text-driven stylization of video objects. In *European Conference on Computer Vision*, pages 594–609. Springer. 38, 43
- Lou, X., Guo, J., Zhang, J., Wang, J., Huang, K., and Du, Y. (2023). Pecan: Leveraging policy ensemble for context-aware zero-shot human-ai coordination. 46, 49
- Luther, T., Kimmerle, J., and Cress, U. (2024). Teaming up with an ai: Exploring human-ai collaboration in a writing scenario with chatgpt. *AI*, 5(3):1357–1376. 42
- Maathuis, C. (2024). Human-centered ai in military cyber operations. In *International Conference on Cyber Warfare and Security*, volume 19, pages 121–128. 36
- Manion, M. (2002). Ethics, engineering, and sustainable development. *IEEE Technology and Society Magazine*, 21(3):39–48. 52
- Marangunić, N. and Granić, A. (2015). Technology acceptance model: a literature review from 1986 to 2013. *Universal access in the information society*, 14:81–95. 5, 93
- Marathe, A. R., Schaefer, K. E., Evans, A. W., and Metcalfe, J. S. (2018). Bidirectional communication for effective human-agent teaming. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 338–350. Springer. 61
- Marri, S. 12 conversational archetypes for human-ai interaction. 43

- Masson, C. B., Martin, D., Colombino, T., and Grasso, A. (2016). “the device is not well designed for me” on the use of activity trackers in the workplace? In *COOP 2016: Proceedings of the 12th International Conference on the Design of Cooperative Systems, 23-27 May 2016, Trento, Italy*, pages 39–55. Springer. 59
- Masters, K. L., Team, G. Z., et al. (2019). Twelve years of galaxy zoo. *Proceedings of the International Astronomical Union*, 14(S353):205–212. 19
- Mathewson, K. W. (2019). A human-centered approach to interactive machine learning. *arXiv preprint arXiv:1905.06289*. 60
- McCallum, L. and Fiebrink, R. (2019). Supporting feature engineering in end-user machine learning. 58
- McCamish, B., Termehchy, A., and Touri, B. (2017). A game-theoretic approach to data interaction: A progress report. In *Proceedings of the 2nd Workshop on Human-In-the-Loop Data Analytics*, pages 1–4. 46, 50, 62
- Mehak, S., Kelleher, J. D., Guilfoyle, M., and Leva, M. C. (2024). Action recognition for human–robot teaming: Exploring mutual performance monitoring possibilities. *Machines*, 12(1):45. 46, 48
- Mehrotra, S., Degachi, C., Vereschak, O., Jonker, C. M., and Tielman, M. L. (2024). A systematic review on fostering appropriate trust in human-ai interaction: Trends, opportunities and challenges. *ACM Journal on Responsible Computing*. 56
- Meitei, M. G. and Marchang, N. (2020). A user-centric task allocation approach in mobile crowdsensing. In *Electronic Systems and Intelligent Computing: Proceedings of ESIC 2020*, pages 627–639. Springer. 27
- Meitei, M. G. and Marchang, N. (2024a). Altruistic user-oriented task allocation techniques for mobile crowdsensing. *CCF Transactions on High Performance Computing*, 6(4):378–396. 27

- Meitei, M. G. and Marchang, N. (2024b). Provisioning load balancing in time-sensitive task allocation for mobile crowdsensing. *Journal of Network and Systems Management*, 32(1):13. 1, 5, 27
- Meng, Z., Yu, C., and Qian, Y. (2023). Privacy-preserving task allocation and decentralized dispute protocol in mobile crowdsourcing. In *ICC 2023-IEEE International Conference on Communications*, pages 1579–1584. IEEE. 23, 26
- Meza Martínez, M. A., Nadj, M., and Maedche, A. (2019). Towards an integrative theoretical framework of interactive machine learning systems. 58
- Milani, S., Topin, N., Veloso, M., and Fang, F. (2024). Explainable reinforcement learning: A survey and comparative review. *ACM Computing Surveys*, 56(7):1–36. 61
- Mohammad, Y. and Nishida, T. (2008). Human adaptation to a miniature robot: Precursors of mutual adaptation. In *RO-MAN 2008-The 17th IEEE International Symposium on Robot and Human Interactive Communication*, pages 124–129. IEEE. 55
- Mousa, H., Mokhtar, S. B., Hasan, O., Younes, O., Hadhoud, M., and Brunie, L. (2015). Trust management and reputation systems in mobile participatory sensing applications: A survey. *Computer Networks*, 90:49–73. 1, 4, 27
- Mozafari, B., Sarkar, P., Franklin, M., Jordan, M., and Madden, S. (2014). Scaling up crowd-sourcing to very large datasets: a case for active learning. *Proceedings of the VLDB Endowment*, 8(2):125–136. 38, 41
- Murthy, S. and Murthy, S. S. (2024). Sapphire vs conventional youtube video ranking. 31, 32
- Musick, G., O’Neill, T. A., Schelble, B. G., McNeese, N. J., and Henke, J. B. (2021). What happens when humans believe their teammate is an ai? an investigation into humans teaming with autonomy. *Computers in Human Behavior*, 122:106852. 36

- Mysore, S., Jasim, M., McCallum, A., and Zamani, H. (2023). Editable user profiles for controllable text recommendations. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 993–1003. 56
- Nassar, Y., Albeaino, G., Jeelani, I., Gheisari, M., and Issa, R. R. (2024). Human-robot collaboration levels in construction: Focusing on individuals’ cognitive workload. In *Construction Research Congress*, pages 639–648. 48, 53
- Nguyen, A., Ilesanmi, F., Dang, B., Vuorenmaa, E., and Järvelä, S. (2024). Hybrid intelligence in academic writing: Examining self-regulated learning patterns in an ai-assisted writing task. In *HHAI 2024: Hybrid Human AI Systems for the Social Good*, pages 241–254. IOS Press. 42
- Nikolaidis, S. (2017). *Mathematical Models of Adaptation in Human-Robot Collaboration*. PhD thesis, Carnegie Mellon University. 53, 55
- Nikolaidis, S., Hsu, D., and Srinivasa, S. (2017a). Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research*, 36(5-7):618–634. 46, 49, 53, 54
- Nikolaidis, S., Nath, S., Procaccia, A. D., and Srinivasa, S. (2017b). Game-theoretic modeling of human adaptation in human-robot collaboration. In *Proceedings of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 323–331. 49, 55, 57
- Nikolaidis, S., Zhu, Y. X., Hsu, D., and Srinivasa, S. (2017c). Human-robot mutual adaptation in shared autonomy. In *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 294–302. IEEE. 6, 46, 49, 54, 57
- O’Brien, H. L. and Toms, E. G. (2008). What is user engagement? a conceptual framework for defining user engagement with technology. *Journal of the American society for Information Science and Technology*, 59(6):938–955. 59

- Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., and Peters, C. (2020). Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI*, 7:92. 60
- Ostheimer, J., Chowdhury, S., and Iqbal, S. (2021). An alliance of humans and machines for machine learning: Hybrid intelligent systems and their design principles. *Technology in Society*, 66:101647. 56
- Ostrom, E. (1996). Crossing the great divide: Coproduction, synergy, and development. *World development*, 24(6):1073–1087. 21
- Pandya, R. E. (2012). A framework for engaging diverse communities in citizen science in the us. *Frontiers in Ecology and the Environment*, 10(6):314–317. 20
- Paripati, L., Hajari, V. R., Narukulla, N., Prasad, N., Shah, J., and Agarwal, A. (2024). Ai algorithms for personalization: Recommender systems, predictive analytics, and beyond. *Darpan International Research Analysis*, 12(2):51–63. 56
- Peng, T., Zhong, W., Wang, G., Zhang, S., Luo, E., and Wang, T. (2024). Spatiotemporal-aware privacy-preserving task matching in mobile crowd-sensing. *IEEE INTERNET OF THINGS JOURNAL*, 11(2):2394–2406. 23, 26
- Pérez-Quñones, M. A. and Sibert, J. L. (1996). A collaborative model of feedback in human-computer interaction. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pages 316–323. 63
- Pescetelli, N., Cebrian, M., and Rahwan, I. (2020). Beeme: Real-time internet control of situated human agents. *Computer*, 53(8):49–58. 35, 37
- Petri, L. (2010). Concept analysis of interdisciplinary collaboration. In *Nursing forum*, volume 45, pages 73–82. Wiley Online Library. 6, 28

- Pocock, M. J., Adriaens, T., Bertolino, S., Eschen, R., Essl, F., Hulme, P. E., Jeschke, J. M., Roy, H. E., Teixeira, H., and De Groot, M. (2024). Citizen science is a vital partnership for invasive alien species management and research. *Iscience*. 21
- Pollefeys, M. and Gool, L. V. (2002). From images to 3d models. *Communications of the ACM*, 45(7):50–55. 38, 43
- Puerta-Beldarrain, M., Gómez-Carmona, O., Casado-Mansilla, D., and López-de-Ipiña, D. (2022). Human-ai collaboration to promote trust, engagement and adaptation in the process of pro-environmental and health behaviour change. In *Proceedings of the International Conference on Ubiquitous Computing & Ambient Intelligence*, pages 381–392. Springer. 39
- Puerta-Beldarrain, M., Gómez-Carmona, O., Casado-Mansilla, D., and López-de-Ipiña, D. (2023). Human-ai collaboration to promote trust, engagement and adaptation in the process of pro-environmental and health behaviour change. In *Proceedings of the International Conference on Ubiquitous Computing & Ambient Intelligence (UCAmI 2022)*, pages 381–392. Springer International Publishing. 5, 21
- Qin, R.-J. and Yu, Y. (2024). Learning in games: a systematic review. *SCIENCE CHINA-INFORMATION SCIENCES*, 67(7). 49
- Radlinski, F., Balog, K., Diaz, F., Dixon, L., and Wedin, B. (2022). On natural language user profiles for transparent and scrutable recommendation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2863–2874. 56
- Rafter, R. and Smyth, B. (2005). Conversational collaborative recommendation—an experimental analysis. *Artificial Intelligence Review*, 24:301–318. 56
- Ramchurn, S. D., Wu, F., Jiang, W., Fischer, J. E., Reece, S., Roberts, S., Rodden, T., Greenhalgh, C., and Jennings, N. R. (2016). Human-agent collaboration for disaster response. *Autonomous Agents and Multi-Agent Systems*, 30(1):82–111. 35, 36, 37, 62

- Ramesh, A., Dhariwal, P., Nichol, A., Chu, C., and Chen, M. (2022). Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125*, 1(2):3. 43
- Ramesh, A., Pavlov, M., Goh, G., Gray, S., Voss, C., Radford, A., Chen, M., and Sutskever, I. (2021). Zero-shot text-to-image generation. In *International Conference on Machine Learning*, pages 8821–8831. PMLR. 43
- Ramos, G., Suh, J., Ghorashi, S., Meek, C., Banks, R., Amershi, S., Fiebrink, R., Smith-Renner, A., and Bansal, G. (2019). Emerging perspectives in human-centered machine learning. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, pages 1–8. 59
- Ren, H., Dai, H., Dai, Z., Yang, M., Leskovec, J., Schuurmans, D., and Dai, B. (2021). Combiner: Full attention transformer with sparse computation cost. *Advances in Neural Information Processing Systems*, 34:22470–22482. 43
- Reverberi, C., Rigon, T., Solari, A., Hassan, C., Cherubini, P., and Cherubini, A. (2022). Experimental evidence of effective human-ai collaboration in medical decision-making. *Scientific reports*, 12(1):14952. 39
- Rezwana, J. and Maher, M. L. (2022). Understanding user perceptions, collaborative experience and user engagement in different human-ai interaction designs for co-creative systems. In *Creativity and Cognition*, pages 38–48. 60
- Riveiro, M. and Thill, S. (2021). “that’s (not) the output i expected!” on the role of end user expectations in creating explanations of ai systems. *Artificial Intelligence*, 298:103507. 61
- Rosero, A., Dinh, F., de Visser, E. J., Shaw, T., and Phillips, E. (2021). Two many cooks: Understanding dynamic human-agent team communication and perception using overcooked 2. *arXiv preprint arXiv:2110.03071*. 49

- Rotman, D., Preece, J., Hammock, J., Procita, K., Hansen, D., Parr, C., Lewis, D., and Jacobs, D. (2012). Dynamic changes in motivation in collaborative citizen-science projects. In *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pages 217–226. 21
- Saeed, W. and Omlin, C. (2023). Explainable ai (xai): A systematic meta-survey of current challenges and future opportunities. *Knowledge-Based Systems*, 263:110273. 57, 58, 62
- Salas, E., Shuffler, M. L., Thayer, A. L., Bedwell, W. L., and Lazzara, E. H. (2015). Understanding and improving teamwork in organizations: A scientifically based practical guide. *Human resource management*, 54(4):599–622. 61
- Sarker, S., Susarla, A., Gopal, R., and Thatcher, J. B. (2024). Democratizing knowledge creation through human-ai collaboration in academic peer review. *Journal of the Association for Information Systems*, 25(1):158–171. 42
- Satcher, D. (2005). *Methods in community-based participatory research for health*. John Wiley & Sons. 21
- Sawaragi, T. (2005). Dynamical and complex behaviors in human-machine co-adaptive systems. *IFAC Proceedings Volumes*, 38(1):94–99. 54
- Schaefer, K. E., Perelman, B. S., Brewer, R. W., Wright, J. L., Roy, N., and Aksaray, D. (2018). Quantifying human decision-making: Implications for bidirectional communication in human-robot teams. In *International Conference on Virtual, Augmented and Mixed Reality*, pages 361–379. Springer. 61
- Scharowski, N. D. (2024). *Trust in artificial intelligence: understanding and calibrating trust and distrust in the human-AI interaction*. PhD thesis, University_of_Basel. 57

- Schelble, B., Flathmann, C., Canonico, L.-B., and Mcneese, N. (2021). Understanding human-ai cooperation through game-theory and reinforcement learning models. In *Proceedings of the Annual Hawaii International Conference on System Sciences*. 46, 49
- Schmitt, O. and Buschek, D. (2021). Characterchat: Supporting the creation of fictional characters through conversation and progressive manifestation with a chatbot. In *Creativity and Cognition*, pages 1–10. 38, 43
- Schmitt, V., Villa-Arenas, L.-F., Feldhus, N., Meyer, J., Spang, R. P., and Möller, S. (2024). The role of explainability in collaborative human-ai disinformation detection. In *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, pages 2157–2174. 57
- Schroder, A., Constantiou, I. D., Tuunainen, V., and Austin, R. D. (2022). Human-ai collaboration-coordinating automation and augmentation tasks in a digital service company. In *HICSS*, pages 1–10. 32
- Schulman, J., Zoph, B., Kim, C., Hilton, J., Menick, J., Weng, J., Uribe, J. F. C., Fedus, L., Metz, L., Pokorny, M., et al. (2022). Chatgpt: Optimizing language models for dialogue. *OpenAI blog*. 38, 42, 56
- Schurr, N., Marecki, J., Kasinadhuni, N., Tambe, M., Lewis, J. P., and Scerri, P. (2005a). The defacto system for human omnipresence to coordinate agent teams: The future of disaster response. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 1229–1230. 35
- Schurr, N., Marecki, J., Tambe, M., Scerri, P., Kasinadhuni, N., and Lewis, J. P. (2005b). The future of disaster response: Humans working with multiagent teams using defacto. In *AAAI spring symposium: AI technologies for homeland security*, pages 9–16. 35
- Sefeni, D., Johnson, M., and Lee, J. (2024). Game-theoretic approaches for step-wise controllable text generation in large language models. *Authorea Preprints*. 42

- Shao, L., Chen, B., Zhang, Z., Zhang, Z., and Chen, X. (2024). Artificial intelligence generated content (aigc) in medicine: A narrative review. *Mathematical Biosciences and Engineering*, 21(1):1672–1711. 41
- Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., and Althoff, T. (2023). Human–ai collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence*, 5(1):46–57. 38, 42, 63
- Sheard, J. K., Adriaens, T., Bowler, D. E., Büermann, A., Callaghan, C. T., Camprasse, E. C., Chowdhury, S., Engel, T., Finch, E. A., von Gönner, J., et al. (2024). Emerging technologies in citizen science and potential for insect monitoring. *Philosophical Transactions of the Royal Society B*, 379(1904):20230106. 21
- Sherson, J., Rafner, J., and Büyükgüzél, S. (2024). Operational criteria of hybrid intelligence for generative ai virtual assistants. In *HHAI 2024: Hybrid Human AI Systems for the Social Good*, pages 475–477. IOS Press. 28
- Shi, F., Zhou, F., Liu, H., Chen, L., and Ning, H. (2022). Survey and tutorial on hybrid human-artificial intelligence. *Tsinghua Science and Technology*, 28(3):486–499. 28
- Shirk, J. L., Ballard, H. L., Wilderman, C. C., Phillips, T., Wiggins, A., Jordan, R., McCallie, E., Minarchek, M., Lewenstein, B. V., Krasny, M. E., et al. (2012). Public participation in scientific research: a framework for deliberate design. *Ecology and society*, 17(2). 21
- Siau, K. and Wang, W. (2020). Artificial intelligence (ai) ethics: ethics of ai and ethical ai. *Journal of Database Management (JDM)*, 31(2):74–87. 52
- Skarlatidou, A., Ponti, M., Sprinks, J., Nold, C., Haklay, M., and Kanjo, E. (2019). User experience of digital technologies in citizen science. *Journal of Science Communication*, 18(01). 5, 21

- Smith, A., Kumar, V., Boyd-Graber, J., Seppi, K., and Findlater, L. (2018). Closing the loop: User-centered design and evaluation of a human-in-the-loop topic modeling system. In *23rd International Conference on Intelligent User Interfaces*, pages 293–304. 32
- Song, H. F., Abdolmaleki, A., Springenberg, J. T., Clark, A., Soyer, H., Rae, J. W., Noury, S., Ahuja, A., Liu, S., Tirumala, D., et al. (1999). V-mpo: On-policy maximum a posteriori policy optimization for discrete and continuous control. In *International Conference on Learning Representations*. 51
- Song, H. F., Abdolmaleki, A., Springenberg, J. T., Clark, A., Soyer, H., Rae, J. W., Noury, S., Ahuja, A., Liu, S., Tirumala, D., Heess, N. M. O., Belov, D., Riedmiller, M. A., and Botvinick, M. M. (2020). V-mpo: On-policy maximum a posteriori policy optimization for discrete and continuous control. *ArXiv*, abs/1909.12238. 48
- Song, X., Wang, E., Liu, W., Liu, Y., and Dong, Y. (2023). Fairness task assignment strategy with distance constraint in mobile crowdsensing. *CCF TRANSACTIONS ON PERVASIVE COMPUTING AND INTER-ACTION*, 5(2):184–205. 26
- Stevens, M. and D’Hondt, E. (2010). Crowdsourcing of pollution data using smartphones. In *Workshop on ubiquitous crowdsourcing*, pages 1–4. ACM. 2, 23
- Stock, P., Fan, A., Graham, B., Grave, E., Gribonval, R., Jegou, H., and Joulin, A. (2021). Training with quantization noise for extreme model compression. In *ICLR 2021-International Conference on Learning Representations*. 43
- Strouse, D., McKee, K., Botvinick, M., Hughes, E., and Everett, R. (2021). Collaborating with humans without human data. *Advances in Neural Information Processing Systems*, 34:14502–14515. 46, 48
- Tabari, N., Deshmukh, A. A., Kang, W.-C., Zamani, H., Gangadharaiyah, R., McAuley, J., and Karypis, G. (2024). First workshop on generative ai

- for recommender systems and personalization. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 6737–6738. 56
- Tahboub, K. A. (2019). Human-machine coadaptation based on reinforcement learning with policy gradients. In *2019 8th International Conference on Systems and Control (ICSC)*, pages 247–251. IEEE. 49, 54
- Talbot, J., Lee, B., Kapoor, A., and Tan, D. S. (2009). Ensemblematrix: interactive visualization to support machine learning with multiple classifiers. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1283–1292. 51
- Taneja, S. et al. (2024). User experience and interaction in information applications: Advanced human-machine interfaces. In *Algorithmic Approaches to Financial Technology: Forecasting, Trading, and Optimization*, pages 197–206. IGI Global. 4, 21
- Tao, L., Bowman, M., Zhang, J., and Zhang, X. (2021). Forming real-world human-robot cooperation for tasks with general goal. *IEEE Robotics and Automation Letters*, 7(2):762–769. 46, 47
- Tegen, A., Davidsson, P., Mihailescu, R.-C., and Persson, J. A. (2019). Collaborative sensing with interactive learning using dynamic intelligent virtual sensors. *Sensors*, 19(3):477. 38, 41
- Teng, M., Elmustafa, A., Akera, B., Bengio, Y., Radi, H., Larochelle, H., and Rolnick, D. (2024). Satbird: a dataset for bird species distribution modeling using remote sensing and citizen science data. *Advances in Neural Information Processing Systems*, 36. 19
- Termehchy, A. and Touri, B. (2015a). A signaling game approach to databases querying and interaction. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 361–364. 46, 50

- Termehchy, A. and Touri, B. (2015b). A signaling game approach to databases querying and interaction. In *Proceedings of the 2015 International Conference on The Theory of Information Retrieval*, pages 361–364. 50, 62
- Teso, S., Alkan, Ö., Stammer, W., and Daly, E. (2023). Leveraging explanations in interactive machine learning: An overview. *Frontiers in Artificial Intelligence*, 6:1066049. 57, 58
- Texler, O., Futschik, D., Kučera, M., Jamriška, O., Sochorová, Š., Chai, M., Tulyakov, S., and Sỳkora, D. (2020). Interactive video stylization using few-shot patch-based training. *ACM Transactions on Graphics (TOG)*, 39(4):73–1. 38, 43
- Thoppilan, R., De Freitas, D., Hall, J., Shazeer, N., Kulshreshtha, A., Cheng, H.-T., Jin, A., Bos, T., Baker, L., Du, Y., et al. (2022). Lamda: Language models for dialog applications. *arXiv preprint arXiv:2201.08239*. 38, 43
- Tolmeijer, S., Gadiraju, U., Ghantasala, R., Gupta, A., and Bernstein, A. (2021). Second chance for a first impression? trust development in intelligent system interaction. In *Proceedings of the 29th ACM Conference on user modeling, adaptation and personalization*, pages 77–87. 57
- Tong, Y., Zhou, Z., Zeng, Y., Chen, L., and Shahabi, C. (2020). Spatial crowdsourcing: a survey. *The VLDB Journal*, 29:217–250. 2, 22, 24
- Tutul, A. A., Nirjhar, E. H., and Chaspari, T. (2024). Investigating trust in human-ai collaboration for a speech-based data analytics task. *International Journal of Human-Computer Interaction*, pages 1–19. 58
- Van Den Bosch, K. and Bronkhorst, A. (2018). Human-ai cooperation to benefit military decision making. NATO. 62
- Van Zoelen, E. M., Barakova, E. I., and Rauterberg, M. (2020). Adaptive leader-follower behavior in human-robot collaboration. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pages 1259–1265. IEEE. 55

- Van Zoelen, E. M., Van Den Bosch, K., and Neerincx, M. (2021). Becoming team members: identifying interaction patterns of mutual adaptation for human-robot co-learning. *Frontiers in Robotics and AI*, 8, 54
- Vayadande, K., Pednekar, C. B., Khune, P. A., Prabhavalkar, V. S., and Dange, V. R. (2024). Gpt-3-and dall-e-powered applications: A complete survey. *How Machine Learning is Innovating Today's World: A Concise Technical Guide*, pages 329–341. 43
- Vicente, L. and Matute, H. (2023). The inherited bias effect: the propagation of artificial intelligence biases to human decisions. 6, 28, 29
- Vlachogianni, P. and Tselios, N. (2022). Perceived usability evaluation of educational technology using the system usability scale (sus): A systematic review. *Journal of Research on Technology in Education*, 54(3):392–409. 93
- Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., and Wagenknecht, K. (2021). *The science of citizen science*. Springer Nature. 1, 4, 18, 19, 64
- Vössing, M., Köhl, N., Lind, M., and Satzger, G. (2022). Designing transparency for effective human-ai collaboration. *Information Systems Frontiers*, 24(3):877–895. 59
- Wan, T., Yue, S., and Liao, W. (2021). Privacy-preserving incentive mechanism for mobile crowdsensing. *Security and Communication Networks*, 2021(1):4804758. 23, 26
- Wang, C. and Zhao, J. (2024). Role dynamic assignment of human-robot collaboration based on target prediction and fuzzy inference. *IEEE TRANSACTIONS ON INDUSTRIAL INFORMATICS*, 20(1):471–481. 46, 48
- Wang, D., Churchill, E., Maes, P., Fan, X., Shneiderman, B., Shi, Y., and Wang, Q. (2020). From human-human collaboration to human-ai collaboration: Designing ai systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*, pages 1–6. 60

- Wang, P. and Ding, H. (2024). The rationality of explanation or human capacity? understanding the impact of explainable artificial intelligence on human-ai trust and decision performance. *Information Processing & Management*, 61(4):103732. 57
- Wang, Y., Yingxin, L., Weilong, W., Maradapu Vera Venkata Sai, A., Cai, Z., et al. Mobile crowdsourcing based on 5g and 6g: A survey. 24
- Wang, Y., Zhu, J., Liu, R., and Jiang, Y. (2024). Enhancing recommendation acceptance: Resolving the personalization–privacy paradox in recommender systems: A privacy calculus perspective. *International Journal of Information Management*, 76:102755. 56
- Wessman, A. P. F., Thomas, S. E., and Rohiola, V. (2019). Digital archaeology and citizen science:: Introducing the goals of findsampo and the sualt project. *SKAS*, 2019(1):2–17. 19
- Whittaker, L., Kietzmann, T. C., Kietzmann, J., and Dabirian, A. (2020). “all around me are synthetic faces”: the mad world of ai-generated media. *IT Professional*, 22(5):90–99. 38, 43
- Wiethof, C. and Bittner, E. A. (2022). Toward a hybrid intelligence system in customer service: collaborative learning of human and ai. 51
- Wu, J., Gan, W., Chen, Z., Wan, S., and Lin, H. (2023a). Ai-generated content (aigc): A survey. *arXiv preprint arXiv:2304.06632*. 6, 28, 29
- Wu, S. A., Wang, R. E., Evans, J. A., Tenenbaum, J. B., Parkes, D. C., and Kleiman-Weiner, M. (2021). Too many cooks: Bayesian inference for coordinating multi-agent collaboration. *Topics in Cognitive Science*, 13(2):414–432. 49
- Wu, T., He, S., Liu, J., Sun, S., Liu, K., Han, Q.-L., and Tang, Y. (2023b). A brief overview of chatgpt: The history, status quo and potential future development. *IEEE/CAA Journal of Automatica Sinica*, 10(5):1122–1136. 42, 43, 56

- Xing, X., Li, W., Yuan, S., and Li, Y. (2024). Fuzzy logic-based arbitration for shared control in continuous human-robot collaboration. *IEEE Transactions on Fuzzy Systems*. 46, 48
- Xu, Y., Ohmoto, Y., Okada, S., Ueda, K., Komatsu, T., Okadome, T., Kamei, K., Sumi, Y., and Nishida, T. (2012). Formation conditions of mutual adaptation in human-agent collaborative interaction. *Applied Intelligence*, 36(1):208–228. 46, 47, 54
- Xu, Y., Ueda, K., Komatsu, T., Okadome, T., Hattori, T., Sumi, Y., and Nishida, T. (2009). Woz experiments for understanding mutual adaptation. *Ai & Society*, 23(2):201–212. 47
- Yang, D., Zhou, Y., Zhang, Z., Li, T. J.-J., and LC, R. (2022). Ai as an active writer: Interaction strategies with generated text in human-ai collaborative fiction writing. In *Joint Proceedings of the ACM IUI Workshops*, volume 10. CEUR-WS Team. 38, 42
- Yang, Q., Suh, J., Chen, N.-C., and Ramos, G. (2018). Grounding interactive machine learning tool design in how non-experts actually build models. In *Proceedings of the 2018 designing interactive systems conference*, pages 573–584. 57, 58
- Yang, R., Shin, E., Newman, M. W., and Ackerman, M. S. (2015). When fitness trackers don’t fit’ end-user difficulties in the assessment of personal tracking device accuracy. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 623–634. 57
- Yang, Y., Kandogan, E., Li, Y., Sen, P., and Lasecki, W. S. (2019). A study on interaction in human-in-the-loop machine learning for text analytics. In *IUI Workshops*. 32, 34
- Yeh, C., Ramos, G., Ng, R., Huntington, A., and Banks, R. (2024). Ghostwriter: Augmenting collaborative human-ai writing experiences through personalization and agency. *arXiv preprint arXiv:2402.08855*. 32, 33

- Yu, C., Gao, J., Liu, W., Xu, B., Tang, H., Yang, J., Wang, Y., and Wu, Y. (2023). Learning zero-shot cooperation with humans, assuming humans are biased. In *The Eleventh International Conference on Learning Representations*. 46, 49
- Yuan, Y., Liu, H., Liu, X., Huang, Q., Plumbley, M. D., and Wang, W. (2024). Retrieval-augmented text-to-audio generation. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 581–585. IEEE. 44
- Yucel, F. and Bulut, E. (2020). User satisfaction aware maximum utility task assignment in mobile crowdsensing. *Computer Networks*, 172:107156. 23, 25
- Zakershahrak, M. (2021). *Towards Human-Machine Symbiosis: Design for Effective AI Facilitation*. PhD thesis, Arizona State University. 35, 36
- Zbyszynski, M., Di Donato, B., and Tanaka, A. (2019). The effect of co-adaptive learning & feedback in interactive machine learning. 63
- Zeng, C., Yang, C., and Chen, Z. (2020). Bio-inspired robotic impedance adaptation for human-robot collaborative tasks. *Science China Information Sciences*, 63(7):1–10. 55
- Zerilli, J., Bhatt, U., and Weller, A. (2022). How transparency modulates trust in artificial intelligence. *Patterns*, page 100455. 57
- Zhan, Y., Xia, Y., and Zhang, J. (2018). Incentive mechanism in platform-centric mobile crowdsensing: A one-to-many bargaining approach. *Computer Networks*, 132:40–52. 23, 25
- Zhang, C., Zhang, C., Zhang, M., and Kweon, I. S. (2023a). Text-to-image diffusion model in generative ai: A survey. *arXiv preprint arXiv:2303.07909*. 43

- Zhang, C., Zhang, C., Zheng, S., Zhang, M., Qamar, M., Bae, S.-H., and Kweon, I. S. (2023b). A survey on audio diffusion models: Text to speech synthesis and enhancement in generative ai. *arXiv preprint arXiv:2303.13336*, 2. 44
- Zhang, J., Ding, Q., Li, B., and Ye, X. (2023c). Bidirectional k-nearest neighbor spatial crowdsourcing allocation protocol based on edge computing. *PeerJ Computer Science*, 9:e1244. 90
- Zhang, Q., Kang, Y., and Roth, R. (2023d). The ethics of ai-generated maps: Dall·e 2 and ai’s implications for cartography (short paper). In *12th International Conference on Geographic Information Science (GIScience 2023)*. Schloss Dagstuhl-Leibniz-Zentrum für Informatik. 38, 43
- Zhang, R., Duan, W., Flathmann, C., McNeese, N., Freeman, G., and Williams, A. (2023e). Investigating ai teammate communication strategies and their impact in human-ai teams for effective teamwork. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2):1–31. 61
- Zhang, S., Wang, X., Zhang, W., Chen, Y., Gao, L., Wang, D., Zhang, W., Wang, X., and Wen, Y. (2024a). Mutual theory of mind in human-ai collaboration: An empirical study with llm-driven ai agents in a real-time shared workspace task. *arXiv preprint arXiv:2409.08811*. 46, 48
- Zhang, S., Yu, J., Xu, X., Yin, C., Lu, Y., Yao, B., Tory, M., Padilla, L. M., Caterino, J., Zhang, P., et al. (2024b). Rethinking human-ai collaboration in complex medical decision making: A case study in sepsis diagnosis. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, pages 1–18. 38, 39
- Zhang, Y., Ying, Z., and Chen, C. P. (2022). Achieving privacy-preserving multitask allocation for mobile crowdsensing. *IEEE Internet of Things Journal*, 9(18):16795–16806. 23, 26
- Zhao, M., Simmons, R., and Admoni, H. (2022). The role of adaptation in collective human-ai teaming. *Topics in Cognitive Science*. 53

- Zheng, Q., Tang, Y., Liu, Y., Liu, W., and Huang, Y. (2022). Ux research on conversational human-ai interaction: A literature review of the acm digital library. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–24. 60
- Zhou, H., Wei, D., Chen, Y., and Wu, F. (2021). Promoting mutual adaptation in haptic negotiation using adaptive virtual fixture. *Industrial Robot: the international journal of robotics research and application*. 46, 47, 54
- Zhou, R., Li, Z., and Wu, C. (2017). A truthful online mechanism for location-aware tasks in mobile crowd sensing. *IEEE Transactions on Mobile Computing*, 17(8):1737–1749. 23, 25
- Zoelen, E. M. v., Cremers, A., Dignum, F. P., Diggelen, J. v., and Peeters, M. M. (2020). Learning to communicate proactively in human-agent teaming. In *International Conference on Practical Applications of Agents and Multi-Agent Systems*, pages 238–249. Springer. 62



Examples of Campaign Shapes

In this appendix, we present several examples of campaigns with different shapes. These shapes are used to define the area where the tasks are distributed. The examples presented in this appendix are a cross shape, a rectangular shape, an elliptical shape, a square shape, a diamond shape and a prominent rectangular shape. The examples are shown in Figure A.1.



(a) An example of a campaign with a cross shape.



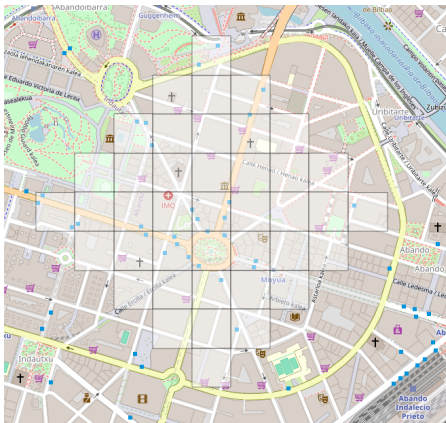
(b) An example of a campaign with a rectangular shape.



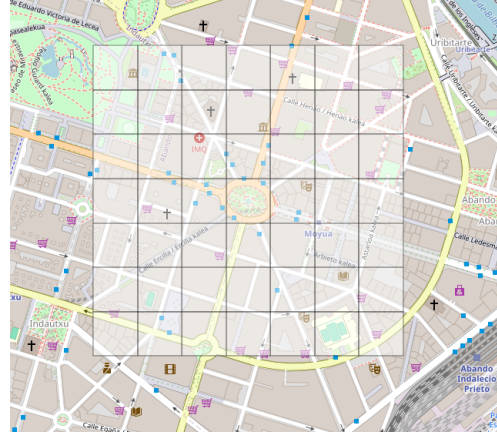
(c) An example of a campaign with an elliptical shape.



(d) An example of a campaign with a square shape.



(e) An example of a campaign with a diamond shape.



(f) An example of a campaign with a prominent rectangular shape.

Figure A.1: Several campaigns with different shapes.

Deusto Use Case Questionnaire

The questionnaire presented here is designed to gather feedback on the use case overall experience, satisfaction, perceived personalization of the Telegram bot, and potential future adoption. This feedback will be utilised to enhance the bot's performance, improve user experience, and better integrate users into the SC problem in future iterations. This questionnaire is inspired by the Technology Acceptance Model (TAM) (Davis et al., 1989) and the System Usability Scale (SUS) (Grier et al., 2013). TAM is a theoretical framework for evaluating users' acceptance of a technology based on two key factors: perceived usefulness (how much users believe the technology will enhance their performance) and perceived ease of use (how easy they find it to use). These factors shape users' attitudes, intentions, and actual adoption of the technology. TAM is widely applied in research to predict and understand user behaviour toward new tools and systems. SUS, developed by John Brooke in 1986, is a standardized questionnaire that assesses the perceived usability of a system, product, or service, such as an application, website, or software. It is one of the most widely used methods for evaluating user experience.



Figure B.1: The logo, created by a participant for this experiment, symbolizes both the concept of the experiment and the potato omelet that was offered to the participants at the end of the experiment.

In this appendix, we present this questionnaire used in the Deusto Uses case before the Deusto campaign. The results are analysed in Subsection 4.4.2 and 5.3.2. The questionnaire comprises seven sections: demographic information, satisfaction and activity information, adoption information, perceived personalization information, recommendations and feedback information, feedback about the Telegram bot, and a comparison of conducted campaigns. Before all these questions, all the users accepted the confect form to participate in the use case and were given the authority to use this data for the research.

B.1 Demographic information

This section aims to gather information about the participants' demographic characteristics, such as age, occupation and so on. This information will help us understand the profile of the participants and how it may influence their experience with the Telegram bot. The asked question and their possible response are presented below.

Question 1 - What is your age?

- | | | |
|--------------------------------|--------------------------------|----------------------------------|
| ☐ 12-14 | ☐ 19-35 | ☐ 66-70 |
| ☐ 15-16 | ☐ 36-50 | ☐ 71-75 |
| ☐ 17-18 | ☐ 51-65 | ☐ Over 75 |

Question 2 - What is your gender?

- | | |
|--|---------------------------------------|
| ☐ Female | ☐ Male |
| ☐ I do not want to answer | ☐ Others _____ |

Question 3 - What is your level of education?

- | | |
|---|--|
| ☐ Secondary school | ☐ Master's degree |
| ☐ High school diploma | ☐ PhD |
| ☐ Vocational training | ☐ Others _____ |
| ☐ Undergraduate degree | |

Question 4 - Could you indicate your work status?

- | | |
|--|---------------------------------------|
| ☐ Unemployed | ☐ Retired |
| ☐ Self-employed | ☐ Student |
| ☐ Employed (private sector) | ☐ Others _____ |
| ☐ Employed (public sector) | |

Question 5 - What is your level of knowledge in the usage of digital tools?

- ☐ Zero level (I do not use any digital tool including a smartphone)
- ☐ Basic level (I use WhatsApp, I browse a little through websites, ...)
- ☐ Intermediate level (I browse websites, I watch videos on YouTube, ...)
- ☐ Advanced level (I make purchases electronically, I use spreadsheets, ...)

B.2 Satisfactory and activity information

This section aims to gather information about participants' satisfaction with the activity and its perceived usefulness. This will help us understand their views on the activity and how it might affect their future adoption of the Telegram bot. The questions and possible responses are provided below.

Question 1 - I think this activity was interesting to me.

	1	2	3	4	5	
Strongly disagree	☐	☐	☐	☐	☐	Strongly agree

Question 2 - I think this activity was useful to me.

	1	2	3	4	5	
Strongly disagree	☐	☐	☐	☐	☐	Strongly agree

Question 3 - I think that this activity dealt with an important topic for me.

	1	2	3	4	5	
Strongly disagree	☐	☐	☐	☐	☐	Strongly agree

Question 4 - Comments or recommendations regarding this activity that you would like to share.

B.3 Adoption information

This section aims to gather information about participants' intentions to adopt the Telegram bot in the future. This will help us understand their perceptions of the bot and how these perceptions might influence their future adoption. The questions and possible responses are provided below.

Question 1 - Would you recommend the recommendation system to others?
Or to use in other applications or propose?

☐ Yes

☐ Maybe

☐ No

Question 2 - Do you see yourself continuing to use the recommender system
in the future?

☐ Yes

☐ Maybe

☐ No

Question 3 - If no, what would be the primary reason for discontinuing the
use of the recommender system?

Question 4 - How likely are you to adopt a Telegram bot or participate in
the next experiment?

	1	2	3	4	5	
Very unlikely	☐	☐	☐	☐	☐	Very likely

Question 5 - What factors influence your intention to adopt the bot services
in your daily life?

B.4 Perceived personalization information

This section aims to gather information about participants' perceptions of the personalization in the recommendations provided by the Telegram bot. This will help us understand how participants view the bot's personalization and how it might impact their future adoption. The questions and possible responses are provided below.

Question 1 - How personalised do you find the recommendations provided by the Telegram bot?

Not personalised at all 1 2 3 4 5 Extremely personalised
 ☐ ☐ ☐ ☐ ☐

Question 2 - To what extent do you feel that the recommendations align with your interests/preferences?

Not at all 1 2 3 4 5 Completely
 ☐ ☐ ☐ ☐ ☐

Question 3 - How do you perceive the role of personalised recommendations in enhancing your overall experience with the Telegram bot?

Not significant 1 2 3 4 5 Extremely significant
 ☐ ☐ ☐ ☐ ☐

Question 4 - Are there any specific instances where you felt the recommendations were particularly well-tailored to your needs?

Question 5 - What improvements, if any, would you suggest to enhance the personalization of recommendations provided by the Telegram bot?

B.5 Recommendations and feedback information

This section aims to gather information about participants' experiences with the recommender system and the Telegram bot. This information will help us understand their perceptions of the system and how it might influence their future adoption. The questions and possible responses are presented below.

Question 1 - Is there anything else you would like to share about your experience with the recommender system?

Question 2 - What works well?

Question 3 - What could be improved/added?

Question 4 - What is missing?

Question 5 - How would you improve it?

B.6 Feedback about the Telegram Bot

This section aims to gather information about participants' experiences with the Telegram bot. This information will help us understand their perceptions of the bot and how it might influence their future adoption. The questions and possible responses are presented below.

Question 1 - What works well?

Question 2 - What does not work?

Question 3 - What could be improved/added?

Question 4 - What is missing?

Question 5 - How would you improve it?

B.7 Comparison of Conducted Pilots

This section aims to gather insights into participants' experiences with both the first and second pilots. Only those who participated in both pilots are required to answer these questions. The information collected will help us understand participants' perceptions of the system and how it might influence their future adoption. The questions and their possible responses are presented below.

Question 1 - Have you noticed any difference between the recommendations obtained in the first and second experiment?

Question 2 - Which recommender do you find most usable in your day-to-day life?

- ☐ First campaign/experiment
- ☐ Second campaign/experiment
- ☐ I don't know

Question 3 - What do you think could be the determining factor for this choice? Explain your last answer.

B.8 Thank you for your participation

Thanks for your participation in the campaign and your paciést to respond to this questionnaire. Your feedback is very important to us.

Telegram Bot

This appendix presents the Telegram bot used in the Deusto use case for user interaction. Throughout this section, the interaction process with the Telegram bot employed in the second Deusto use case, where the Bio-VTAS was implemented, is detailed.

The bot was designed to provide cell recommendations to users, allowing them to accept or reject suggestions. Additionally, it offered information on campaign progress, including visual updates, and guided users through the steps of the interactive workflow.

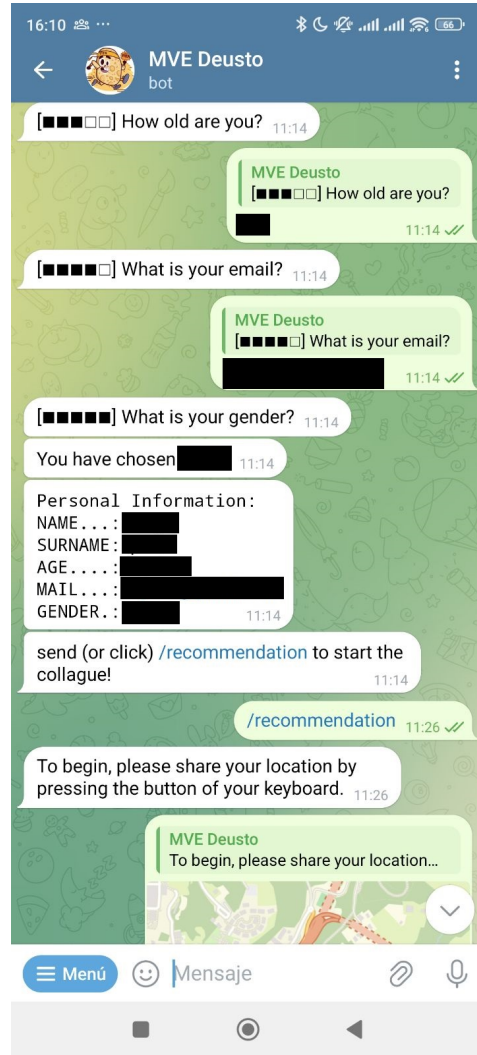
Through text commands, users could perform various actions, such as requesting recommendations, accepting or rejecting suggestions, viewing a map composed of photographs, and sending images of the accepted recommended cells. The bot interpreted these commands and responded efficiently, enhancing the user experience.

The explanation of the bot provided in this section is divided into three parts: Section C.1 explains the initial collection of personal data carried out by the bot. Section C.2 describes the core interaction process between users and the bot during the photo collection process. Finally, Section C.3 explains how the bot reacts when the task-completion goal has been achieved, informing users that there are no further recommendations available.

C.1 Personal Information Collection



(a) Initial interaction of the Telegram bot requesting personal information.



(b) Continuation of the interaction where the Telegram bot confirms and completes personal information.

Figure C.1: Telegram bot requesting and verifying personal information.

During the initial interaction, and only in the first iteration with the bot, users are asked to provide some basic personal information, such as their first name, last name, age, email, and gender. These questions are presented sequentially, as shown in Figures C.1a and C.1b.

Once this data has been collected, the bot sends a summary of the gathered information for the user to verify its accuracy. It then requests confirmation to ensure the details provided are correct. If the user confirms, the bot guides them on how to proceed and removes the confirmation question from the chat to maintain a cleaner and more organized interface, improving navigation. As shown in Figure C.1b, the summary of personal information remains visible while the confirmation question has been removed to enhance clarity and comprehension.

If the user wishes to modify their personal information, they can do so at any time using the command */personal_information*, accessible from the lower-left menu. This menu operation was previously explained to users during training.

Finally, the bot sends a message instructing the user that to request a recommendation, they must enter the command */recommendation*, available in the menu buttons. Section C.2 demonstrates how the user makes this request and provides a detailed explanation of the image capture process, which will contribute to the final collage for the Deusto use case.

C.2 Bot Interaction Process for Image Capture

This section describes the interaction process between the bot and the user for image capture within the Deusto use case. The complete interaction flow, from requesting a recommendation to confirming the submitted image, can be observed in Figures C.2, C.3, C.4, and C.5.

The process begins when the user requests a recommendation from the bot using the command */recommendation*. The execution of this command is shown in Figure C.1b, while the bot's full response can be seen at the end of this image and at the beginning of Figure C.2.

Once the command is received, the bot requests the user's location. The keyboard is replaced with a button that allows direct location sharing with a single click. Once the user shares their location, the bot generates three cell recommendations, as shown in Figures C.2 and C.3.

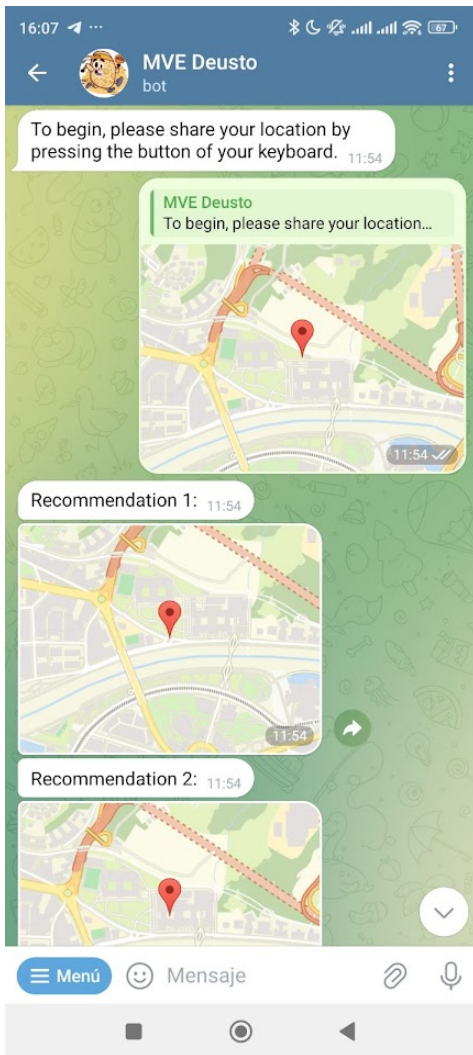


Figure C.2: The bot provides the user with three location-based recommendations.

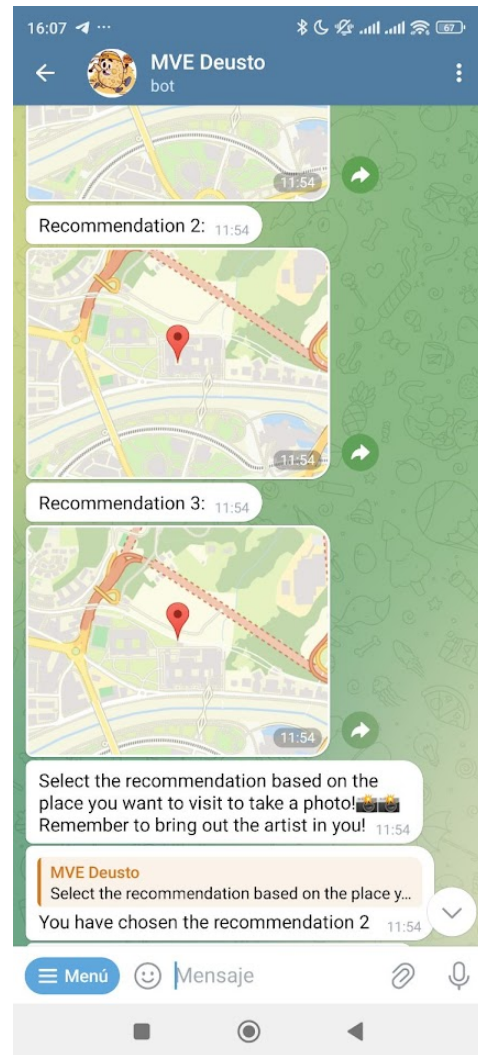


Figure C.3: The user selects "Recommendation 2" and the bot sends the corresponding location.

The three recommended cells are labelled as "Recommendation 1", "Recommendation 2" and "Recommendation 3". Simultaneously, the user's keyboard is replaced with three buttons, each corresponding to one of the available options, allowing selection with a single click.

When the user selects a recommendation, the bot interprets this as an acceptance and requests an image capture at the chosen cell. Figures C.3 and C.4 illustrate how the user accepts "Recommendation 2". To facilitate

navigation, the bot resends the location of the selected cell, enabling the user to reach the specified location easily.

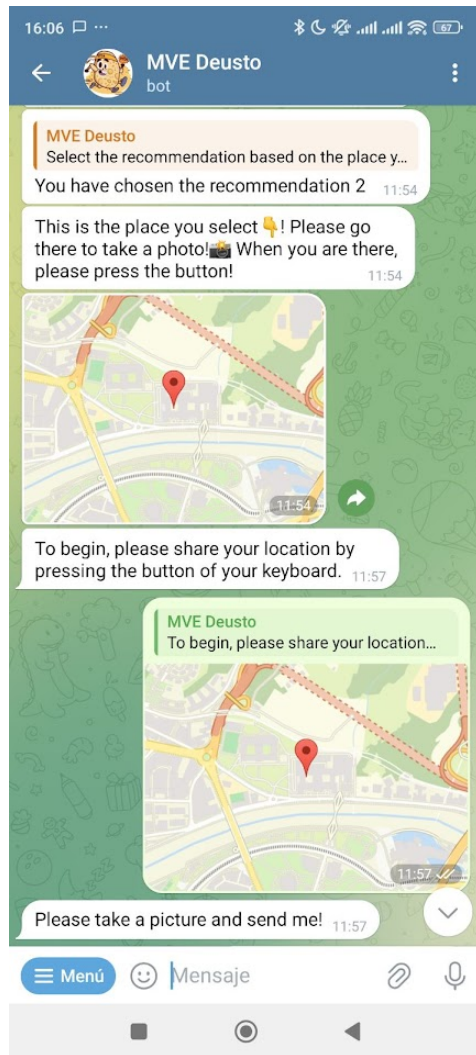


Figure C.4: The bot verifies that the user's location matches the selected recommendation before requesting an image.



Figure C.5: The user submits the captured image, and the bot confirms task completion.

After sending the location, the bot instructs the user to share their location again upon arrival at the selected cell using the keyboard button. This step allows the bot to verify the user is at the correct location before proceeding with the photo capture.

If the user's submitted location matches the accepted cell, as shown in Figure C.4, the bot requests an image capture. If the location does not match, the bot resends the correct location and instructs the user to go to the correct place before sending the photo.

The user's image submission is shown in Figure C.5. Once the bot receives the photo, it confirms successful task completion. Additionally, it provides a button that the user can press to view the map and track real-time collage progress. Furthermore, it invites the user to continue participating in the campaign by submitting more images if they wish.

C.3 End of the Experiments

This is a special case where the campaign's task-completion goal has already been achieved, and the user has no remaining tasks. In this case, as seen in Figure C.6, the bot sends a message informing the user that no further recommendations are available close to him/her.

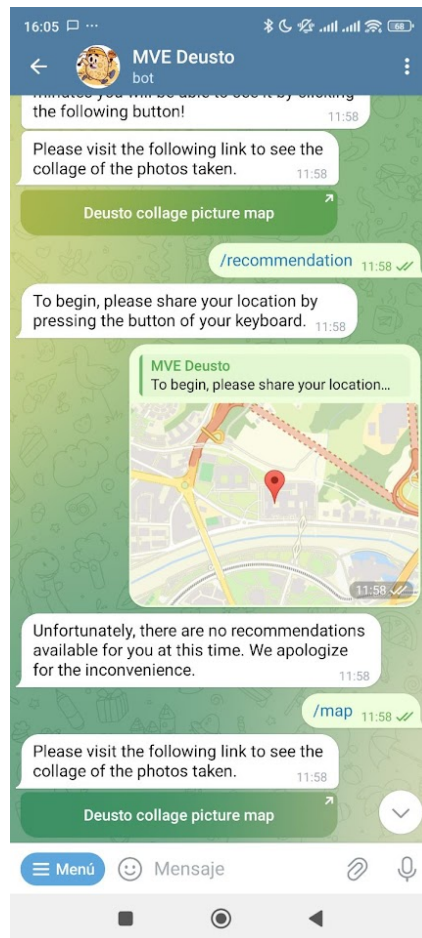


Figure C.6: The bot informs the user that no further recommendations are available, as the campaign’s task-completion goal has been achieved.

Deusto personal face-to-face interview

This appendix presents the questionnaire administered face-to-face with participants in the use cases at the University of Deusto. The primary goal of these interviews is to gain a detailed understanding of each participant's unique perspective on interacting with the technology used in citizen science, to enhance their engagement and comfort. The feedback obtained will be instrumental in developing new strategies to increase user involvement and motivation in Citizen Science projects.

The interviews were conducted individually with each participant, ensuring that each conversation was private and focused on the personal perspective of the user. Since the participants are the same individuals who took part in the Deusto use case to test the Bio-VTAS, they were first presented with the collage created during this campaign (Figure D.1) as a token of appreciation.

All interviews were recorded, and the analysis of these interviews is presents in Section 6.1.

The questionnaire is organized (and presented here) into five distinct sections: usage patterns, frequency of use, feedback, motivation, and interview



Figure D.1: A picture of the collage created by the users during the Deusto use case to evaluate the Bio-VTAS, which was presented to them at the beginning of the interview.

conclusions. Each section is designed to elicit specific information relevant to understanding and improving user experience in future developments.

D.1 Introduction to the interview

First of all, thank you very much for participating in this test. If you do not mind, we will record the audio to better analyse the responses later. The objective of these interviews is to understand what users need in their interaction with the technology used in citizen science. Citizen science seeks to involve ordinary people in the scientific process so they can help find solutions to their problems or identify and quantify such problems. The experiment conducted last Friday was a proof of concept, but similar tools can be used in citizen science. For example, in Copenhagen, there are citizen science projects to count birds and track migratory processes, detect inaccessible areas for

people with disabilities, create new environmental laws, etc. The goal of these projects is usually to do science for the citizens so that citizens can generate new laws, change things, or expand their knowledge on a topic of interest to them.

In this interview, we are not looking to focus on this development or application, but rather on what you need to get involved. Therefore, we want to know your opinion on how this interaction should be. Do you have any questions so far? Then let's begin. This interview has 5 blocks of questions.

D.2 Block 1 - Way of use

Let's start with the first block. This block is related to the way you would use the system.

Question 1 – How do you think you would use this system in your daily life? Would it be during your free time, while getting around the city, on your way to work, to explore new places, or a mix depending on the situation?

Question 2 – When using the system, would you prefer to visit nearby places, go further away, pick a specific area yourself, or have no restrictions?

Question 3 – Would you like the system to learn from your choices and suggest options based on what you've liked before?

Question 4 – Do you prefer having multiple options to choose from?

Question 5 – Do you think your preferences for distance and usage might change in the future?

D.3 Block 2 - Frequency of use

We finished the first block of questions and now we are going to start the second block of questions that are related to the frequency of use.

Question 1 – How often do you think you would use this system?

Question 2 – Would you like the system to notify you when it has something relevant for you, or would you rather decide when to use it yourself? Would you want to set your own rules for when you get notifications?

Question 3 – Do you think how often you use the system might change in the future? What factors do you think could influence this change?

D.4 Block 3 - Feedback

We are going to start the third block of questions that are related to the feedback you would like to give.

Question 1 – Would you like to rate how much you liked the task, the distance, and the place?

Question 2 – How would you prefer to give feedback? For example, a 5-star rating, a thumbs-up, or writing your own comments?

Question 3 – How often would you like to give feedback? Every few days, once a month, or something else?

D.5 Block 4 - Motivation

We are going to start the fourth block of questions that are related to the motivation you would like to have.

Question 1 – What would keep you interested in using the system over time? For example:

- Tracking your participation and sharing it on social media (WhatsApp, Telegram, X, TikTok, Facebook, etc.)
- Competing with others on a ranking list

- A built-in social network
- Getting rewards like interesting activities or useful information

Question 2 – What would encourage you to try new ways of using the system beyond what you’re used to? Or do you think your habits wouldn’t change?

D.6 Block 5 - Conclusions of the interview

We are nearing the end of the interview. This final block of questions serves as a summary of the interview.

Question 1 – Out of everything we’ve talked about (how to use the system, giving feedback, etc.), what do you think is most important?

Question 2 – Is there anything about using the system that we haven’t discussed but should?

Question 3 – If everything we talked about works as expected, would that make you more likely to keep using the system long-term?

Question 4 – Do you think your usage habits will change over time? If so, how? For example, would you change how often you use it, how you give feedback, or what you use it for?

Question 5 – Overall, do you see yourself using this system in the long run? Is there anything else you’d like to share about your experience or this interview?

Lastly, is there anything else you would like to share? Anything you want to comment on, whether about the experiment or this interview?

Thank you for your time and valuable feedback.

Visual representation of the LFI metric through the experiments

This appendix presents the visual representation of the LFI metric for the experiments conducted in this thesis. The LFI (Eq. 4.4) metric serves as a key performance indicator, measuring the degree of equitable task distribution achieved by the algorithm. The higher number of colours in the campaign visualization corresponds to a higher value of the LFI, as it signifies a more diverse range of tasks performed in each cell. All pictures are obtained at the end of the first slot of the campaign.

E.1 Experiment 1

This section provides the visual representation of the LFI metric for the first experiment, with 100, 75, 50, and 25 users, and for all algorithms used throughout this thesis. In these simulations, the User Task Completion Probability was set at 75%, reflecting a high level of user commitment to their assigned tasks, while the User Availability Probability was fixed at 50%.

E.1.1 100 Users

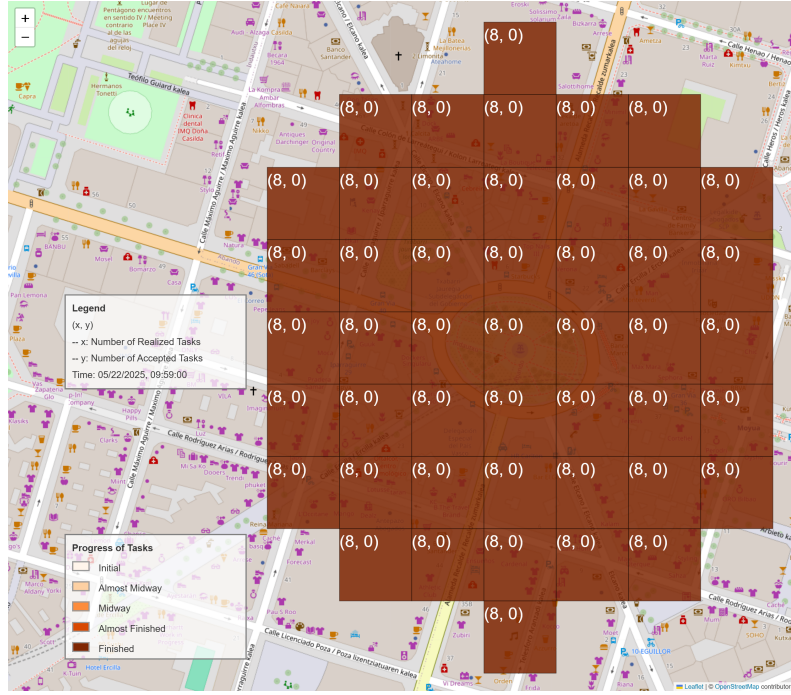


Figure E.1: Hybrid-VTAS - 100 Users

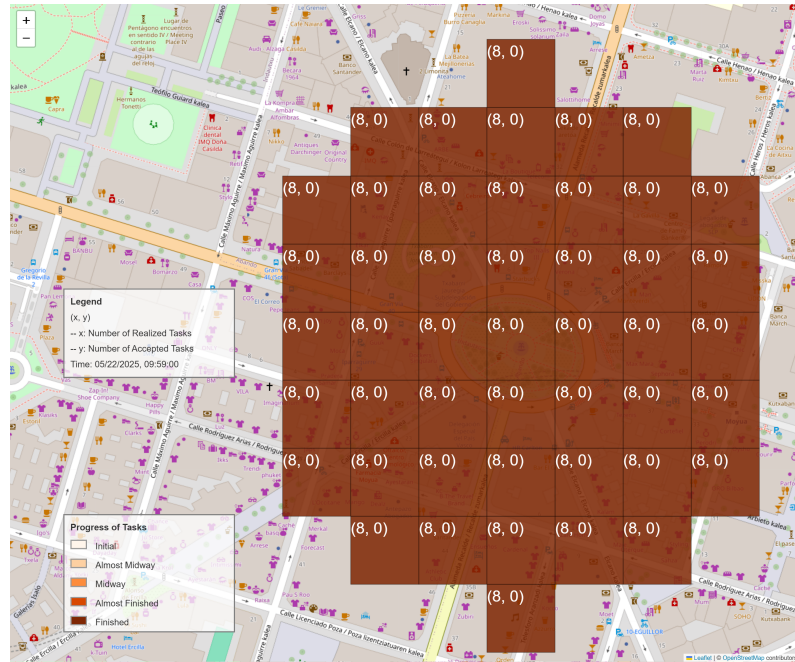


Figure E.2: VTAS - 100 Users

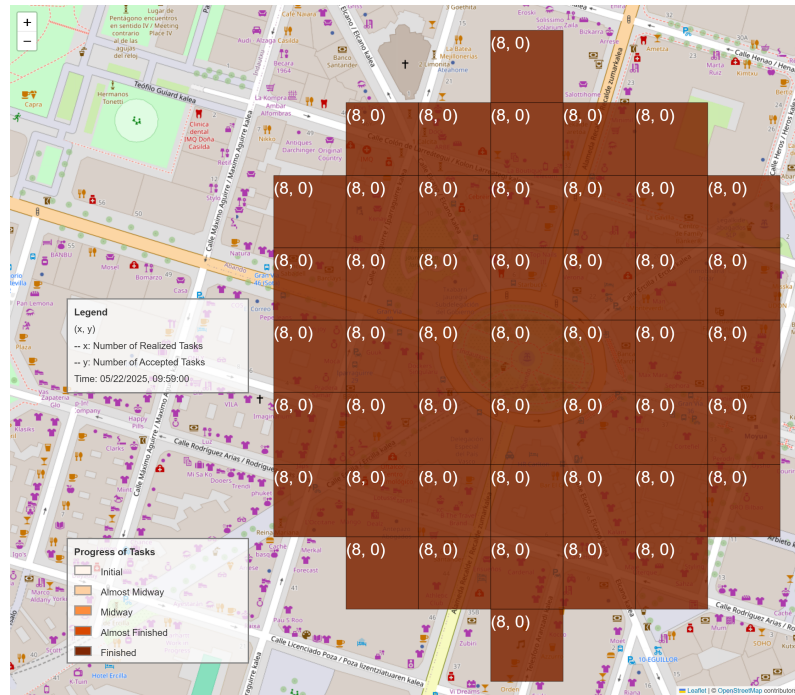


Figure E.3: Bio-VTAS - 100 Users

E.1.2 75 Users

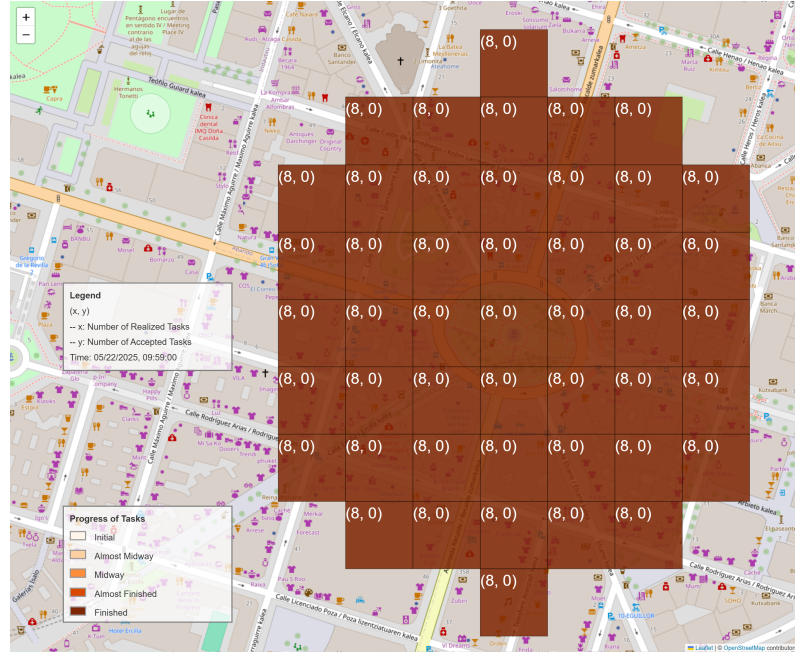


Figure E.6: Hybrid-VTAS - 75 Users

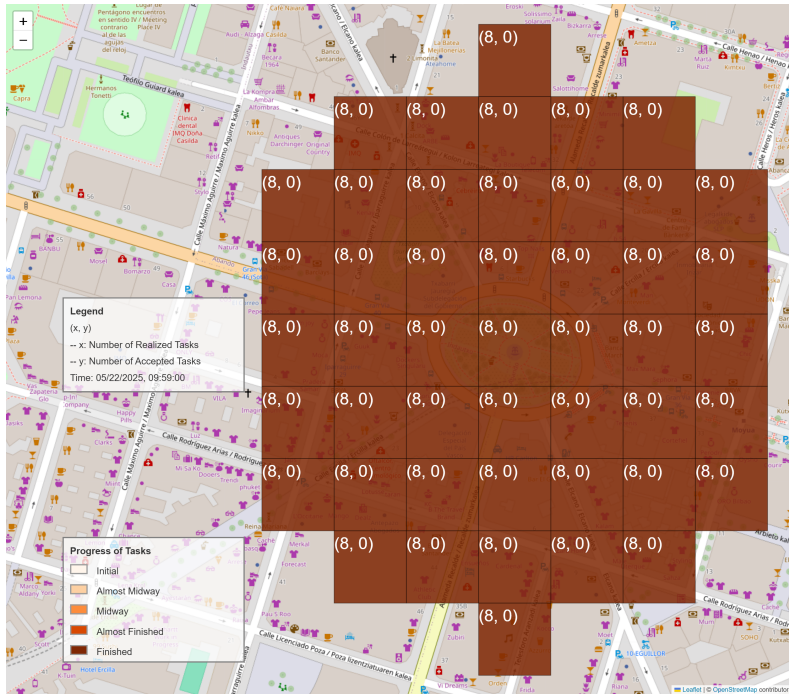


Figure E.7: VTAS - 75 Users

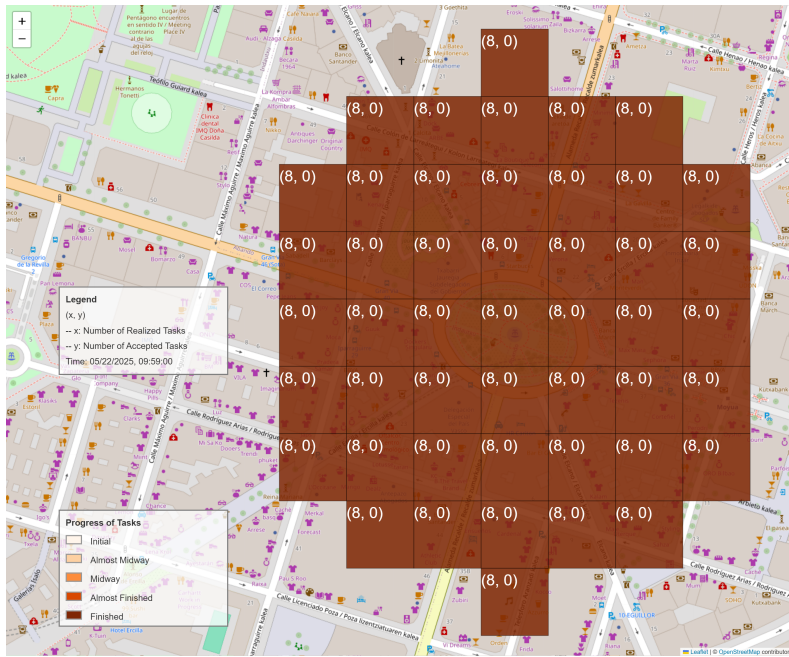


Figure E.8: Bio-VTAS - 75 Users

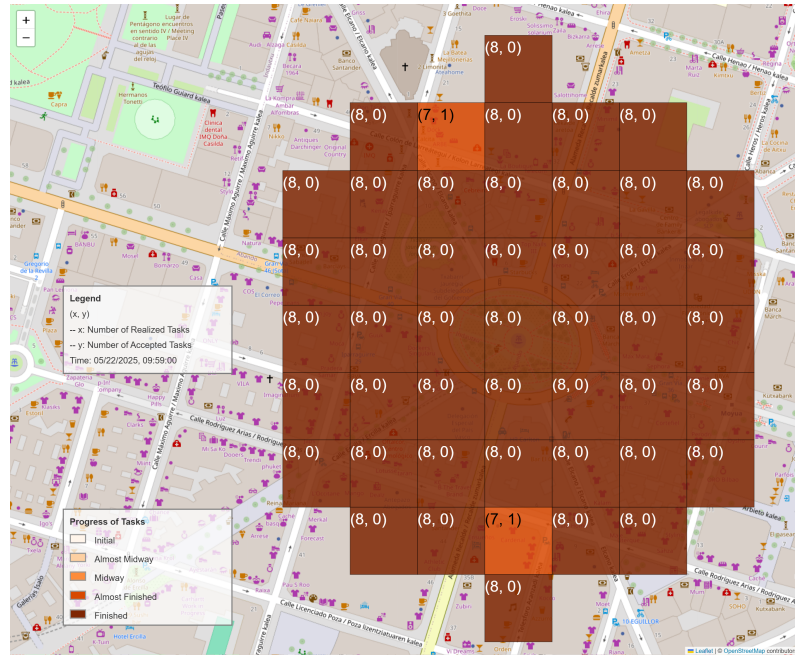


Figure E.9: LNN - 75 Users

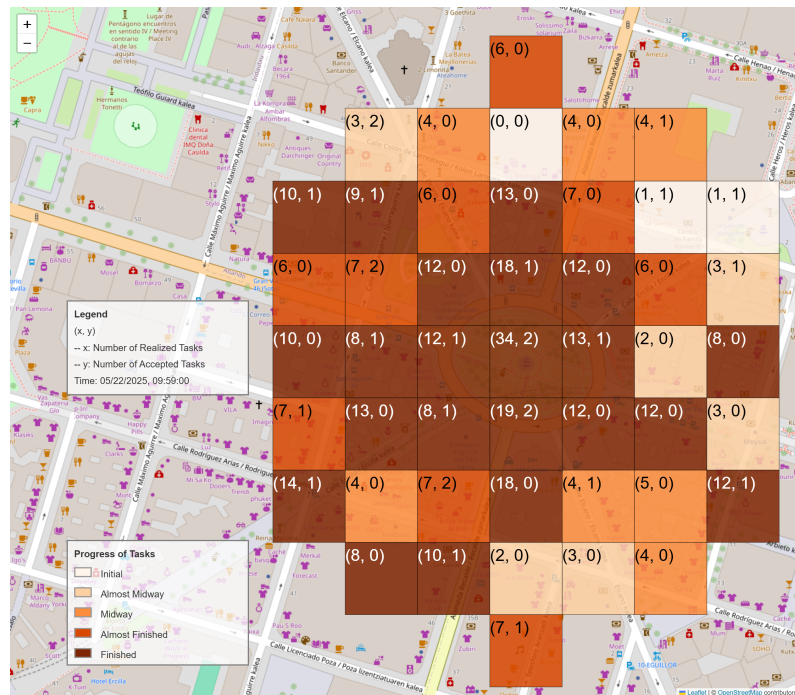


Figure E.10: NN - 75 Users

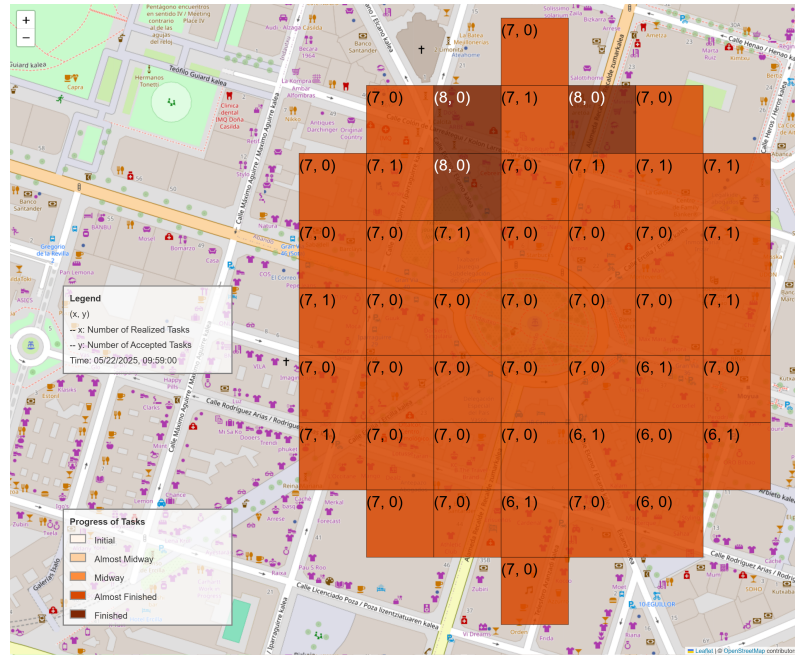


Figure E.12: VTAS - 50 Users

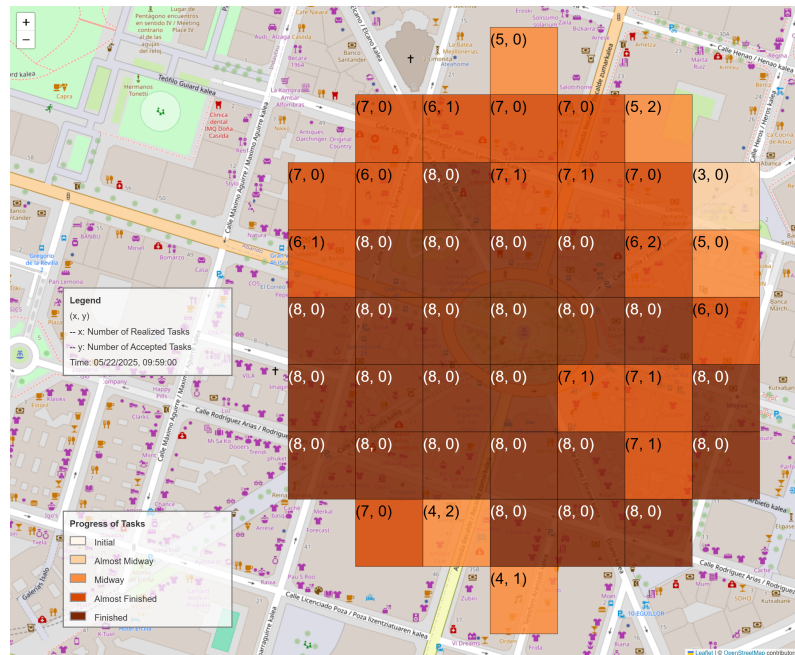


Figure E.13: Bio-VTAS - 50 Users

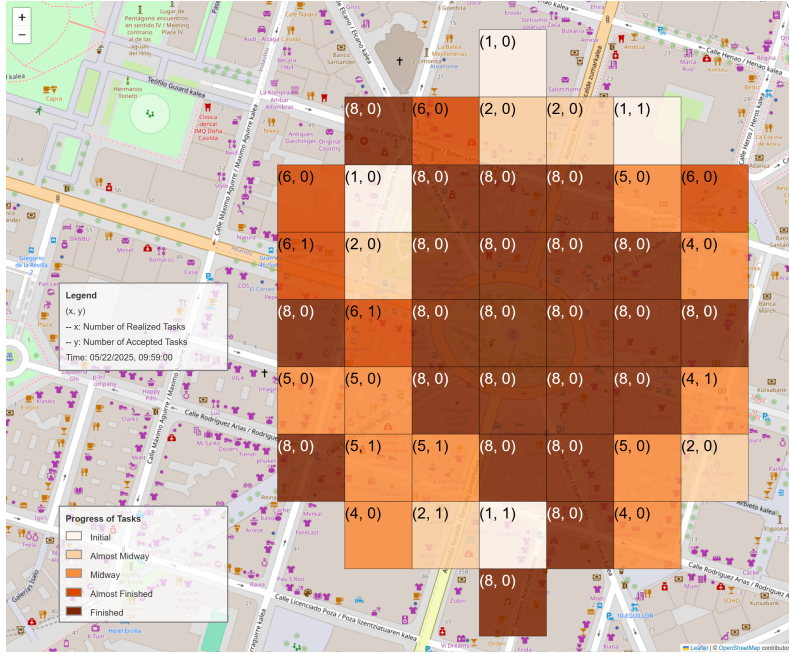


Figure E.14: LNN - 50 Users

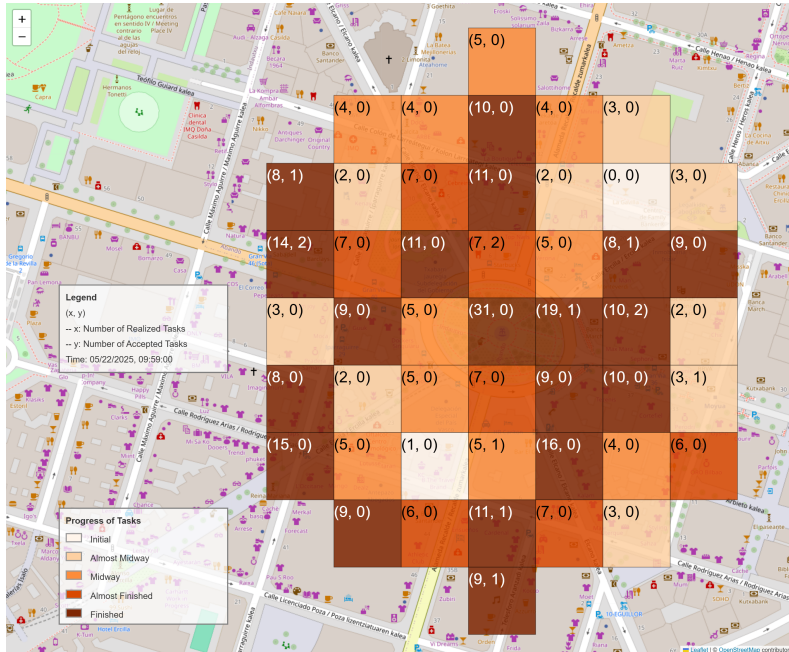


Figure E.15: NN - 50 Users

E.1.4 25 Users

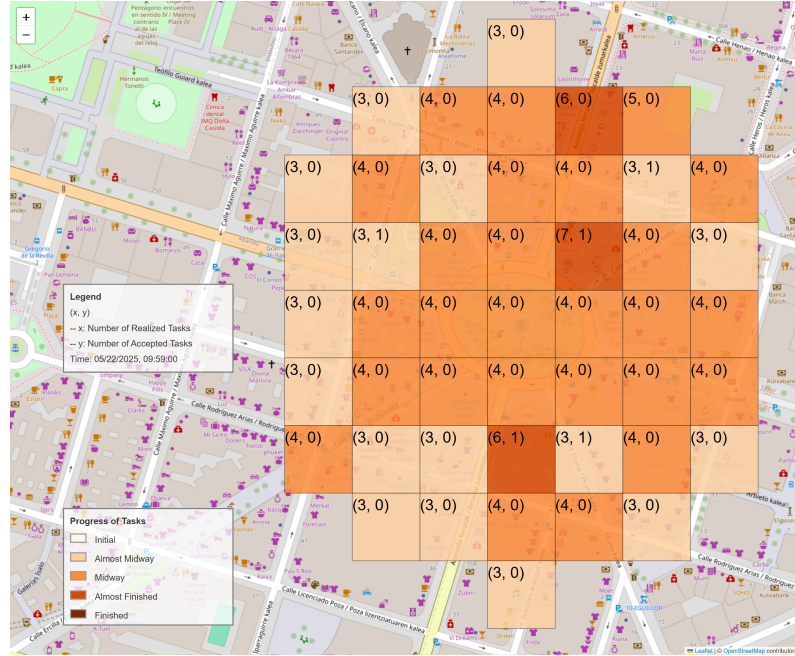


Figure E.16: Hybrid-VTAS - 25 Users

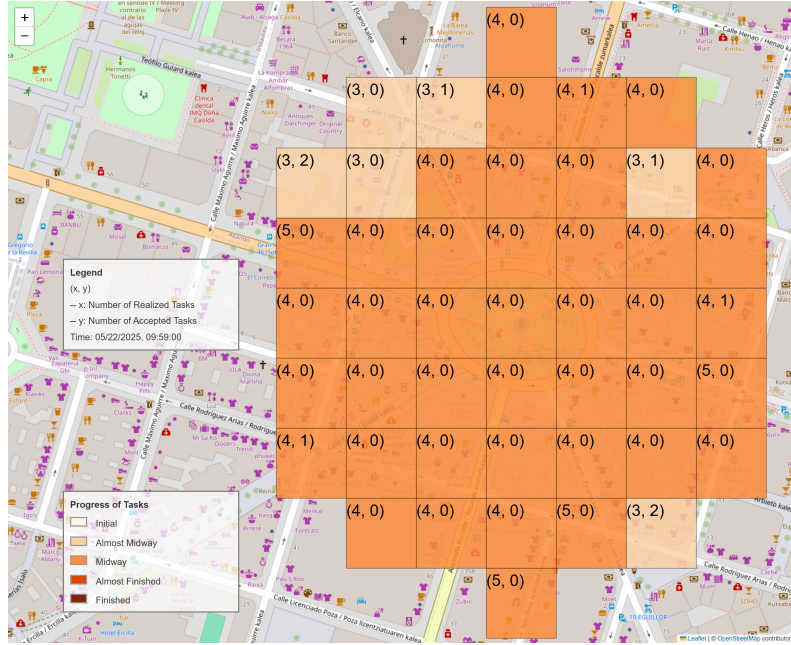


Figure E.17: VTAS - 25 Users

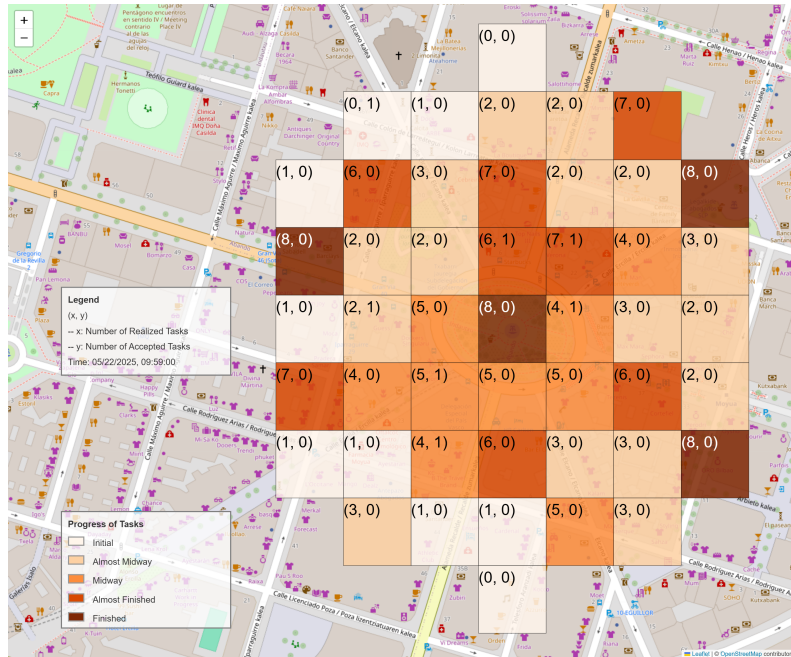


Figure E.18: Bio-VTAS - 25 Users

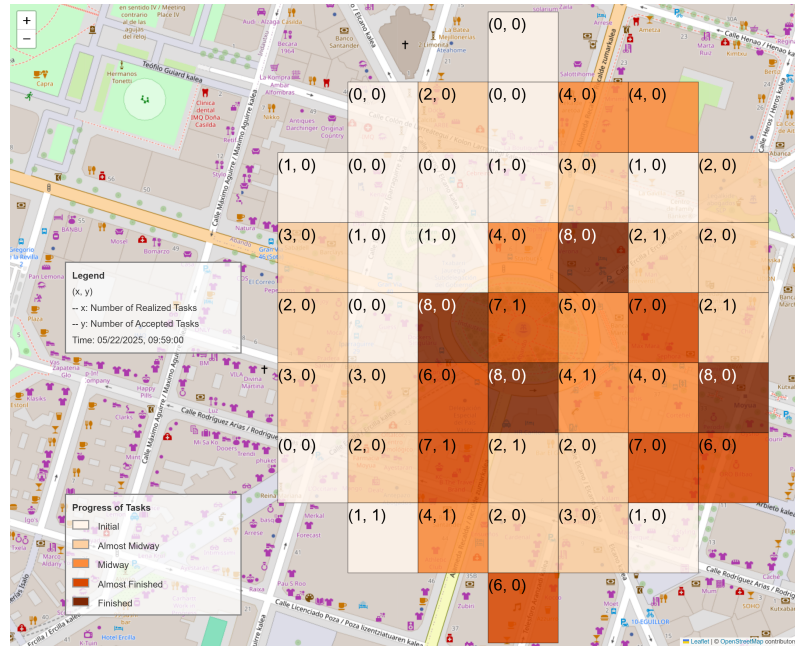


Figure E.19: LNN - 25 Users

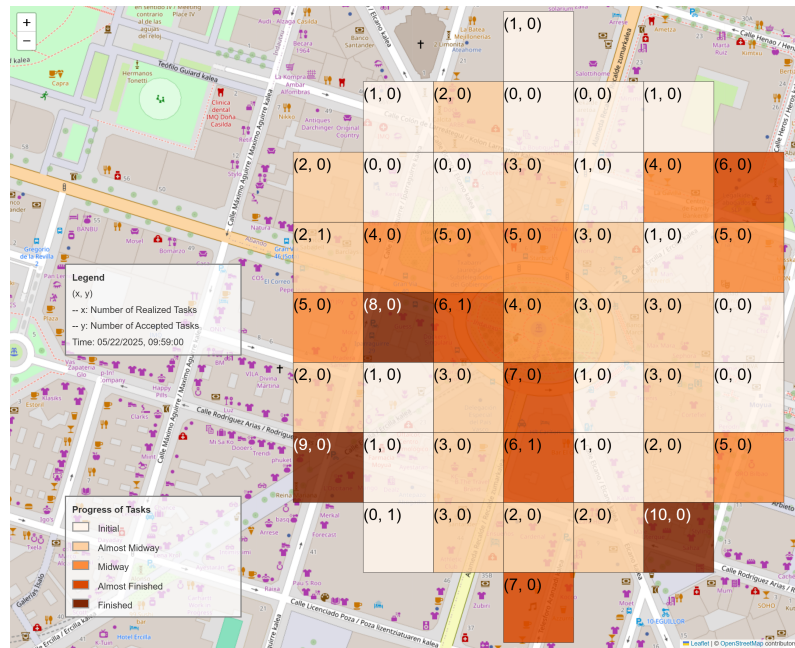


Figure E.20: NN - 25 Users

E.2 Experiment 2 - User Availability Probability

This section provides the visual representation of the LFI metric for the second experiment, with 75%, 50%, 25% User Availability Probability and for all algorithms used throughout this thesis. In these simulations, the User Task Completion Probability was set at 75%, reflecting a high level of user commitment to their assigned tasks, while the number of users was fixed at 50.

E.2.1 75% User Availability Probability

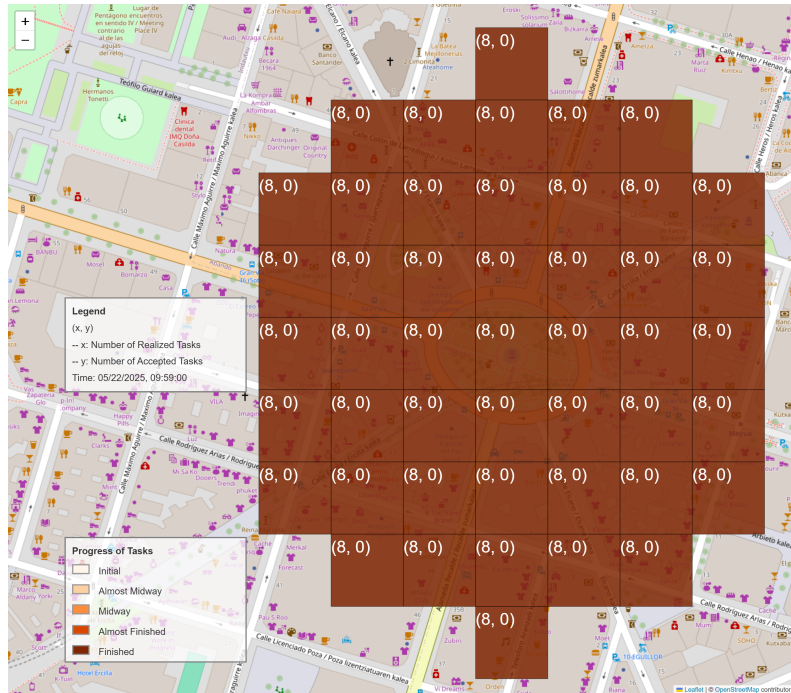


Figure E.21: Hybrid-VTAS - 75% User Availability Probability

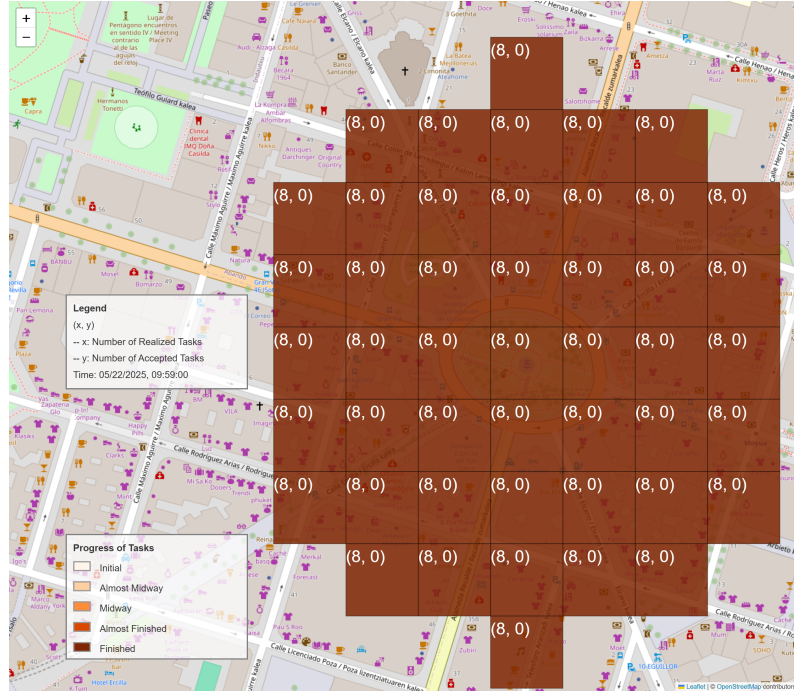


Figure E.22: VTAS - 75% User Availability Probability

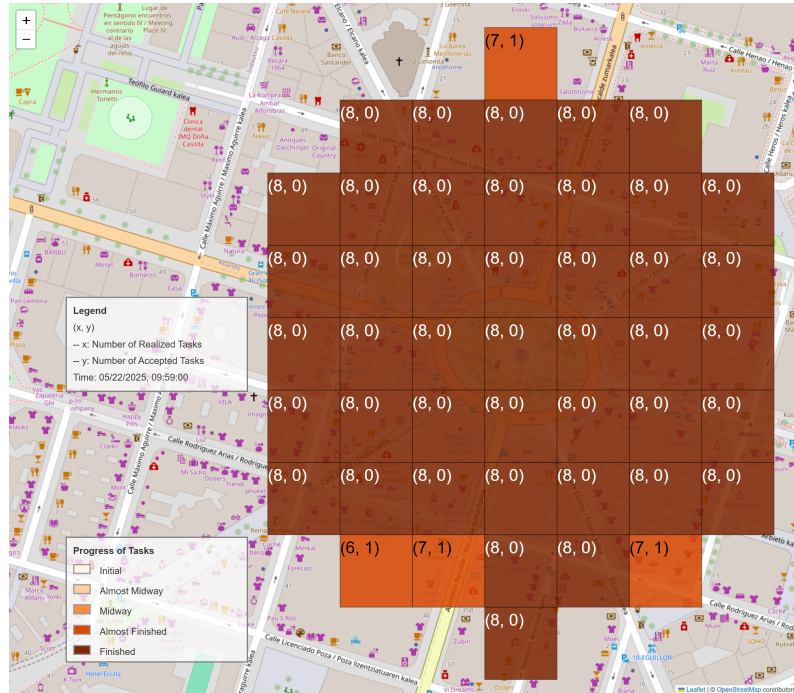


Figure E.23: Bio-VTAS - 75% User Availability Probability

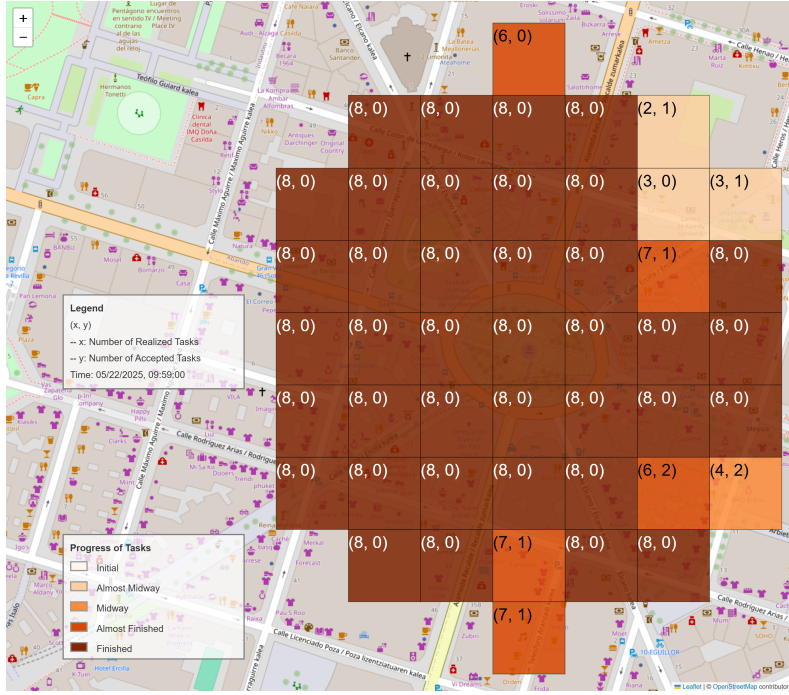


Figure E.24: LNN - 75% User Availability Probability

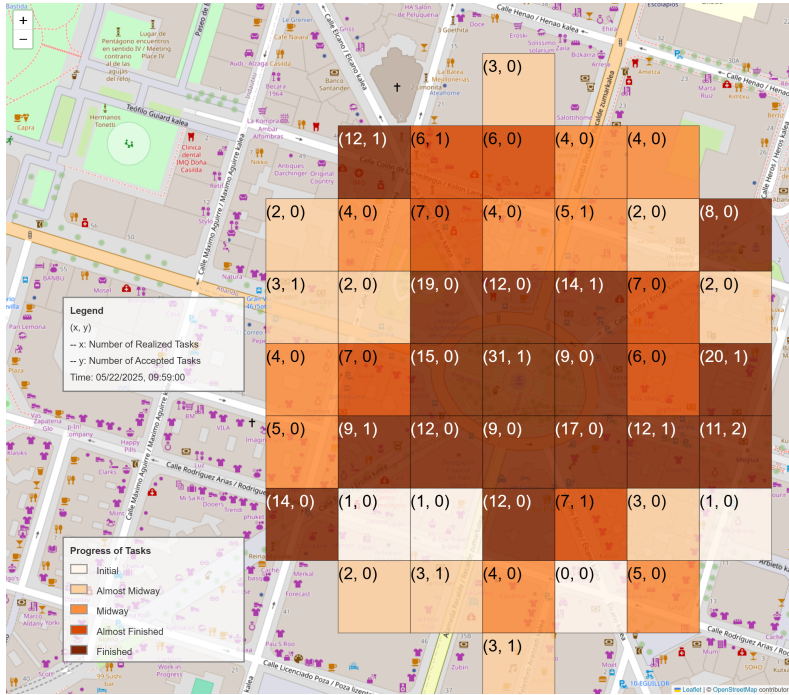


Figure E.25: NN - 75% User Availability Probability

E.2.2 50% User Availability Probability

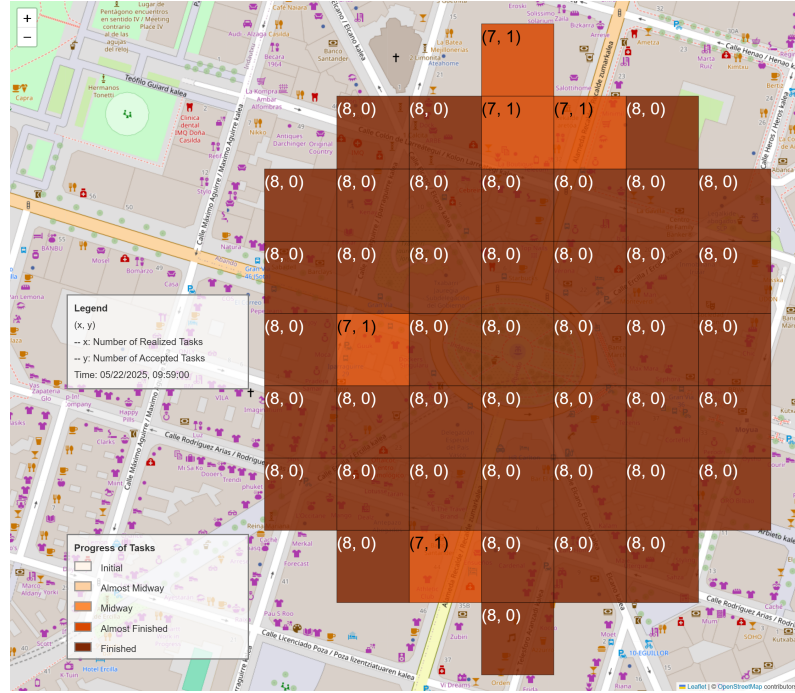


Figure E.26: Hybrid-VTAS - 50% User Availability Probability

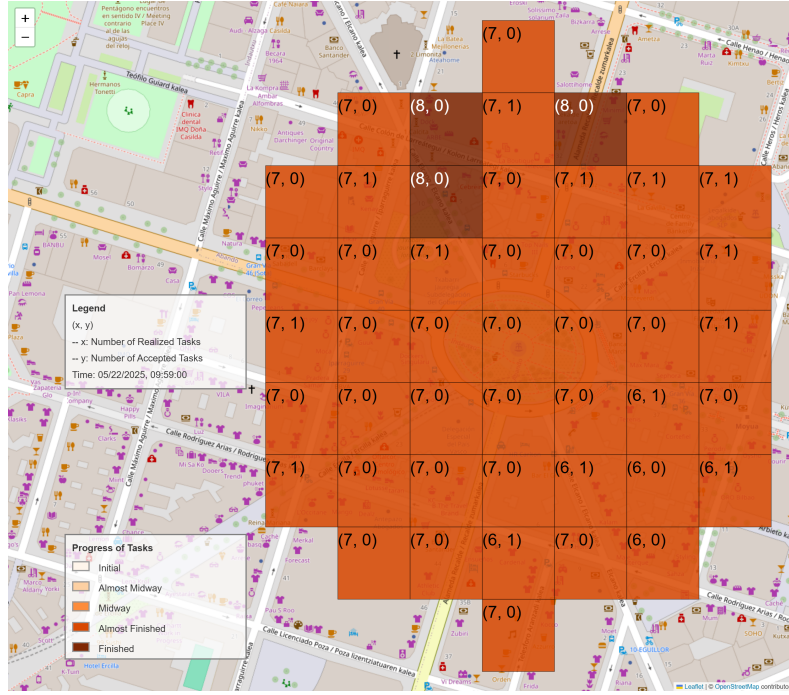


Figure E.27: VTAS - 50% User Availability Probability

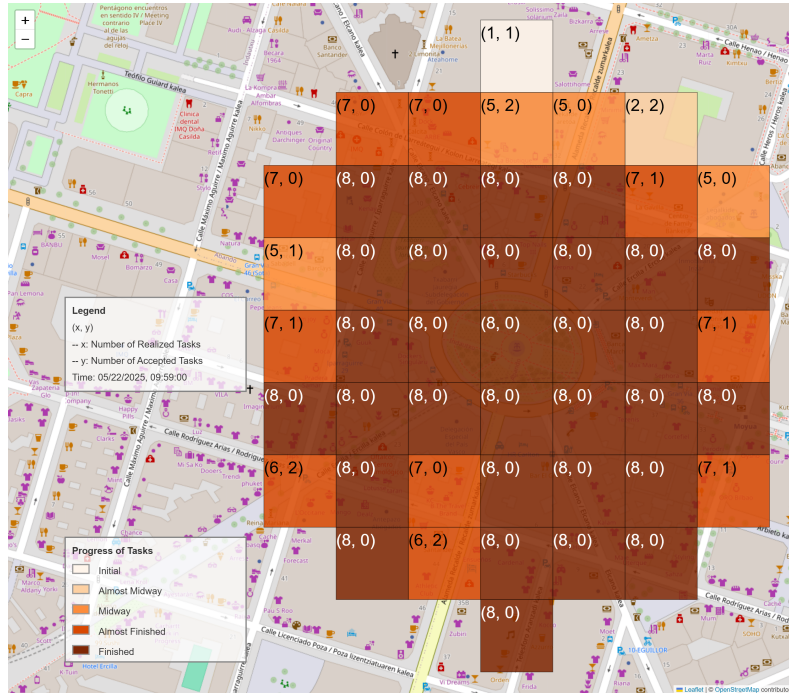


Figure E.28: Bio-VTAS - 50% User Availability Probability

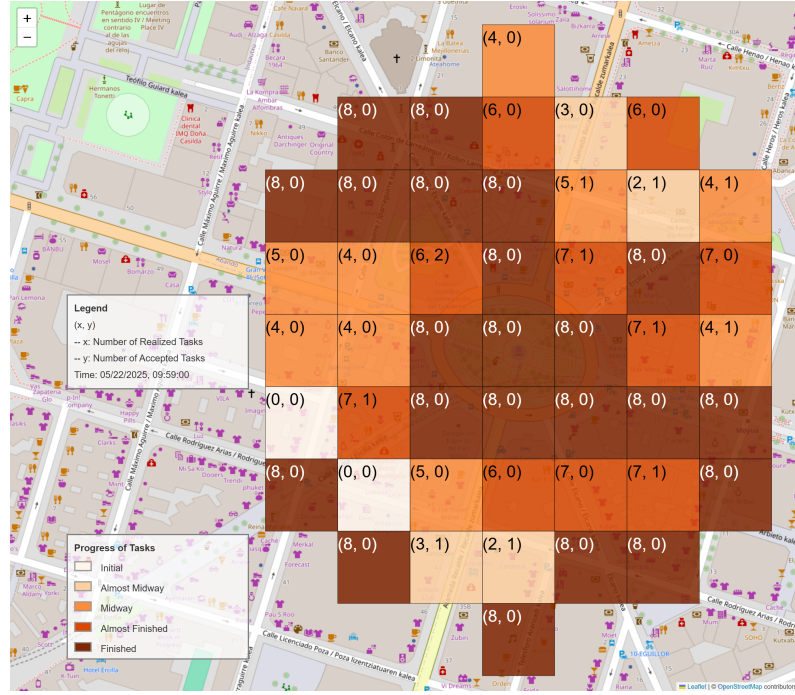


Figure E.29: LNN - 50% User Availability Probability

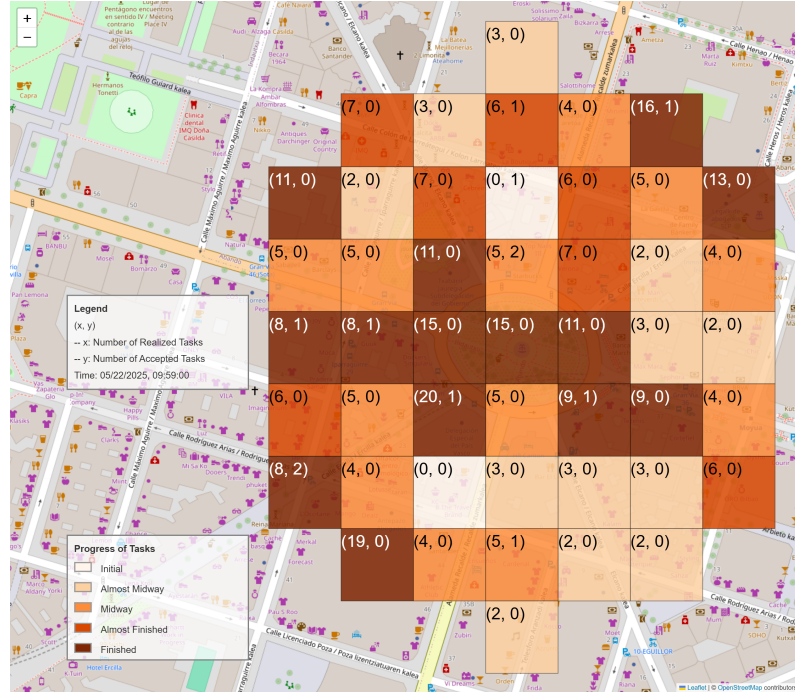


Figure E.30: NN - 50% User Availability Probability

E.2.3 25% User Availability Probability

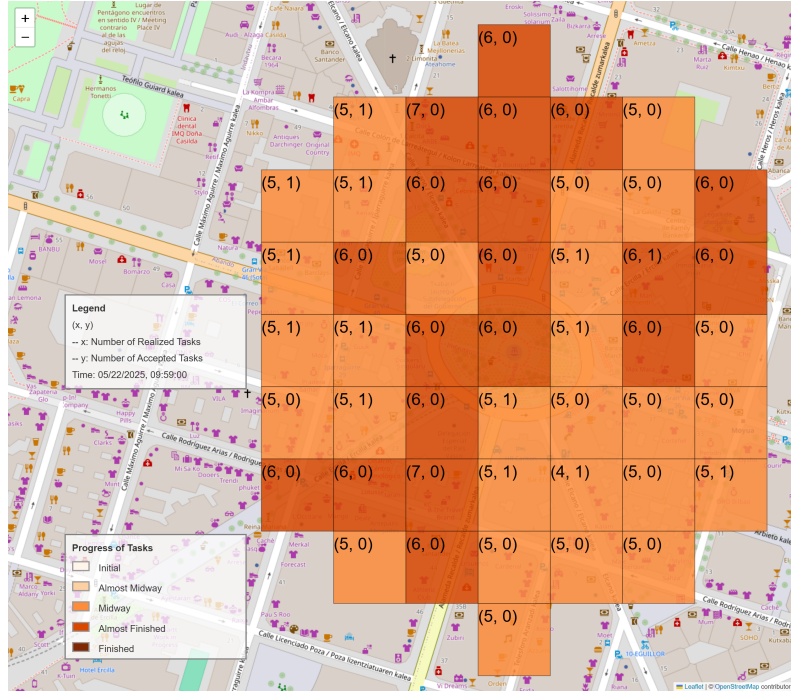


Figure E.31: Hybrid-VTAS - 25% User Availability Probability

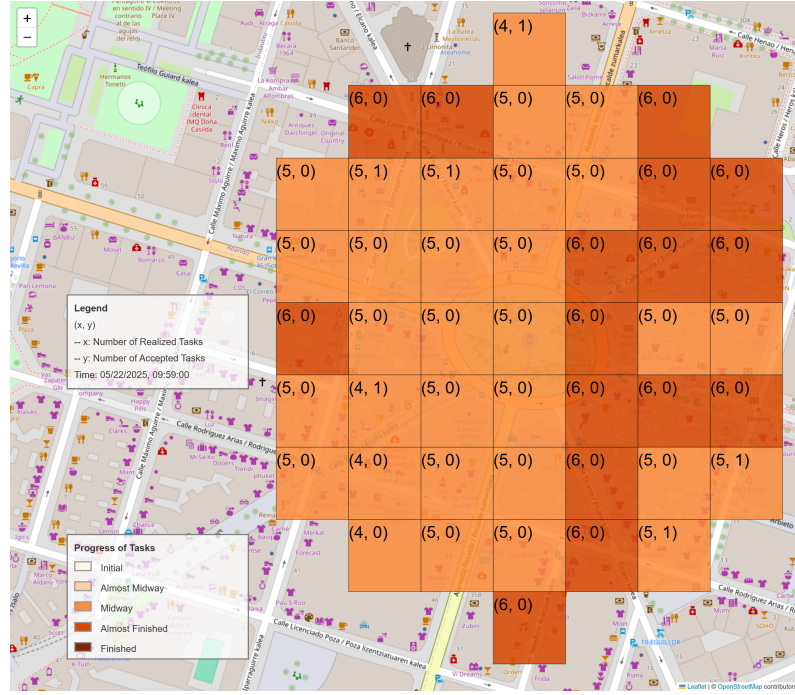


Figure E.32: VTAS - 25% User Availability Probability

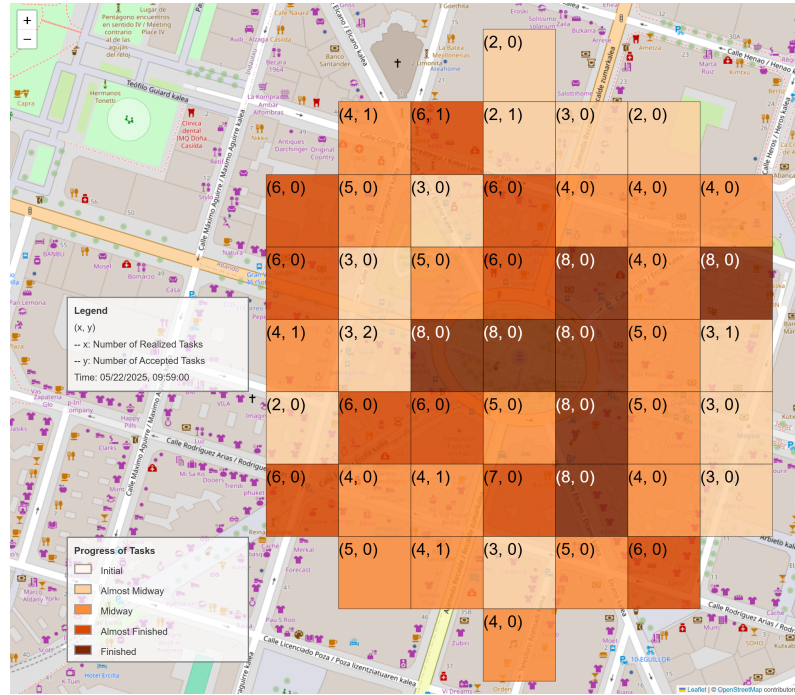


Figure E.33: Bio-VTAS - 25% User Availability Probability

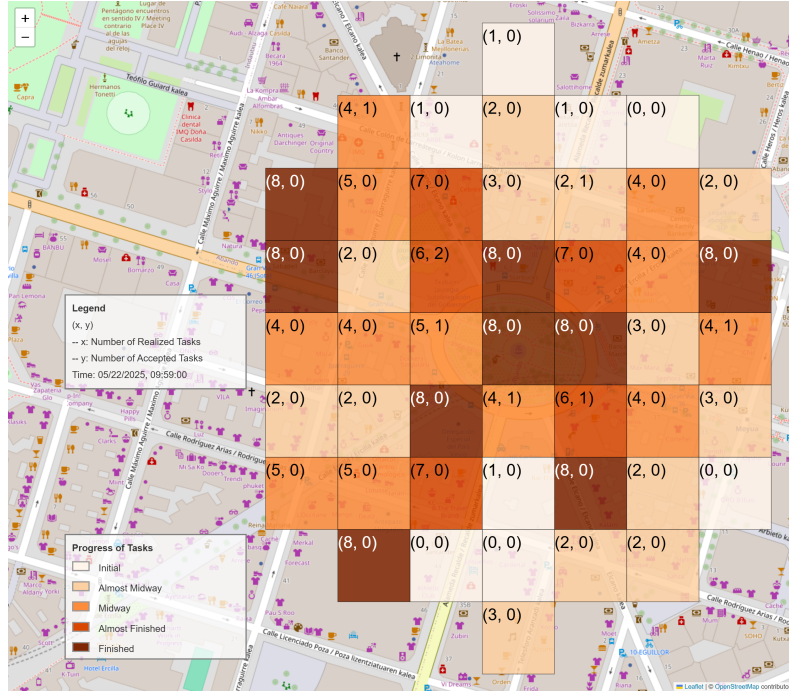


Figure E.34: LNN - 25% User Availability Probability

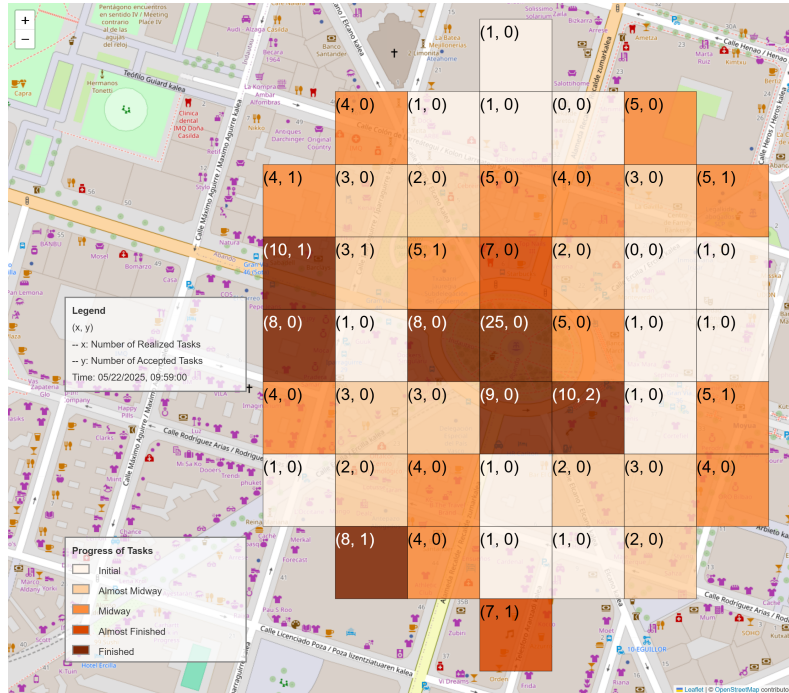


Figure E.35: NN - 25% User Availability Probability

E.3 Experiment 3 - User Tasks Completion Probability

This section provides the visual representation of the LFI metric for the last experiment, with 75%, 50%, 25% User Tasks Completion Probability and for all algorithms used throughout this thesis. In these simulations, the User Availability Probability was set at 50%, while the number of users was fixed at 50.

E.3.1 75% User Task Completion Probability

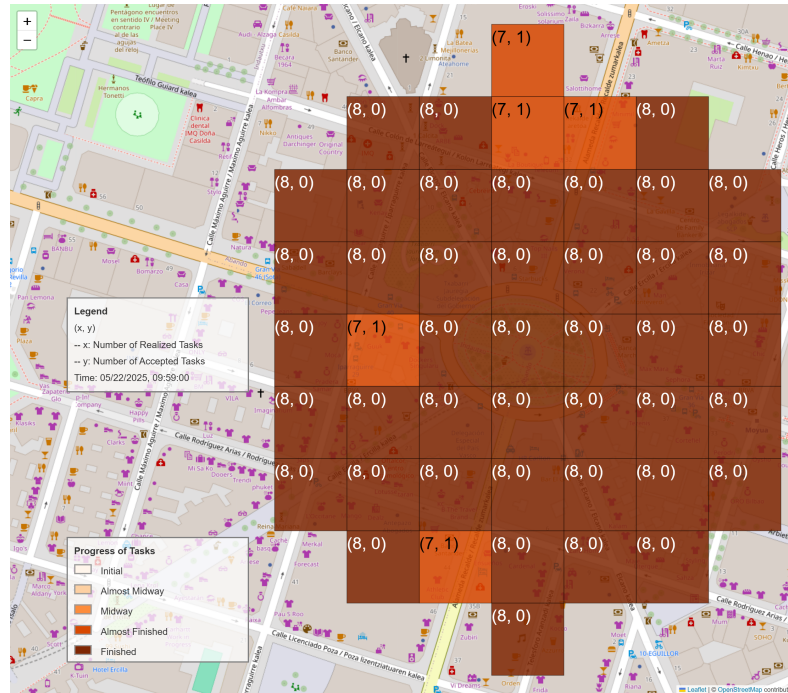


Figure E.36: Hybrid-VTAS - 75% User Task Completion Probability

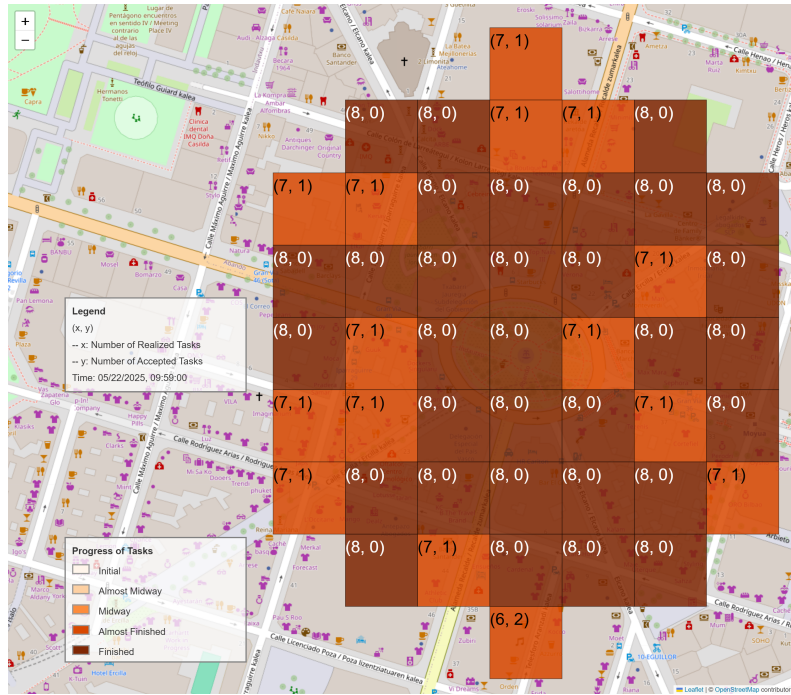


Figure E.37: VTAS - 75% User Task Completion Probability

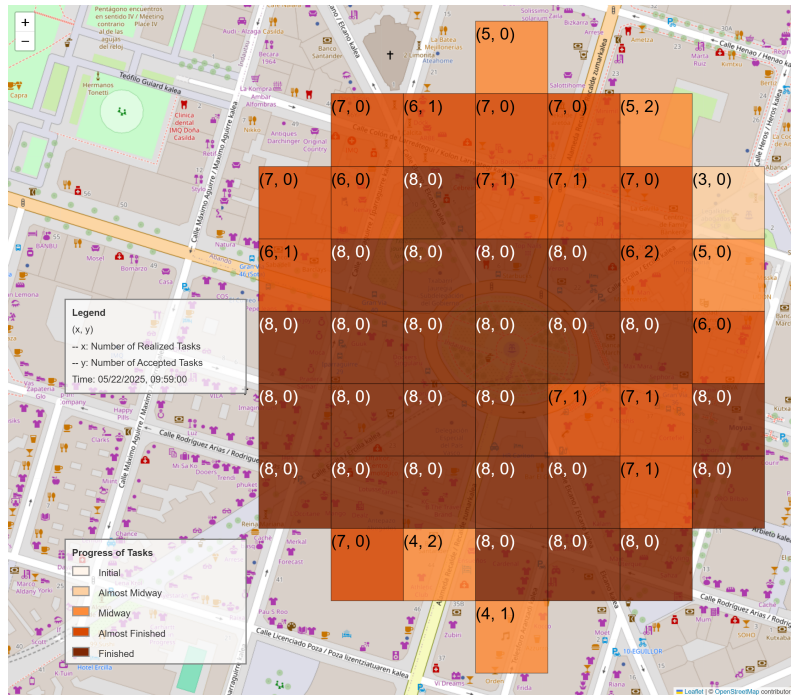


Figure E.38: Bio-VTAS - 75% User Task Completion Probability

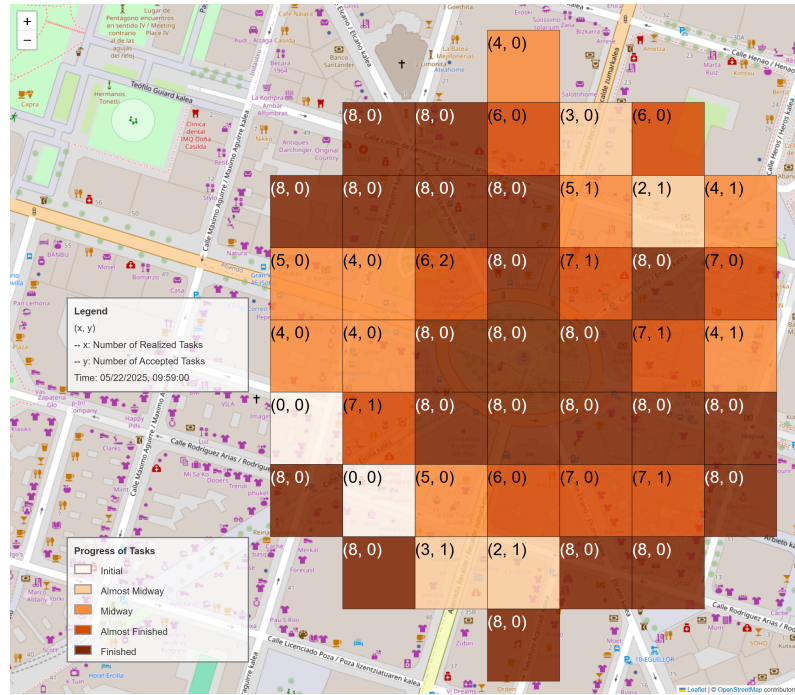


Figure E.39: LNN - 75% User Task Completion Probability

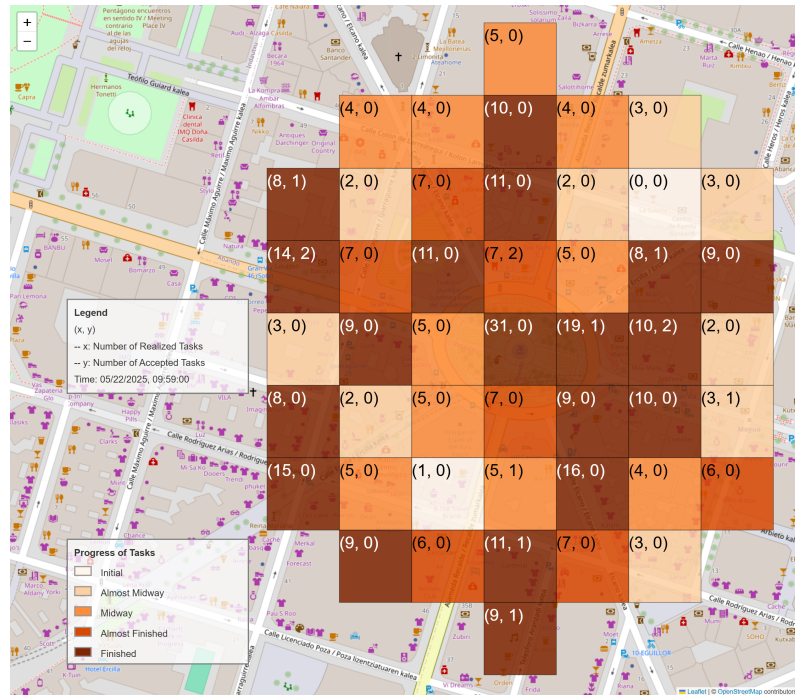


Figure E.40: NN - 75% User Task Completion Probability

E.3.2 50% User Task Completion Probability

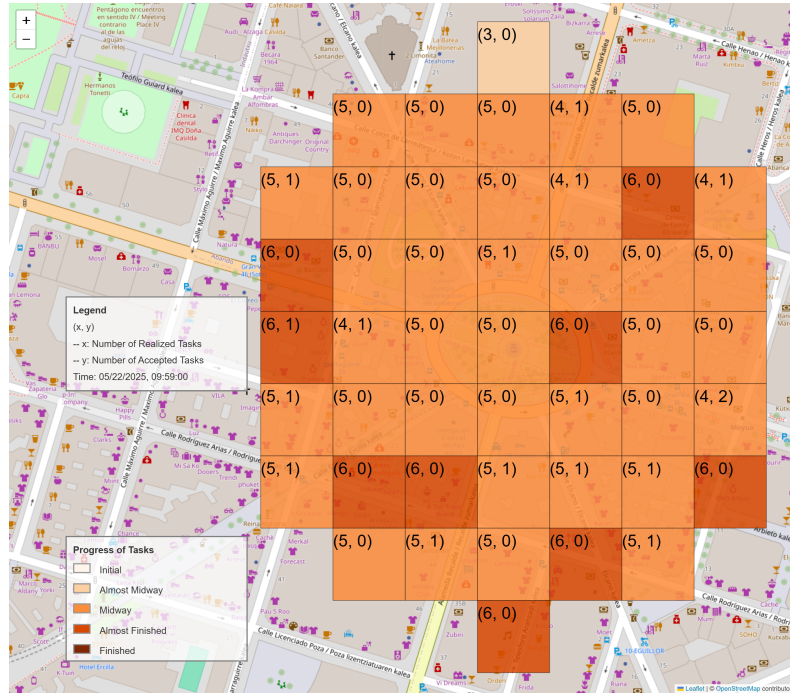


Figure E.41: Hybrid-VTAS - 50% User Task Completion Probability

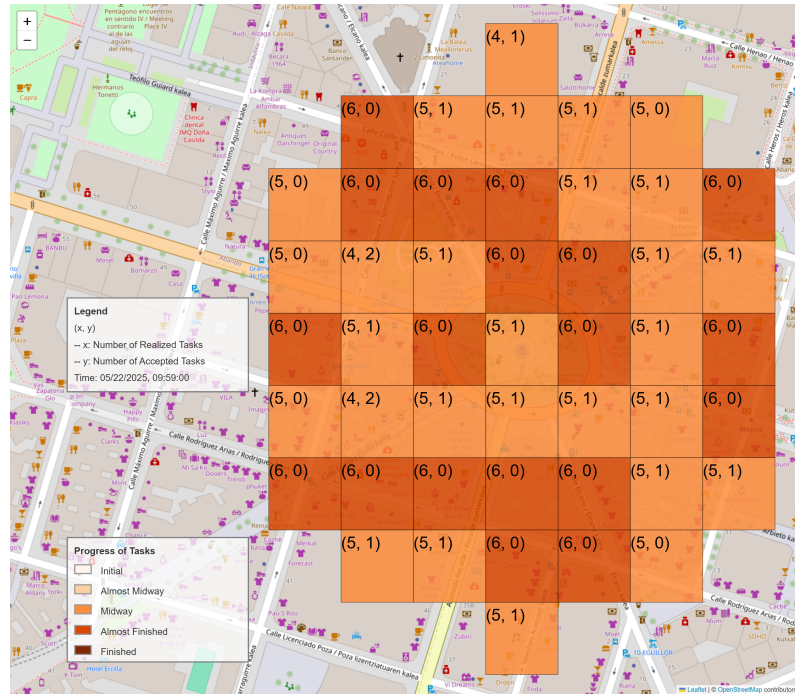


Figure E.42: VTAS - 50% User Task Completion Probability

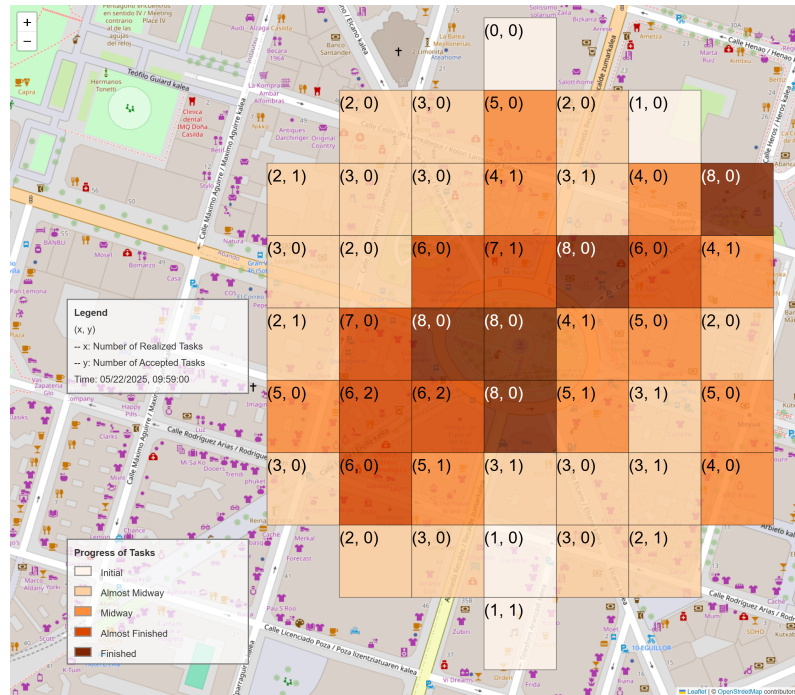


Figure E.43: Bio-VTAS - 50% User Task Completion Probability

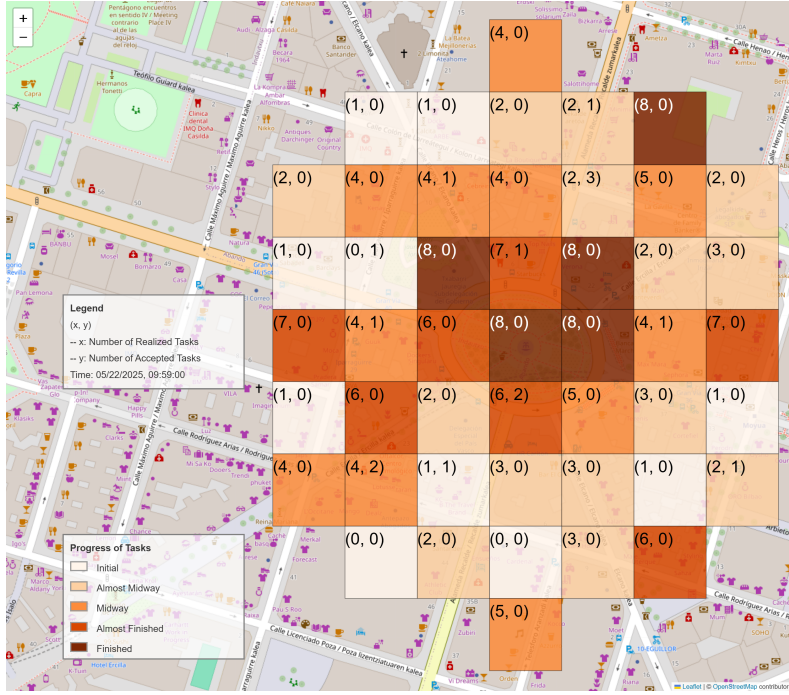


Figure E.44: LNN - 50% User Task Completion Probability

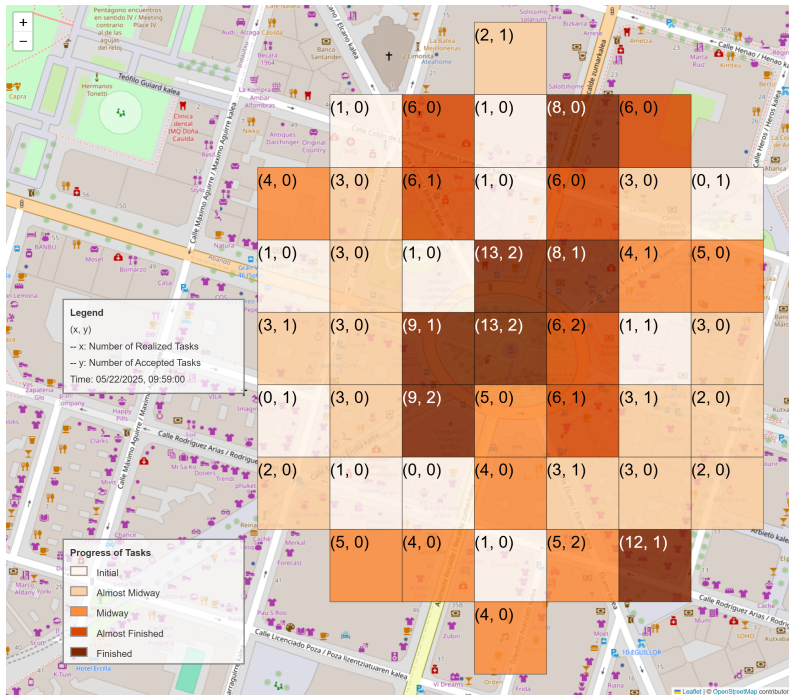


Figure E.45: NN - 50% User Task Completion Probability

E.3.3 25% User Task Completion Probability

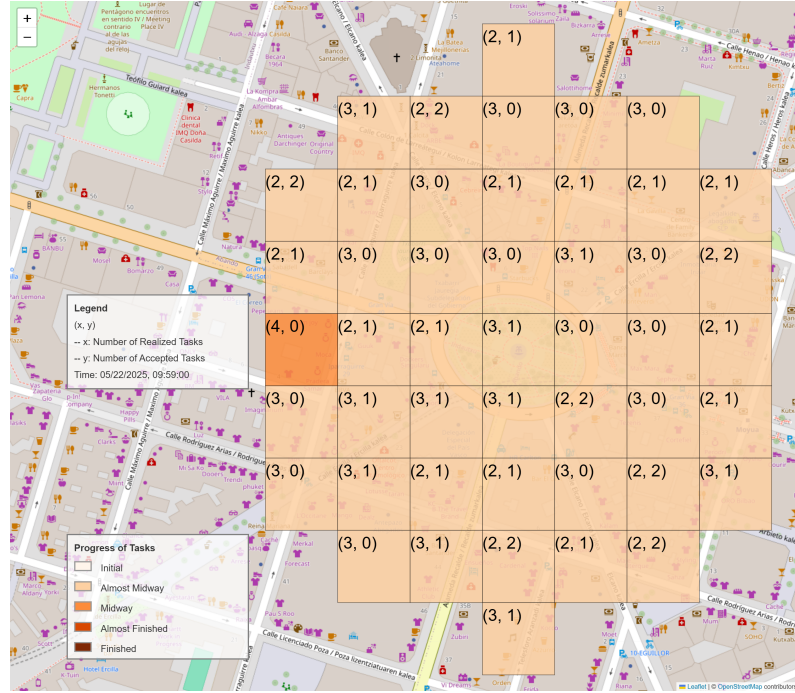


Figure E.46: Hybrid-VTAS - 25% User Task Completion Probability

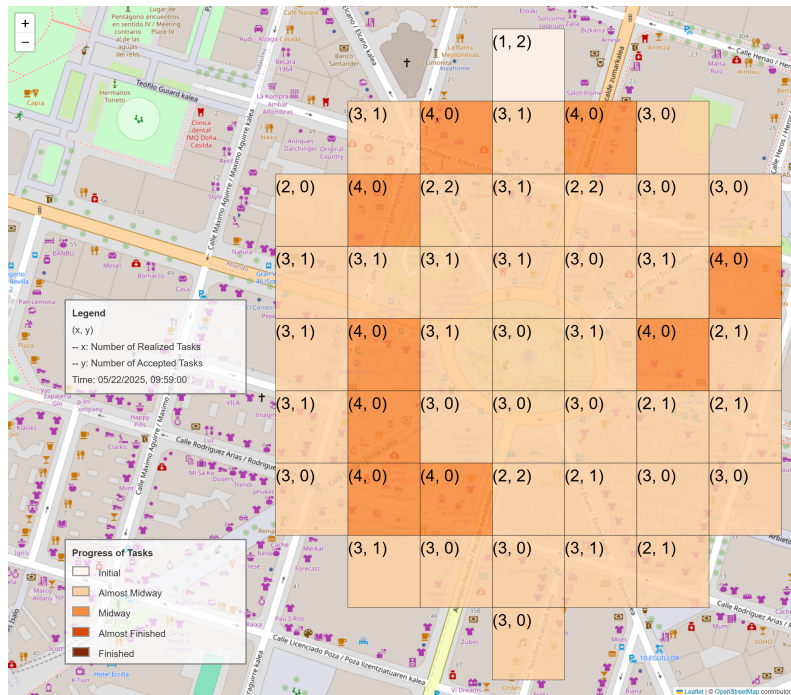


Figure E.47: VTAS - 25% User Task Completion Probability

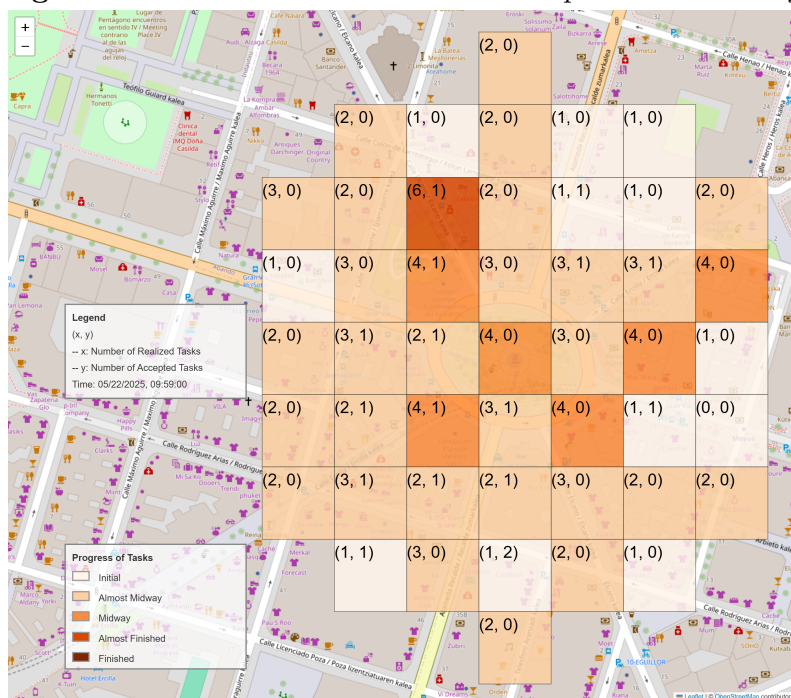


Figure E.48: Bio-VTAS - 25% User Task Completion Probability

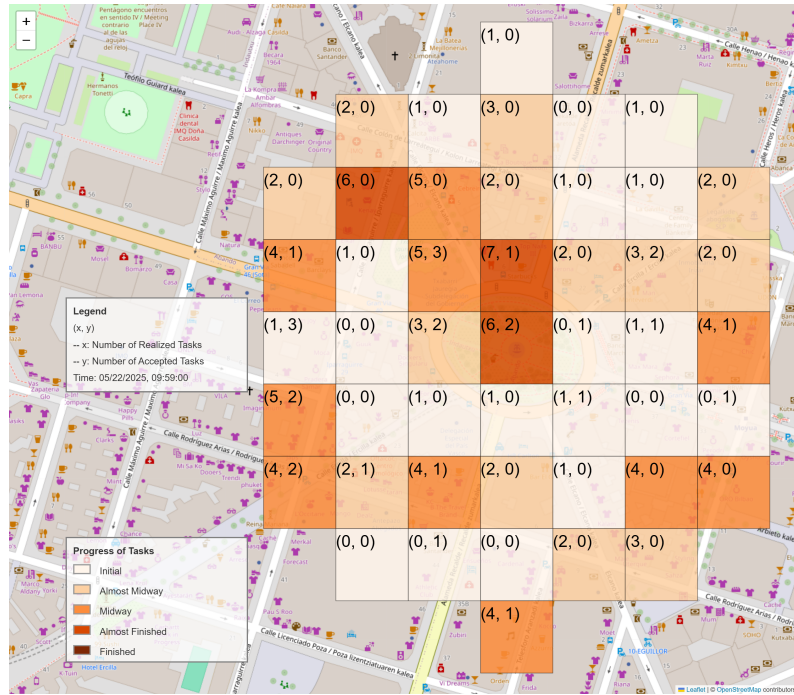


Figure E.49: LNN - 25% User Task Completion Probability

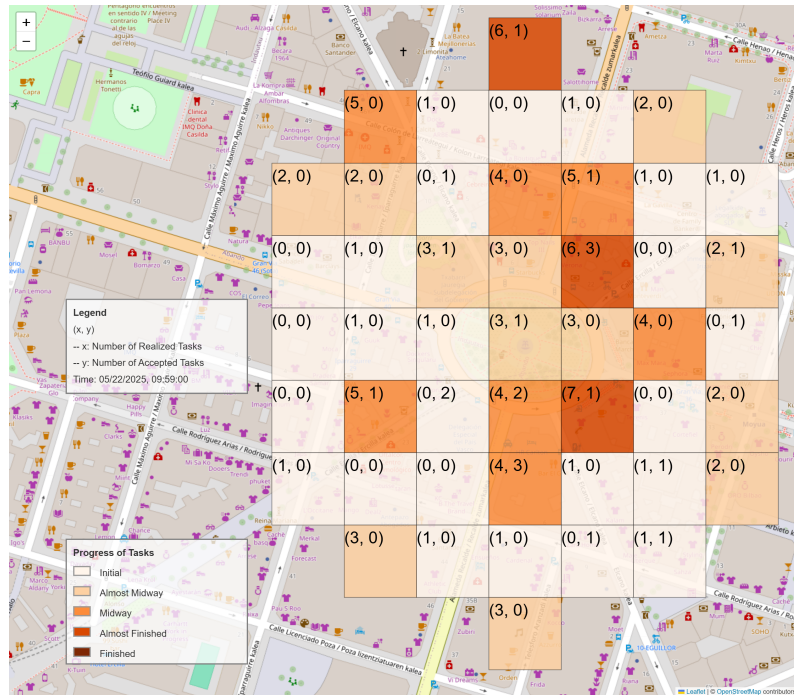


Figure E.50: NN - 25% User Task Completion Probability

Declaration

I, Maite Puerta Beldarrain, herewith declare that this dissertation is my own original work, carried out as a doctoral student at the University of Deusto. All assistance received and notions from other sources have been identified as such, acknowledging their correspondent contributions and citing them properly.

This work contains no material which has been presented in identical or similar form to any examination board, except where due acknowledgement is made in the dissertation.

This dissertation was finished writing on March 17th, 2025.