

KT4D

Knowledge Technologies
for Democracy

Project Number: 101094302

Start Date of Project: 01/02/2023

Duration: 36 months

Deliverable 1.2

Consolidated Report on the Results of the Use Cases:

Coordination of the Combined Use Case Methodology

Dissemination Level	PU
Due Date of Deliverable	30/11/2023
Actual Submission Date	30/11/2023
Work Package	WP 1 Management and Coordination
Task	T1.5
Type	Report
Version	V1.0
Number of Pages	p.1 – p.74



Funded by the
European Union

KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Deliverable Abstract

This report will document the Use Cases, including the specific materials and variations presented at each of the Use Case meetings, number and target groups of participants, anonymised feedback, etc. The first version of this Deliverable is due at Month 10 with further updates planned at Month 20 and Month 32.

The information in this document reflects only the author's views and the European Community is not liable for any use that may be made of the information contained therein. The information in this document is provided "as is" without guarantee or warranty of any kind, express or implied, including but not limited to the fitness of the information for a particular purpose. The user thereof uses the information at his/her sole risk and liability. This deliverable is licensed under a Creative Commons Attribution 4.0 International License.



DELIVERY SLIP

Date	Name	Partner/Activity	Date
From:			
Moderated by:	Eleonora Lima	TCD	27/11/2023
Reviewed by:	Eva Power	TCD	28/11/2023
Approved by:	Jennifer Edmond	TCD	29/11/2023

DOCUMENT LOG

Issue	Date	Comment	Author	ORCID ID
v.1	14/08/2023	Initial draft and layout	Eva Power	0009-0000-5118-0417
v.2	15/08/2023	Adding content and editing	Jennifer Edmond	0000-0001-9991-1637
v.3	21/08/2023	Adding content and editing	Eleonora Lima	0000-0001-7578-8005
v.4	25/08/2023	Adding content and editing	Eva Power	0009-0000-5118-0417
v.5	25/08/2023	Adding content and editing	Eleonora Lima	0000-0001-7578-8005
v.6	19/09/2023	Adding content	Ziemowit Stanczyk	
v.7	04/10/2023	Editing	Eva Power	0009-0000-5118-0417
v.8	05/10/2023	Adding content	Lucia Garcia	0009-0006-4257-7993
v.9	09/10/2023	Adding content	Andres del Alamo	0000-0003-4438-5312
v.10	27/10/2023	Adding content	Anna Björk	0000-0002-2008-0920
v.11	01/11/2023	Adding content	Johannes Anttila	
v.12	01/11/2023	Adding content	Atte Ojanen	0000-0002-7543-5821
v.13	01/11/2023	Adding content	Anna Björk	0000-0002-2008-0920

v.14	12/11/2023	Adding content	Jennifer Edmond	0000-0001-9991-1637
v.15	03/11/2023	Editing	Eva Power	0009-0000-5118-0417
v.16	03/11/2023	Adding content and editing	Eleonora Lima	0000-0001-7578-8005
v.17	14/11/2023	Adding content and editing	Olha Zadorozhna	0000-0002-2160-6904
v.18	14/11/2023	Adding content and editing	Piotr Lis	0000-0001-6060-2423
v.19	20/11/2023	Adding content and editing	Lucia Garcia	0009-0006-4257-7993
v.20	21/11/2023	Editing	Eva Power	0009-0000-5118-0417
V.21	23/11/2023	Peer review	Joanne Ahern	0000-0001-5793-7687
v.22	26/11/2023	Peer review	Furkan Akar	0000-0002-6723-6996
v.23	27/11/2023	Adding content and editing	Eleonora Lima	0000-0001-7578-8005
v.24	27/11/2023	Final edits	Jennifer Edmond	0000-0001-9991-1637
v.24	28/11/2023	Formatting	Eva Power	0009-0000-5118-0417

TERMINOLOGY

Terminology/Acronym	Definition
AI	Artificial Intelligence
AKT	Advanced Knowledge Technology
GPT	Generative Pre-Trained Transformer
CIB	Fundación Cibervoluntarios
CSO	Civil Society Organisation
Demos	Demos Helsinki
EU	European Union
GDPR	General Data Protection Regulation
HLEG	High Level Expert Group
IRMiR	The Institute of Urban and Regional Development, Warsaw
KER	Key Exploitative Result
KT4D	Knowledge Technologies For Democracy
LGTB	Lesbian, Gay, Trans, Bisexual
LLM	Large Language Models
NGO	Non-Governmental Organisation
TCD	Trinity College Dublin
US	United States
WP	Work Package

Table of Contents

1 Introduction	9
2 KT4D’s Use Cases.....	9
2.1 Purpose of the KT4D Use Cases.....	9
2.2 General Goals and Structure of KT4D’s Use Cases	10
2.3 Factors considered in the planning and recruitment phase.....	10
2.3.1 Representativity.....	10
2.3.2 Logistics.....	10
2.3.3 External Factors	11
2.3.4 Need for Flexibility.....	11
2.4 Temporal Structure of the Meetings of Each Use Case.....	12
2.5 Differences Between the Four Use Cases.....	13
2.6 Cross Project Embedding of the the KT4D Use Cases.....	13
3 Key Concepts and Shared Glossary	15
4 Identifying Stakeholders and Participant Recruitment.....	16
4.1 Plain Language Information Sheets and Consent Forms.....	16
4.2 Pseudonymisation	17
5 Descriptions of the Use Cases.....	17
5.1 Use Case 1 (Brussels): Governance Framework, Policy Roadmap and Recommendations.....	17
5.1.1 General Goal and Structure	17
5.1.2 Outline of Use Case 1 Meeting 1.....	18
5.1.3 Methodology and Activities of Use Case 1 Meeting 1	19
5.1.4 Results of Use Case 1 Meeting 1:.....	22
5.1.5 Plans for future interactions	23
5.2 Use Cases 2 & 3 (Madrid and Warsaw): Narrative Based Simulation Game and Two Interactive Explainers	23
5.2.1 General Goal and Structure	23
5.2.1.1. Session’s objectives	23
5.2.1.2. Outline of the Use Cases 2 and 3.....	23
5.2.3 Methodology and Activities	24
5.2.3.1 Prior to the Use Case meeting.....	24
5.2.3.2 Welcome	25
5.2.3.3 Warm up.....	25
5.2.3.4 Discussions - in groups.....	26
5.2.3.5 Co-creation	28
5.2.4 Results of Use Case 2 Meeting 1: Madrid	29
5.2.4.1 Participants and their demographics:	29
5.2.4.2 Focus groups results.....	30
5.2.4.2.1 General discussion about the AI and what it is.....	30
5.2.4.2.2 In group discussions about different areas AI will have an impact in 10 years	30

5.2.4.3 Materials validation exercise results	32
5.2.4.4 Co-creation exercise results	36
5.2.5 Plans for future interactions	39
5.2.6 Results of Use Case 3 Meeting 1: Warsaw	39
5.2.6.1 Participants and their demographics	39
5.2.6.2 Attitudes towards AI at the start of the session.....	39
5.2.6.3 General discussion about the AI and what it is	39
5.2.6.4 Choosing the AI technologies/uses that will have the greatest impact on democracy and civil rights in the future	40
5.2.6.5 In group discussions about different areas AI will have an impact in 10 years.....	40
5.2.6.6 Materials about AI exercise results	45
5.2.6.7 Co-creation exercise results	46
5.2.7 Plans for future interactions	49
5.3 Use Case 4 (Dublin): Social Computing Compass for self-assessment	50
5.3.1 Goal.....	50
5.3.2 Outline of Use Case 4 Meeting 1.....	51
5.3.3 Methodology and Activities of Use Case 4 Meeting 1	53
5.3.3.1 Activities	53
5.3.3.2 Methodology	57
5.3.4 Results Use Case 4 Meeting 1	58
5.3.4.1 Activity 1: The Conversation stations.....	58
5.3.4.2 Activity 2: Co-creation activity.....	61
5.3.5 Plans for future interactions	63
6 Conclusion.....	63
Appendix I. Consent Form Template	67
Appendix II. Information Sheet Template.....	70

List of Figures

Figure 1 KT4D’s Use Cases	9
Figure 2 Typology of Risks Caused by AI to Democracy.....	19
Figure 3 Current Policy and Regulatory Landscape.....	20
Figure 4 Gaps and Weak Spots in Current EU Approach	21
Figure 5 Fostering Democracy Through AI.....	21
Figure 6 Technologies/uses of AI that will have the greatest impact on democracy and civil rights	26
Figure 7 In-group discussions.....	27
Figure 8 First Co-creation Exercise.....	36
Figure 9 Virtual Insanity	37
Figure 10 KT4All	38

Figure 11 Learning with Fanti	39
Figure 12 Votes on AI Technology with the Greatest Impact on Democracy	40
Figure 13 In-group discussion results of the impact of AI on “Education” in 10 years.	41
Figure 14 In-group discussion results of the impact of AI on “Entertainment” in 10 years.....	42
Figure 15 In-group discussion results of the impact of AI on “Health” in 10 years	43
Figure 16 In-group discussion results of the impact of AI on “Work” in 10 years	44
Figure 17 In-group discussion results of the impact of AI on “Civil society, politics, and democracy” in 10 years.....	45
Figure 18 Jurassic Park	47
Figure 19 Submarine	48
Figure 20 Desert Island - First Version	48
Figure 21 Deserted Island - Second Version	49
Figure 22 Haunted House	49
Figure 23 Use Case 4 Meeting 1 Main Poster	54
Figure 24 Use Case 4 Meeting 1 Station 2 Main Poster	55
Figure 25 Use Case 4 Meeting 1 Station 3 Main Poster	56
Figure 26 Co-Creation Exercise CULTURaLi Tool	61
Figure 27 Co-creation Exercise ZombAI	62
Figure 28 KT4D KERs (highlight KERs 2, 3 & 4)	64
Figure 29 KT4D KERs (Highlight KERs 5 & 6).....	65

List of Tables

Table 1 KT4D Tasks Contributing to the Use Cases.....	14
Table 2 KT4D Glossary.....	15
Table 3 Group 1 Participant Feedback.....	33
Table 4 Groups 2 Participant Feedback	33
Table 5 Group 3 Participant Feedback.....	34
Table 6 Group 4 Participant Feedback.....	35

Executive Summary

This Deliverable describes the KT4D project's four Use Cases, from the point of their conceptual design through the results of the first meetings of each. The Use Cases are an intrinsic part of the project, conceptually aligned, but practically different in that they face different user communities, highlight different project results, and take place in different cities (and in some cases in different languages).

The first section of the deliverable covers the conceptual and practical framework that unites the four use cases, including the manner in which the tasks to achieve them have been placed under distributed responsibility throughout the project. It also looks at the generic framework for the three meetings of each cohort, and our approach to recruitment and handling of personal data of the participants (see [KT4D's Pseudonymisation Guidelines](#) and Deliverables 8.1 and 8.3). The report continues to give an account of the first meeting of each of the four Use Cases, which were modelled according to the principles of participatory design.

Each of these Use Cases reports successful and highly useful insights, albeit with a different focus in each case.

Use Case 1 focuses on navigating the complexities of AI governance and formulating a democratic approach to regulate general-purpose AI. The primary goal is to address challenges posed by emerging technologies, such as power centralization, opacity, and the rapid evolution of AI, which threaten democratic values. During their roundtable in Brussels, policymakers and policy-facing CSOs were engaged to explore AI's impact on democracy, identify gaps in EU governance, and propose democratic AI governance solutions. The session also examined risks to democracy, the second evaluated gaps in current EU regulations, and the third explored potential governance responses. Key risks highlighted include mass manipulation, concentration of power in US AI companies, and misuse of open-source models. Identified gaps in EU regulations involved enforcement challenges, compatibility issues with collective bargaining models, and unclear liability for AI providers. The findings from this meeting will inform a Delphi survey in 2024, extending outreach to key stakeholders. Results will contribute to refining the governance framework and roadmap for democratic AI regulation, aligning with the EU's commitment to global leadership in AI regulation.

Use Cases 2 & 3, which are being conducted in Madrid and Warsaw, focus on exploring citizens' perceptions of knowledge technologies and the development of educational materials and games to enhance digital literacy. The goals include understanding threats and opportunities related to AI, big data, and democracy, as well as fostering critical literacy for democratic challenges tied to technology. The first sessions in these Use Cases were very well-attended, focussing on group discussions about AI's future impact, and feedback collection on existing educational materials. A Co-creation activity allowed participants to develop scenarios for escape room games addressing AI-related challenges. Key points for effective AI education materials included a stepped approach, concrete examples, consultation with older adults for intuitiveness, tailoring information depth to the audience, and a multi-modal approach for engagement.

Use Case 4 invited a group of software developers and people in positions of overseeing software development to explore the limits of current approaches to ethical AI development and to envision new, more effective ones that could take into consideration the cultural dimension of these software and systems.

The first meeting was shaped by two activities: the first employed “conversation stations” to explore challenges, tools, and cultural considerations, highlighting issues such as innovation hindrance, the need for clear ethical tools, and the struggle to integrate cultural aspects. Activity 2 was a co-creation exercise in which participants created prototypes for AI ethics tools targeting cultural differences, and focusing on pre, during, and post-development ethical checks. The workshop gathered nuanced perspectives, providing a foundation for further exploration and development of practical, culturally aware AI ethics tools.

1 Introduction

The overall aim of KT4D’s Use Cases, and the participatory design sessions that will comprise them, is to understand current and desired future user practices so as to align design to their requirements and identify knowledge gaps in relation to big data and AI. The Use Cases are at the heart of the KT4D value proposition, and therefore, rather than risking that they might become disconnected from the heart of the project’s research and development, the tasks to optimally deliver these key exploitation points are distributed throughout the WPs, with leadership passing from partner to partner as appropriate from a knowledge creation point of view, even while the physical locations of the meetings remain stable over the course of the project. As such, the Use Cases will be one of the points in the project most dependent on inter- and transdisciplinary exchange. This Deliverable will be released in three iterations in order to capture and disseminate the underpinning structure and methodology for delivering a coherent and yet flexible set of interactions to bridge KT4D research and development with user practices. Each iteration will be released following one of the three cycles of interaction that will be instigated for each cohort, capturing not only the underlying design philosophy, planning measures, and structure of the events, but also their results.

2 KT4D’s Use Cases

2.1 Purpose of the KT4D Use Cases

The complexity and the many cultural, political, technological and institutional factors we expect to encounter in the baseline state of the art we will draw from makes for a challenging environment in which to advance the state of the art: according to the values of what fields? Starting where? Benefitting whom? In our assessment, progress in this space must be underpinned by a fundamental understanding of three decision-making positions: that of the individual citizen user (or object) of technology; that of the regulator seeking to optimise control of technology for the good of democratic society; and the software developer creating the pre-conditions for both use and control. Levers for change within these three broad actor groups have also been defined as (respectively, though not exclusively) education, regulation, and innovation. Culture change is difficult, so a strong user-centric approach, covering these key stakeholder communities, will be essential to the successful outcomes of KT4D. Each of these use cases is based on a user scenario that crosses over at least two of the project’s stakeholder subgroups, as seen in Figure 1.



Figure 1 KT4D’s Use Cases

2.2 General Goals and Structure of KT4D's Use Cases

Each Use Case will be hosted in a different city, and face different target groups, with the exception of Use Cases #2 and #3, which will both engage citizens and citizen-facing CSOs. The reason for this is to account for the fact that democracy and civic participation are in essence local phenomena, requiring that participants be able to participate in a concrete rather than abstract context, their local language, and within cultural norms and expectations. This is particularly the case for citizen participants, whose participation will be far more contingent on their personal and individual identities, rather than their professional roles. The two parallel citizen-facing use cases in Madrid and Warsaw have therefore been chosen to give us a sense of the extent to which attitudes and effective measures toward the safeguarding of democracy in the face of AI and big data are reliant upon culturally specific linguistic and contextual signals.

Each use case meeting will seek to recruit a cohort of participants to engage actively with the conceptual framework and results of the KT4D project. More information on the structure and recruitment considerations for each meeting follow in section 2.3 below.

2.3 Factors considered in the planning and recruitment phase

In the process of designing the meetings for each of the four Use Cases, several key factors were taken into account to ensure optimal outcomes, as detailed further in Section 4.

2.3.1 Representativity

A significant challenge encountered during the design of the four Use Cases revolved around stakeholder mapping and the identification of a representative sample within each Use Case. To establish a comprehensive cross-section of the three categories involved in the Use Cases, deliberate efforts were made to invite participants from diverse profiles.

For instance, in Use Cases 2 and 3, citizens participating encompassed varying levels of digital literacy, diverse social backgrounds, different age groups, and disparate purposes for using AI technologies (such as work, leisure, and education). In the context of Use Case 4, software developers were invited from both academia and the industry sector, holding distinct roles and positions of authority.

2.3.2 Logistics

Logistical considerations encompassed a range of aspects aimed at fostering active engagement from participants. The decision was made to involve a group of 10-12 individuals for each Use Case meeting. This approach facilitated meaningful discussions among participants and ensured ample space for the expression of ideas and opinions, an endeavour that might be challenging in larger gatherings. This aligns with the qualitative approach of the KT4D project, which emphasises personal insights derived from situated experiences as the central observational standpoint for analysis. Should the need arise to engage more than 10-12 people for each phase of the Use Cases, the approach of maintaining multiple cohorts with smaller numbers was adopted.

While the ideal scenario involves the same group of individuals participating in all three planned meetings for each Use Case, the reality acknowledges that scheduling conflicts might prevent consistent attendance across several months. As a solution, the plan involves substituting participants with similar profiles, thereby

maintaining a consistent sample of individuals throughout each phase.

Regarding duration, it was determined that each meeting should approximately span half a day, either in the morning leading up to lunch or in the afternoon preceding dinner. This consideration respects participants' time, given their voluntary involvement, and prevents overwhelming participants, ensuring sustained engagement throughout the meeting. Whenever possible, arrangements were made to provide lunch or dinner at the conclusion of the meeting, fostering an environment for conviviality and informal exchange of ideas, factors vital for ensuring participants feel welcomed and engaged.

In contrast to the initial Grant Agreement, the inaugural meetings for the four Use Cases are scheduled for October 2023, rather than September 2023. This decision takes into account the challenge of recruiting participants during the summer for a September meeting, due to people's unavailability during the holiday season, which could result in a reduced participation.

2.3.3 External Factors

Each Use Case incorporates a consideration of external factors that could shape discussions during specific meetings. For instance, the first meeting of Use Case 3 is set in Warsaw in the weeks leading up to the 2023 Polish parliamentary election. Similarly, the first meeting of Use Case 2 is located in Madrid, following a recent general election (August 2023). These diverse political contexts are likely to influence people's AI-related concerns, such as the use of Russian bots in influencing elections in Poland, thus impacting the meeting's content.

Likewise, the releases of ChatGPT (November 2022) and GPT-4 (March 2023) are expected to dominate discussions during the initial meeting of Use Case 4, understandably considering the significance of these developments for software developers.

These contextual and external factors, rather than posing a problem, offer an important opportunity for the KT4D team to investigate the main assumptions guiding the Use Cases: that because democracy and civic participation are in essence local and situated phenomena, a strong User-centric approach is essential.

2.3.4 Need for Flexibility

Given the distinct groups involved in the three types of Use Cases - regulators, individual citizens, and software developers - and the inherent variations in roles and perspectives even within these broader categories, a level of flexibility needed to be factored into the design of the Use Cases' general framework and structure. This was essential while still maintaining an overall sense of consistency and coherence, as elaborated in Section 2.4 below.

To manage this, a series of Use Case meetings were dedicated to navigating this tension. During these sessions, all partnering entities collaboratively established a shared structure and deliberated on the customisation of specific activities. For further insight into how both flexibility and consistency are upheld, refer to Section 5 below, which provides an extensive account of the first meetings of the four Use Cases.

Regarding flexibility, another aspect taken into consideration is the need for the materials provided during

meetings to avoid being overly directive. This measure aims to mitigate confirmation bias and the potential to stifle certain viewpoints. To foster critical thinking and accommodate various perspectives, including those that might be unexpected yet highly valuable, activities (such as prototyping exercises) and questions were designed to promote participants' free contributions within a non-judgmental and collaborative environment.

2.4 Temporal Structure of the Meetings of Each Use Case

The four use cases will follow a common methodology, based upon three interaction points between the project and their target stakeholder groups, at the start (M10), middle (M20) and end of the project development (M32). The groups will be built up out of the host partners' current projects and networks, and have been assigned to localities to capitalise on partner networks and local community assets, such as the presence of technology multinationals in Dublin or of policymakers in Brussels.

These first interactions will have the dual purpose of making the policy and software development target groups aware of the KT4D framing of issues of advanced knowledge technologies and democracy, and of soliciting from these individuals and groups their input to KT4D's development processes, in particular those of WP5. The first interaction for each group will be based upon the model of Participatory Design, in which participants are guided in the course of a half-day workshop to consider the goals of the project and prior work in KT4D's field of interest as related to the specific KER that user group has been recruited to contribute to (see Figure 1 above for a mapping of these KERs to user groups). Participants will be prompted via a series of prompts and co-creation exercises to offer structured and unstructured input to the project KERs.

The second interaction for each group will be constituted as a one-day Digital Democracy Lab, in which participants will be guided through a series of hands-on exercises working with the components of the WP7 demonstrator platform and actively exploring the ethical and technical issues inherent in such platforms (under a clear data sharing agreement for the purposes of the one-day, closed world, exercise): downloading their own personal data (eg. from Google, or using fake personas); informed anonymised sharing of data through a data commons, creating and understanding ML-based training data, proposing and testing user profiling options, data retention and the deletion of personal data, etc. Each exercise will be accompanied by an opportunity for reflection on the affordances and constraints (technical, security, legal, regulatory, ethical and personal) of the technological components. Each lab will be tailored somewhat to match the needs and interests of the participants (for example software developers versus citizen education providers). The final interaction will complete a Valorisation loop, presenting near-final project results and soliciting feedback to enhance exploitation after the project in a 2-3 hour interactive briefing. We will issue reports to all participants after each workshop, and a final policy brief at the end.

The final interaction between the KT4D Team and the target groups engaged in the Use Cases will in each case be based on the need to valorise and test the components the project team has built. At that point, pre-final versions of our KERs will be available for interrogation and discussion, and our interactions with our established communities will provide us with an essential feedback loop to our future users to ensure the project results at the time of their final submission are in line with their practices and values.

2.5 Differences Between the Four Use Cases

The four Use Cases will each follow a common blueprint, though they will not be identical, due to issues of language and cultural localisation mentioned above, as well as the need to tailor the content to the predicted knowledge baseline and concerns of the target groups involved in each Use Case and event. Each interaction will involve 10-12 participants: keeping the numbers low will allow us to guarantee active attention to participant dynamics, but will allow us to directly access up to 140 individuals in the course of the project. While we may have individuals who take part in all of the events for a given Use Case, we also recognise that it is equally likely that most individuals may only have the time or availability for one, which our contingencies for recruitment account for.

While the shared goal of all the Participatory Design Sessions is to understand current and desired future user practices so as to align design to their requirements, each Use Case has a peculiar structure and a set of goals designed to best suit the participants' needs, interests, and skills. The Brussels-based Use Case will be centred around the concerns, discursive norms and values of policymakers, who are responsible for finding appropriate compromises between differing positions regarding the socially optimal outcomes and the need to keep regulation generalisable and favourable to innovation actors. As such the meetings will not only focus at times of specific policy-facing KERs, but will also take a distinctive approach to the Digital Democracy Lab, observing both the direct level of individual participation but also the meta-level of how these responses and requirements align with or highlight gaps in current regulatory approaches.

The two citizen-facing Use Cases (UC 2 in Madrid and UC 3 in Warsaw) will similarly harness the central framework while also accounting for local differences. The most notable difference will be one of language, with these Use Cases being conducted in Spanish and Polish respectively, so as to allow the participants to approach issues of civic participation in the language of their daily community interactions. The Use Cases may also differ in how they respond to local events (such as the Autumn 2023 elections in Poland) and the affordances of the organisations convening them.

Finally, Use Case 4 will be a particularly sensitive one to convene, as the goal is to ensure feedback from software developers who may not consider the impact of AI and big data as a responsibility of their job role, research area or employer. In this case, the challenge will be to ensure that they feel the meetings are relevant for them, and that the discussions can be embedded not just in meta-level considerations of democracy, but in the details of technological systems, and the epistemic cultures of software designers.

2.6 Cross Project Embedding of the the KT4D Use Cases

When the Use Case structure was being designed for the project, a number of alternative modes for embedding its activities into the project were evaluated. The obvious approach that was finally rejected would have been to have a specific Work Package bringing together all of the Use Case- related activities. This structure was rejected, however, as the cost of this clarity and coherence from a structural point of view would have come as the cost of a risk that the Use Cases would not be intrinsically integrated into the results and development of the project as a whole. It would have been not only detrimental, but ironic indeed, had the value placed on user input and co-creation become diminished because the user interaction itself had become siloed as a particular partner's job, rather than an essential component of

everyone’s work. For this reason, the tasks required to deliver the four Use Cases have instead been fully embedded into the KT4D project structure, with tasks from each Work Package (WP) directly contributing to aspects of the design and delivery of the Use Cases. Table 1 describes each task associated with the Use Cases in KT4D.

Task	Name	Description
T1.5	Coordination of the Combined Use Case Methodology	To ensure coordination between the contributing WPs, towards a coherent and shared approach across the Use Cases, and across the three years of the project.
T2.3	Coordination of the meeting events ensuring coordination between the contributing WPs Events management, including Use Case consultation and co-creation meetings	To support the delivery of the events and share the results with the wider community in post-event reports and dedicated campaigns sharing insights and results in an open manner. This task will also deliver two KT4D Impact Events, aimed to foster wider uptake of the project results among its target groups and wider stakeholders
T3.4	Interaction design for Participatory Design Session, Use Cases 2 and 3	This task will consist of the co-creation of an agenda, discussion guide, and interaction design for the first interactions of the groups to be convened for Use Cases 2 and 3.
T4.4	Interaction design for Participatory Design Session, Use Cases 1 and 4	This task will consist of the co-creation of an agenda, discussion guide, and exercises for the first interactions of the groups to be convened for Use Cases 1 and 3.
T5.3	Interaction design for final Valourisation Meetings, Use Cases 1 and 4	This task will involve the final interaction between the KT4D Team and the target groups engaged in the Use Cases will in each case be based on the need to valourise and test the components the project team has built. This valourisation process for Use Cases 1 and 4 will be designed and delivered within this task
T6.3	Interaction design for final Valourisation Meetings, Use Cases 2 and 3	Task 6.3 will prepare, design and deliver this interaction for Use Case 2 (iterations a and b). To validate the whole framework with advanced developments is fundamental, and so this task will: 1) define a validation methodology; 2) execute the interaction; 3) analyse the results; and, 4) report back to the technical team in order to implement final improvements in the framework.

T7.3	The Digital Democracy Lab	This task will develop and document the Digital Democracy Lab concept and implementation model, including a handbook for deployment across sites and contexts.
------	---------------------------	--

Table 1 KT4D Tasks Contributing to the Use Cases

3 Key Concepts and Shared Glossary

As the Use Cases will be one of the points in the project most dependent on inter- and transdisciplinary exchange, this task will also compile a project level glossary of agreed terms of reference for the purposes of our collaborative work (see Table 2 below).

Term	Definition
Artificial Intelligence (AI)	A particularly challenging term, very meaningful in public and policy discourse but replaced by more precise vocabulary in technical contexts. Although we may shift to meet expectations of these groups, in general we follow the AI HLEG’s publication “A Definition of AI” in speaking of “systems designed by humans that, given a complex goal, act in the physical or digital world by perceiving their environment, interpreting the collected structured or unstructured data, reasoning on the knowledge derived from this data and deciding the best action(s) to take (according to predefined parameters) to achieve the given goal.”
Big data	We take this term to mean, as per the EC’s definition within the ‘Europe’s Digital Future’ Policy area, “large amounts of data produced very quickly by a high number of diverse sources”.
Advanced Knowledge Technology (AKT)	We use this term to refer to the assemblages of advanced processing and big data not according to the kinds of methods that are used to develop them, but rather to those specific implementations or these technologies that are most likely to disrupt civic participation and democratic processes by intervening in the manner in which individuals develop their sense of themselves, others, and the world around them. This term allows us to understand AI and big data in terms of a long history of interactions between technological affordances (writing, printing, television, etc.) and cultural norms, values, and practices.
Democracy	On a basic conceptual level, we use this term to refer to a form of governance in which the will of a majority of citizens is enacted in policies, activities and regulation by their elected representatives. This highly conceptual approach is difficult to apply to the challenges brought on by knowledge technologies, however. In order to focus on the impact of AKTs on democracy, we will use the categories framed by (Bartlett, 2018), which are: active citizens; a shared culture; free elections; stakeholder equality; competitive economy and civic freedom; and trust in authority, which will allow us to see democracy through the lens of the interdependence of technological development and society.
Civic participation	The active manifestation of how democracy is achieved. This term refers to the actions that individuals can take in order to ensure their own opinions, perspectives and experiences are incorporated into the decision-making

	processes that shore up the democracy they live in. It is the interface between the private experience of the individual and their family and the public sphere, in all of its manifestations, that constitutes civic participation.
Disinformation	False, inaccurate, or misleading information designed, presented and promoted to intentionally cause public harm or for profit (European Commission, 2018).
Fake news	False stories that appear to be news, spread on the internet or using other media, sometimes with a satirical undertone, to gain attention and mislead, deceive or damage a reputation or influence political views.
Hate speech	Statements that spread, incite, promote or justify violence, hatred, or intolerance against a person or group of persons. ¹
Polarisation	Division into two sharply contrasting groups or sets of opinions or beliefs.
Deep Fakes	An image, video or audio of a person in which their face, body or voice has been digitally altered so that they appear to be someone else, typically used maliciously or to spread false information.

Table 2 KT4D Glossary

4 Identifying Stakeholders and Participant Recruitment

Engaging stakeholders in a project involves identifying those who have an interest in the proposed research project, have an effect on the project or are affected by its outcomes, and determining what interest they have in the project. It is the process of ascertaining key stakeholders (i.e. individuals or groups with a vested interest in your product or project) and understanding their relationships with each other and to the project. Embedded in KT4D are tasks that involve participant recruitment and such involvement of research participants will be central to the successful implementation of the project.

The following guidelines (D8.1 Section 5 Identification and Recruitment of Research Participants) describe the processes that used by all partners to identify participants for Meeting 1 of all the four Use cases and will be used for the following two meetings:

1. Participants were recruited either through social media and institutional networks, and/or through online platforms that connect researchers with volunteers who wish to participate occasionally in paid scientific studies, e.g., Prolific (<https://www.prolific.co>).
2. Participants such as policymakers and policy experts were mapped and identified based on openly documented existing AI governance networks and online documents, including white papers and legislation from the EU and national level. Policymakers identified were invited to take part in the Use Case or the Delphi study through email invitations based on their organisational affiliation.

4.1 Plain Language Information Sheets and Consent Forms

Templates for the plain language information sheets and consent forms for KT4D research tasks can be found in D8.1 H - Requirement, Appendix 1 and 2. As each Use Case will occur in a different European city, the

¹ <https://ks.echr.coe.int/documents/d/echr-ks/protection-against-hate-speech#:~:text=Statements%20that%20spread%2C%20incite%2C%20promote,of%20the%20rights%20of%20others> (ECHR European Court of Human Rights)

consent form and information sheet templates will be carefully tailored for each location due to variations in focus, audience and language.

4.2 Pseudonymisation

After consulting with KT4D's legal partner ICTLegal Consulting, the partners involved with the Use Cases have unanimously decided to adopt pseudonymisation rather than anonymisation as the optimal way to ensure data protection for all the participants to the various meetings.

Pseudonymisation is a security measure which is being used by the members of the Consortium in order to comply with the data security principle pursuant to Article 32 of the GDPR.² The choice in using pseudonymisation rather than anonymisation techniques is dependent on the specific situation and the level of data privacy protection which is required. Article 32(1)(a) of the GDPR recognises pseudonymisation as a way of achieving the security principle.³ Instead, data is considered to be anonymised only when it is not possible to achieve identifiability of the person to whom the data belongs.⁴

Pseudonymisation entails a process which replaces identifiable information with a key that can be linked back to the original person with extra information. Essentially, this means that the pseudonymisation of data can enable data to be identifiable where more information is provided, however, anonymisation prevents the re-identification of data.⁵ The members of the Consortium will carefully evaluate the differences and feasibility of using such security measures and each Use Case Leader will decide the most suitable method for their Use Case and the iterations involved.

Here are the guidelines and Pseudonymization Guidelines for Use Cases:

[KT4D's Pseudonymisation Guidelines for Use Cases](#)

These documents have been prepared by the ICTLegal team, and all partners involved in the Use Cases are expected to abide by them.

5 Descriptions of the Use Cases

5.1 Use Case 1 (Brussels): Governance Framework, Policy Roadmap and Recommendations

5.1.1 General Goal and Structure

The overall goal and description of Use Case 1 is to identify central questions underlying AI governance, identify emerging priorities and understand how European policymakers can and should regulate the disruptive nature of general-purpose AI democratically. Following the joint methodology of KT4D Use Cases, it consists of three interactions with its key stakeholder group, which includes policymakers and policy-facing CSOs.

² Ibid, 32.

³ Ibid, Art. 32(1)(a)

⁴ L Feiler, N Forgo, M Nebel Article 4(5). Pseudonymisation', in The EU General Data Protection Regulation (GDPR): A Commentary (Christopher Kuner and others (eds), online edn, Oxford Academic, 2020).

⁵ Regulation (EU) 2016/679, Recital 26.

Emerging technologies such as AI not only raise questions in terms of their regulation, but also how these technologies themselves affect governance. Problems generated by these emerging technologies, such as the centralisation of power, opacity and the speed of change inherent to current AI advances, all generate risks to democracy. To tackle these challenges, we need to think carefully about the role of governance at national, supranational and global levels when designing governance accordingly. By engaging with policymakers, the use case contributes to the KT4D's research through generating understanding of how to better address and regulate the societal challenges and potential of AI and big data. On the other hand, it complements the results of Use Cases 2-4 from a top-down perspective, enabling further analysis on how citizens' and developers' concerns come together, or deviate, from policy level questions.

In recent years, the EU has been active in positioning itself as a global regulatory leader on AI and enacting new policies and regulation for data and AI. At the same time, there is national AI legislation and guidelines being introduced outside the EU, making it an intriguing point in time for Europe when it comes to forming its own technology political poise globally. Based on scoping of the recent, ongoing and emerging policy developments, their critical review and assessment of their positive potential, the use case will be developing a framework for democratic AI governance.

The governance framework builds upon existing AI governance and algorithmic accountability guidelines to support democratic and trustworthy use of AI, taking into account fundamental rights and algorithmic bias. As part of the process, we will critically evaluate how democratic values such as equality, transparency and privacy are realised in different AI policies. The governance framework, policy roadmap and recommendations complementing it are a combination of the background research and the findings from Use Case 1. The framework seeks to provide guidelines for assessing risks to democracy when regulating AI in the context of European values. It protects citizens from the possible negative impact of these technologies on fundamental rights and democracy, ensuring public trust in AI. While European citizens stand to benefit from the framework, its main intended audience are policymakers and policy-facing CSOs that are active in AI policy.

5.1.2 Outline of Use Case 1 Meeting 1

The first interaction is framed as a roundtable instead of an intense workshop because the aim is mostly to tease out discussion, inputs and insights on AI regulation and policy instead of co-creating solutions at this point. The aim of the first interaction is to:

- Explore and validate the risks and opportunities AI poses to democracy
- Identify blind spots and gaps in the existing EU governance approach to AI with respect to democracy
- Explore and gather existing positive examples of possible governance responses to AI systems, including general-purpose systems
- Crystallise questions for a future Delphi study on AI governance

Before the event, the Demos Helsinki research team identifies potential participants via its existing networks, the KT4D project networks and stakeholders, via mapping of key EU Commission entities, CSOs working actively with EU digital policy and regulation, and researchers with expertise in European AI policy. The research participants were selected based on the organisation they represent (identifying relevant

organisations), and / or their active engagement with the research topic based on e.g. publication record or other publicly available information. All contact information is collected from publicly available information (e.g. company or organisational information) on websites. The participants are contacted via email without sharing any contact information with other participants, using either a bcc function in an email. The emails and the sign up form include a privacy notice and a link to the Demos Helsinki Privacy Notice, and at the event itself, the participants are asked to sign a consent form.

5.1.3 Methodology and Activities of Use Case 1 Meeting 1

The roundtable took place in Press Club Brussels Europe on 26 October 2023, with the help of the KT4D partner Beyond the Horizon.

Agenda:

- 13.00 – 13.15 Welcome & Introductions
- 13.15 – 13.30 Project introduction and goals
- 13.30 – 14.15 Session 1: Conflict between democracy and AI
- 14.15 - 14.25 Break
- 14.25 – 15.05 Session 2: Challenges, gaps and weak spots in AI regulation
- 15.05 – 15.50 Session 3: Governance solutions
- 15.50 – 16.00 Conclusions and ways forward

The Demos Helsinki team acts as the facilitators of the event. All three team members have individual roles in presenting specific slots throughout the day. The two small groups are facilitated by Johannes Anttila and Atte Ojanen, while Anna Björk has the main responsibility to take field notes from different discussions. The participants write their points down on sticky notes which are added to PESTEL based canvases during the small group discussions. After the event, the sticky notes will be added as digital notes to complement the field notes.

The sessions include the following discussion points and questions:

Session 1: The conflict between democracy and AI

Session opens with a presentation (Figure 2) on a typology of risks caused by AI to democracy, based on literature review and analysis by the Demos Helsinki team.

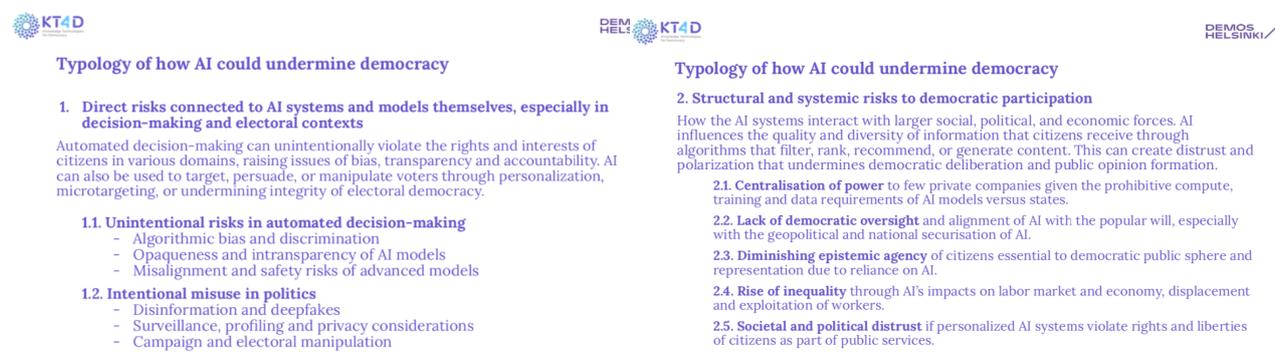


Figure 2 Typology of Risks Caused by AI to Democracy

The participants discussed issues such as:

- Does the typology on direct and structural risks make sense to you? What might we have missed?
- Widely construed, how would you approach the risks of AI to democracy?
 - Can also be approached in terms of a) democratic institutions or processes most impacted by AI or b) in terms of democratic values: e.g. representation, deliberation, participation, transparency, and accountability
- Which of the threats or risks resonate most to you? Which do you think are the most important, pressing or overlooked of them?

Session 2: Governance challenges: Weak spots & gaps

Session to be opened with a presentation (Figure 3) on the current policy and regulatory landscape in the EU and beyond.

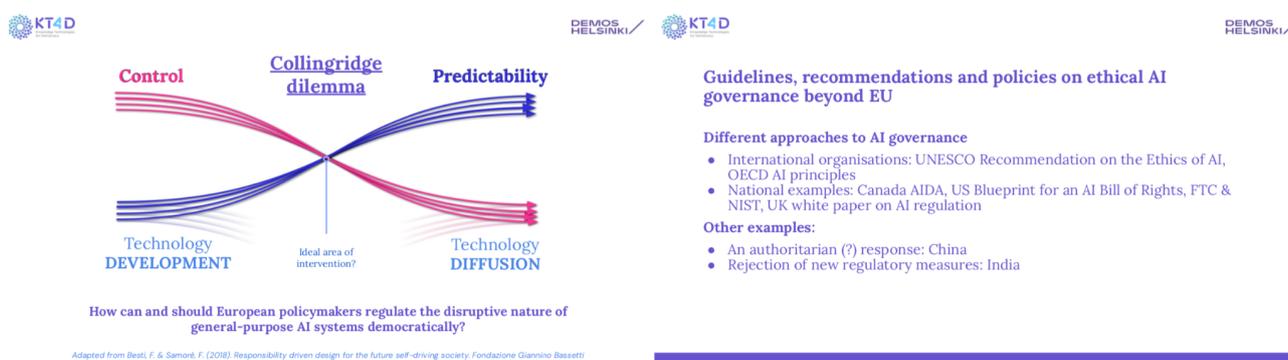


Figure 3 Current Policy and Regulatory Landscape

The presentation laid the ground to facilitate discussion of issues like:

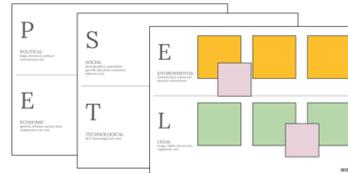
- To what extent is the use case and product legislation approach enshrined in the AI Act fit for general-purpose AI systems?
- Should there be restrictions on the development of increasingly larger models?
- How can we balance technical expertise with effective democratic control in AI regulation; e.g. civil society perspective in standard-setting?
- What should be the role of citizen participation and NGOs in AI governance?
- Can regulation that focuses on individual rights uphold and safeguard collective democratic principles and communicative norms?
- To what extent is the proposed regulation and its enforcement enough (e.g. GDPR vs use of copyrighted data in LLMs) or are new institutions needed (European AI Board vs Agency)?

Main question: In light of the risks identified, what are the gaps and weak spots in the current EU approach?



Discuss in groups

In light of the risks identified, what are the gaps and weak spots in the current EU approach?



Use the same canvas to mark down related gaps and weak spots on new post its

Figure 4 Gaps and Weak Spots in Current EU Approach

Session 3: Potential governance responses

The session was opened with a reference to the KT4D mission and some positive examples of the use of AI to foster democracy (Figure 5).



Fostering democracy through AI: steering our futures?

“Fostering Civic Participation in Democracy by Harnessing the Benefits of Knowledge Technologies” (the KT4D mission)

In essence: thinking about the values underlying the current AI policy and regulation, and about the positive futures we’d like to see, while generating in-depth understanding of the conflicts they include.

Example

AI is an opportunity to improve the democratic processes in our societies and to promote equality.

For example, AI could help citizens to gain a better understanding of politics and engage more easily in democratic debate (knowledge, oversight, platforms). Moreover, AI systems can be utilized to promote more equitable algorithmic decision-making in public administration.

Figure 5 Fostering Democracy Through AI

The discussion covered questions such as:

- What should be the cornerstones of the EU's AI governance approach?
- What’s next? What needs to be done to tackle the risks, the gaps, the weak spots?
- Do you have concrete examples: AI advancing democracy or desirable governance/regulatory approaches?

After the event, the participants received a thank you note including a link to a feedback form, all KT4D social media info and the event slides. They will also be requested to take part in the Delphi survey which will follow in early 2024.

5.1.4 Results of Use Case 1 Meeting 1:

The number of invitations sent to the identified stakeholders was 122. In addition, the event was promoted via KT4D and Demos Helsinki social media channels as a public invitation and in the preceding KT4D Webinar. As a result of the participant recruitment, 23 participants signed up by the deadline. In the morning of the event, 15 participants had indicated their attendance, but the final total of persons taking part was 10 (excluding organisers). Notably, the participants ended up representing CSOs whereas the policy officers from the European Commission mostly did not attend despite registering. While this was regrettable from the perspective of the composition of the group, the agenda of the roundtable and the venue were easily adjusted to suit the number of participants, resulting in more in-depth and analytical discussion on the risks, gaps and future governance possibilities. The participants were very engaged and expressed a keen interest in staying in touch and being part of the following KT4D interactions.

Some of the key points of discussion from the sessions can be summarised as follows:

Session 1: The conflict between democracy and AI

Most prominent risks highlighted in the discussions:

- Mass manipulation (e.g. disinformation, misinformation, psychological manipulation)
- Concentration of power to few US AI companies
- Potential dangers of very powerful open-source models
- Misusing current AI systems based on statistical correlation to determine causality.

Session 2: Governance challenges: Weak spots & gaps

Most prominent gaps and weak spots in the current EU approach highlighted in the discussions:

- Difficulty in enforcement: unsuitability of product legislation for AI and lack of resources for regulatory and enforcement agencies of AI
- Compatibility with collective bargaining and co-determination models in certain national contexts: how can we ensure that good practices around the involvement of workers in certain countries are ensured and taken as an example?
- Power relations and liability: unclear or simply shying away from liability for providers of general-purpose AI systems versus the users/deployers of the systems.
- Sustainability and material footprints of AI systems

Session 3: Potential governance responses

- Given the digital and green transition agenda of the EU, sustainable AI could be a leverage point by which also to advance smaller, more transparent models that do not threaten democratic practices
- Examples could be taken from the regulation of life sciences and medical industry: committees that also involve different stakeholders, including patients.
- Bringing in more participatory approaches to AI governance and standard setting.
- Need to massively increase resources in the field of responsible AI governance.

The first Use Case meeting gave fruitful validation of initial research and also provided data for further analysis and shaping of the Delphi survey on the future of AI governance. The reach and keen engagement

of policy-facing CSOs was positive. The Delphi survey to be sent to the invited participants and others will lend itself to further the reach to other key stakeholders from the policy side. Results of the Delphi survey organised in 2024 will feed into the design of the next interactions of use case 1 as well the framework and roadmap for democratic AI governance.

5.1.5 Plans for future interactions

The Conclusion of this document includes future interactions and deliverables stemming from Meeting 1 of UC 1 and the results of Meetings 2 and 3 in Brussels will be included in future versions of this Deliverable.

5.2 Use Cases 2 & 3 (Madrid and Warsaw): Narrative Based Simulation Game and Two Interactive Explainers

5.2.1 General Goal and Structure

5.2.1.1. Session's objectives

The focus of these sessions, both in Poland and Spain, are twofold. One refers to concept exploration in the field of knowledge technologies. Through qualitative methodologies we want to approach the ideas, fears and hopes of citizens about knowledge technologies in relation with technology. The other is more practical and refers to the format of educational materials and games that the KT4D consortium will develop through collaborative design methodologies and User Research (UX) techniques.

The objectives are to identify perceived threats to but also opportunities for democracy related to AI and big data and, further, to create educational materials and games that help with critical literacy so as to overcome problems in democracy in relation to technology. Meetings are organised in a two-step approach: a pilot study meeting (in Cracow, Poland) and the proper Use Case meetings (Warsaw and Madrid). All processes (participants recruitment, meeting conduct, etc.) adopted for the pilot study helped the researchers to better perform the proper Use Case meetings. Through these actions, we determine the topics and preferred formats among users, to accordingly design the materials and games. Use Cases 2 and 3 are very similar and both will be inspired by the pilot study meeting in Cracow. In addition, the main difference between the use cases in Spain and Poland is that the use case in Spain is conducted without a separate pilot in that country.

5.2.1.2. Outline of the Use Cases 2 and 3.

The total number of participants in the pilot Use Case is 8 and the Warsaw Use Case gathered 20 people, carefully selected based on different criteria (such as age, gender and other relevant background characteristics according to Krueger 2014, Morgan 1996, Onwuegbuzie 2009). These include:

Age diversity: including young adults (18-25), elderly (+65) and general population (25-65).

Vulnerable groups: refugees/migrants, people with disability, and LGTB representatives, considering that two representatives of any of such collectives per session would be highly desirable, although due to the sensitive nature of the questions aimed at recruiting these groups, we did not directly take actions to get these people in the pilot and the Use Cases, and rather we came to CSO's working with and/or representing such individuals.

Politically active/inactive population: measured by identifying themselves as interested in politics vs indifferent or not interested, or other similar categories, usually correlating with age.

Diverse digital skilled population: People with diverse Digital Literacy (which correlates with age).

- Stakeholder mapping and participant recruitment

At the stage of the pilot study, participant selection is conducted in a two-way manner. First, the researchers drew up a list of organisations that may best represent the interests of vulnerable groups (taking into account: ethnic/cultural minorities, sexual orientation minorities, people with disabilities, elderly people). Second, the researchers contacted individuals directly who, due to their engagement and experience, were likely to volunteer to contribute to the Use Case.

The research team reached those people through the Internet by contacting them directly or through decision-makers in chosen organisations. After a rigorous review of the results and lessons from the pilot Use Case meeting, the researchers applied an analogous approach, after required amendments, for the recruitment of participants for the proper Warsaw and Madrid Use Case meetings.

Invitations are sent approximately a month before the session and then followed up by a reminder a week before the session.

- Date and venue

Use Case 3, pilot meeting in Warsaw

The pilot study meeting took place on October 6-7 in Cracow, Poland. The pilot use case study took place during the first day, followed by a methodological workshop conducted the next day. The venue was the conference room at the IRMIR building.

The proper Use Case study took place on November 9 in Warsaw, Poland. The venue was *Wernisaz Cade* in a location that is easily accessible by public transport and car. The event started at 10 am and concluded around 2:30 pm.

Use Case 2, meeting 1 in Madrid

The first UC2 meeting took place the 31st of October in Madrid, in La Casa Encendida, a well-known cultural space in the centre of the city. The session was conducted starting in the late morning and lasted until before lunch time, with catering provided for the participants at the end.

5.2.3 Methodology and Activities

5.2.3.1 Prior to the Use Case meeting

Prior to the session, upon accepting an invitation, participants were sent a brief set of information on what to expect from the Use Case sessions, what contribution is expected from them, why this is important as well as a summary of basic facts about the KT4D Project and the team running/organising the Use Case.

Upon arriving at the Use Case venue, participants were given the privacy statements and consent forms agreeing to take part in the meeting.

5.2.3.2 Welcome

The researcher leading the session shortly presented the research project, its objectives and the project's goals:

- identifying perceived threats to and opportunities for democracy that tackle critical digital literacy,
- The relationship between the rapid development of technology and civic engagement
- The role of values and norms in this complex interaction.

The main goal of the Use Case meeting is to understand the citizens' perceptions and attitudes towards AI and Big Data as well as co-create educational materials that will be easy to understand for the general population.

Organisers then explain what participants will gain from the meeting:

- Expand their knowledge and awareness about digital technologies and their relationship with democracy.
- Learn about co-creation co-design methodology.
- Get hands-on professional experience in understanding the relationship between these processes.
- Empowerment in relation to the challenges to democracy from digital technologies.
- Networking with the rest of participants.

Agenda:

The entire meeting will take the form of a discussion and will also include a series of exercises and several practical tasks.

1. Introductory part: Focus Groups on intuition, definitions, key aspects of AI, the impact of AI on democracy and the real world.
2. Lunch break.
3. Identification of characteristics that educational materials about AI should have through validation of existing materials or searching for new ones.
4. Co-creation of an outline for such materials.
5. Closing

5.2.3.3 Warm up

The researchers asked participants to write down their perception of AI (i.e. positive, negative or neutral). Then their answers were collected and counted. The next question asked was whether the participants have encountered AI in their daily lives, if so, where, and in what form. The moderator collected the answers written on the notes and categorised them on a flip-chart. Next, participants were asked to explain how they

understand AI in their own words. Their discussion was confronted with a definition of AI provided by the chatGPT. After this discussion, the moderator announced the results of the first question asked about the perception of AI. Finally, participants were asked to choose those technologies/uses of AI (displayed on a flip chart) that will have the greatest impact on democracy and civil rights (see Figure 6 below). This exercise is followed by a discussion of the results.

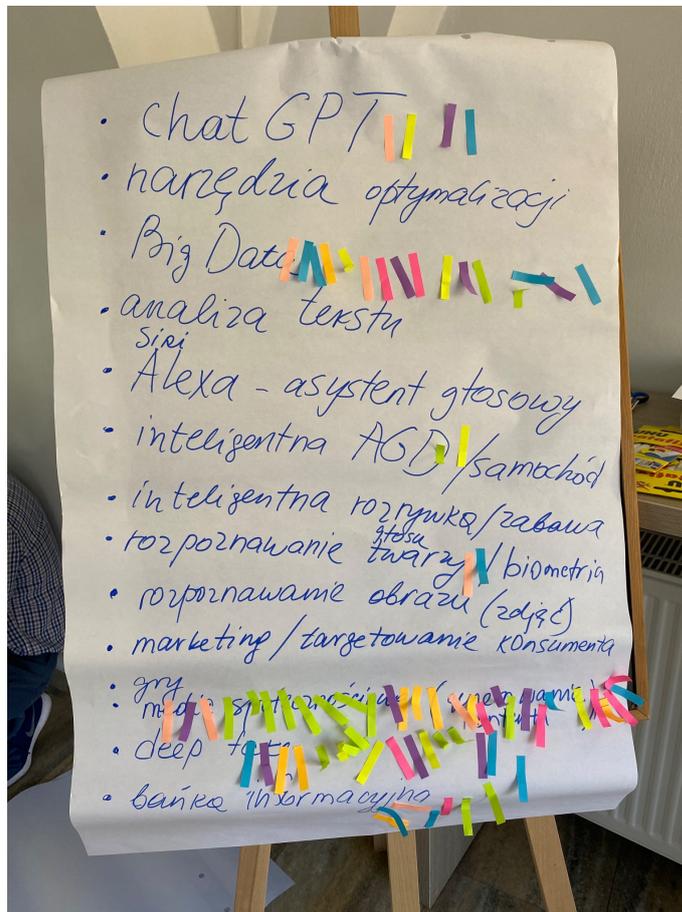


Figure 6 Technologies/uses of AI that will have the greatest impact on democracy and civil rights

Notes: Participants used colourful sticky-notes to indicate the technologies/use of AI that in their opinion would have the greatest impact on democracy and civil rights. The meeting was conducted in Polish and both the participants, and the researchers used Polish to create the flip-charts and the rest of the materials.

5.2.3.4 Discussions - in groups

The researchers split participants into groups of 3-5 people inviting them to discuss the impact that AI will have on the world in 10 years taking into account the following areas (see Figure 7 below):

- Work.
- Education.
- Entertainment.
- Civil society, politics, and democracy.
- Health.

Each group focuses on only one area and a flip-chart to write down their ideas. Afterwards, each group presents their ideas to all the workshop participants and people from other groups supplement their list of ideas with their own. After all the flip-charts are discussed, the moderator leads a discussion on whether these changes are positive (+), negative (-), neutral or perhaps they are ambiguous (+/-). Finally, the researcher initiates the open discussion asking the following questions:

- Should there be limits on the development of AI?
- Who should control AI?
- What do you fear most: the government or corporate control?
- Would you want to be told if something is created by AI (e.g. art)?



Figure 7 In-group discussions

The researchers asked participants to use their own devices (smart phones, tables, laptops) to find informational materials about AI. The specific task was to find materials explaining what artificial intelligence is. Gathering this information will help the researchers to further fit materials/games to the participants' needs.

The following questions were asked in order to gather more specific feedback related to:

Sources of information:

- What sources came up first in your search?
- What platforms/ tools do you use to look for the materials?

Format and content:

- What are your initial impressions on the format and presentation of the educational materials?
- Can you identify any specific elements within the materials or games that you find particularly attractive or thought-provoking?
- Do you think the materials effectively address the issue of artificial intelligence?
- Is the game a good format for such materials?

User interface (UI) and user experience (UX)

- Do you think the interface of these materials is easy to use?
- Is the overall experience pleasant and engaging?
- Have you encountered any problems related to the size of text or graphics on the screen?
- Are there any UI elements that you found distracting or unnecessary?
- How to reach people who are digitally excluded or do not use computers/internet?

Possible changes and improvements

- Is there any specific aspect of the materials that you think needs improvement or modification?
- What changes or additions would improve the educational value or entertainment factor of these resources?
- Were there any technical problems you encountered when using the materials?

Democracy and technology

- How effectively do these materials address the issue of technology's impact on democracy?
- Do you think they provide valuable knowledge or provoke critical thinking on this topic?
- Were there specific examples or scenarios within the materials that stood out to you in terms of their relevance to real-world technology and democracy issues?
- Were there moments when you found yourself questioning your own assumptions or beliefs about democracy and technology?
- Did the materials or games provoke discussions or debates among your group members about the topics they addressed?
- Can you identify any specific insights or insights you gained from these materials or games that you hadn't considered before?
- Who should fund such education materials?

5.2.3.5 Co-creation

The researchers engage participants in a co-creation process to produce an escape room game that helps with critical literacy in order to help address challenges in democracy in relation to technology. First, the

researchers explained the concept of an escape room and the participants randomly drew a convention/basic scenario for an escape room from the following list:

- Deserted island
- Jurassic Park
- Prison
- Space station on Mars
- Big Brother house (reality show)
- House of horrors
- Abandoned office building
- Wild West
- Submarine

Second, the participants were split into groups of 4 persons and were asked to come up with a puzzle, quiz or teaser that is intended to teach people something they should know about artificial intelligence. These puzzles/quizzes should fit into the convention/ basic scenarios each group drew. After all the groups finish their work, they each present their escape room games to the rest of the participants.

Closing exercise

Participants were asked to write down on a note their attitude towards AI again (positive, negative, neutral). The votes are counted and announced.

5.2.4 Results of Use Case 2 Meeting 1: Madrid

5.2.4.1 Participants and their demographics:

The number of stakeholders contacted for the Use Case Meeting was 5 in total.

We used our own networks of Cibervolunteers, vocational training schools, and other organisations, such as an organisation for people with learning disabilities, making sure that everyone attending the event was over eighteen years old.

We had over 46 confirmed participants, and in the end 41 turned up and took part in the workshop, (excluding organisers, facilitators and media colleagues) which was a fair number that enabled a good development of the session.

However, regarding demographics, young people were overrepresented in our sample, since we invited many schools to come to the event to ensure participation, and in the end many of them were interested in attending. This fact biased the results of the meeting towards a younger perspective, but enabled a deeper understanding of their attitudes and opinions on AI, and their input in the development of educational materials and games related to the topic.

Related to the layout of the participants involved, facilitators organised them in four groups, aiming to achieve a heterogeneous sample in each of them, in order to get richer outcomes from the session. These four groups remained throughout the workshop guided by a facilitator from CIB per group, in charge of conducting the workshop and gathering the data/ information needed.

-

5.2.4.2 Focus groups results

5.2.4.2.1 General discussion about the AI and what it is

Participants shared views and personal experiences of their use of AI, and mainly conceived that AI is a technological tool to facilitate daily life, and that it performs according to how you train it.

Participants debated about ethical concerns on AI, leading on the one hand, to a need to regulate the technology, and impose certain limits, and on the other hand to expand through education more knowledge so that the population is well informed about its proper use and potential risks.

However, the majority of the participants agreed that it would have a major positive impact on the societal level, being the area of Health, the one that would benefit the most with the further development of AI.

5.2.4.2.2 In group discussions about different areas AI will have an impact in 10 years

The participants come up with a list of challenges when discussing the impact of AI on “Health” in 10 years.

The main observations from their discussion are:

- Mild illnesses will be easier to self-diagnose.
- Healthcare will be more effective but there will be more restrictions.
- Health personnel will arrive less prepared due to the help of AI.
- Some minor illnesses may be treated entirely by AI, thus leaving more staff free.
- It will help the interpretation of medical tests in order to give a more accurate diagnosis.
- I would like the elderly to facilitate their medical procedures and help them feel more integrated.
- Help the patient understand their diagnosis, situation, and doctor’s explanations.
- Helps diagnose diseases more quickly and find more cures.
- Help or prevent future illnesses.
- I believe that healthcare will benefit the most from these technologies.
- Faster and more accurate diagnoses and surgeries with less intervention and less risk.
- Robots’ inclusion in hospitals.
- Creation of artificial vital organs.
- Help create new prescription drugs.
- Psychological treatment based on the patient's data.
- It will be possible to choose the baby’s features.
- Automate appointments, dates and events in the database.

The participants came up with a list of challenges when discussing the impact of AI on “Civil society, politics and democracy” in 10 years. The main observations from their discussion are:

- AI helps us find people with similar tastes, but at the same time this limits us in our social networks when it comes to learning about fields not related to us.

- It will help complement judges in designing laws and decision making.
- Social benefits will be automated and quicker to obtain.
- Bots on social networks that feed certain political inclinations.
- Phishing will increase.
- AI makes disinformation more massive, but at the same time it can be used to detect fake news.
- Foster connection between people from different places with the same ideology.
- Voting system from home, so nobody could manipulate results, and automate vote counting.
- There will be worse quality of information, due to the ability to segment the type of user and know how to capture their attention.
- It won't change much; it will mainly be used to do illegal things like impersonate someone from the opposite party.
- Optimization of the political system.
- It will help to achieve a more stable economic system.
- To detect corruption.
- It will help find positive political initiatives in place in other countries to apply them nationally.
- It could help to manage overpopulation in the future.
- Crime reduction, through help in the search for criminal objectives and prevention.

The participants came up with a list of challenges when discussing the impact of AI on “work” in 10 years. The main observations from their discussion are:

- Certain jobs are going to disappear, but new ones are also going to be created.
- Provide people with disabilities with more job opportunities by enhancing their abilities.
- It will make work more efficient, and perhaps can help reduce work hours, enhancing productivity.
- Profile analysis in human resources.
- Threat to the most methodical work.
- More restrictions regarding the use of information.
- More remote work, which would help people with reduced mobility, and will help balance work and private life.
- AI integration in daily activities.
- AI will create more jobs, related to communicating with AI to achieve better results.
- More comprehensive data analysis.

The participants came up with a list of challenges when discussing the impact of AI on “Entertainment” in 10 years. The main observations from their discussion are:

- In the field of video games, it will help develop new characters, scenarios, virtual reality games.
- It will help develop code for programming.
- It will help translating to other languages, films, series, etc...
- Media: improve special effects, scenarios, scripts, production assistance.
- Create images for children's books.
- Replace traditional businesses such as nightclubs with a DJ with AI.
- The video game experience will be more personal.

- The work of referees will be completely replaced.
- The deaf will be able to hear with artificial intelligence and will be able to go to concerts.
- New virtual worlds, accessible and entertaining.
- For people with disabilities, it can improve their leisure and make them more autonomous.
- Leisure will be more personalised for each person; we will include machines in our friend's group.
- Reduce the workload of waiters.

The participants came up with a list of challenges when discussing the impact of AI on "Education" in 10 years. The main observations from their discussion are:

- We will have robots doing our homework.
- AI will be able to help people with functional diversity in easy reading (making it lighter), and will put students' documents in Braille.
- Students will be able to have an advisor with artificial intelligence.
- complete courses can be created, relegating the figure of the teacher to that of the tutor.
- Classes will be more dynamic; content can be created to read and listen to in any language and it will be easier to be self-taught
- Responsible teaching of these tools and their appropriate use when studying and working will be necessary.
- The types of tasks and the way of teaching will change.
- Personalise education to meet the specific needs of each student.
- Can help create a simpler way of teaching, for greater performance.

5.2.4.3 Materials validation exercise results

For this validation part of the workshop, we stuck to the same four groups formed before, and we presented two materials, from the proposed list, to each group in order for participants to test them and give feedback regarding the three main topics that are explained in detailed in the previous methodological section; Format and content, User Experience (ux) and User Interface. These materials were discussed and chosen as part of Task 6.1, to be presented to citizens in the sessions of Use Cases 2 & 3.

The four groups were divided in half, so that we could test both materials simultaneously, and later switch to the other resource in the middle of the session. To make it more accessible and engaging for participants, they were asked to note their feedback on the resources in post-its and stick it in three different columns linked to the three topics we were interested in researching.

The following table charts the main takeaways from participant's feedback, related to the educational material and game tested in each group:

GROUP 1	FORMAT	UX	UI
Educational material: Elements of AI ⁶	<ul style="list-style-type: none"> - Simple - A bit intrusive - Not very aesthetic - Small letter size 	<ul style="list-style-type: none"> - Lack of clarity - It doesn't take my attention - Basic - It doesn't ask for consent to save your pictures - Lack of coherence between screens - Well- explained - You can share your results 	<ul style="list-style-type: none"> - Basic navigation system - Weird interface and not easy to use - Simple, well explained but very disorganised - Is not easy to change the language
Game: Orwell ⁷	<ul style="list-style-type: none"> - Intuitive - A lot of text - Good animations - Dialogs should be automatic 	<ul style="list-style-type: none"> - The objective is not clear - Bad translation, repetitive, slow - The introduction of the history is not in line with the game experience - A lot of text 	<ul style="list-style-type: none"> - Good animation and nice drawings

Table 3 Group 1 Participant Feedback

GROUP 2	FORMAT	UX	UI
Educational material: The emotion business: who's	<ul style="list-style-type: none"> - Is not clear what the aim of the game is - It doesn't work properly - Is not available in many 	<ul style="list-style-type: none"> - I understand it better once I play for a second time - Lacks efficiency - I'm worried about my data and what is 	<ul style="list-style-type: none"> - Unintuitive interface and contains errors, - Slow - A bit confusing in the beginning - It is not available in spanish

⁶ Elements of AI is a website to enrol in Online courses about AI, created by the University of Helsinki in Finland.

⁷ Orwell is a series of episodic adventure games where players take on the role of a government surveillance officer. They must use an advanced AI system to analyse citizens' personal data, social media, and online activities to identify potential threats to the state. The game raises ethical questions about mass surveillance, privacy, and the power of technology, created by Osmotic Studios in 2016.

cashing in on your emotions ⁸	languages which is not inclusive.	going to happen with it - Disappointing, boring and lacks instructions.	
Game: Beholder ⁹	<ul style="list-style-type: none"> - Beautiful design. - Designed in 2D. 	<ul style="list-style-type: none"> - It is not saved automatically after each mission - Voice narration is missing 	<ul style="list-style-type: none"> - Non- invasive interface - Efficient

Table 4 Groups 2 Participant Feedback

GROUP 3	FORMAT	UX	UI
Educational material: Online courses ¹⁰	<ul style="list-style-type: none"> - The taking head is not didactic - The lecture’s overview video is in english and can’t be translated - Well structured. - It gives you a preview of what you will learn. - Allows to log in and save the info 	<ul style="list-style-type: none"> - Videos could be interactive. - The discount is attractive. - Is not translated - Easy to use. 	<ul style="list-style-type: none"> - Very classic - Well structured - Preview not attractive, lacks visual content
Game: Democracy 3 ¹¹	<ul style="list-style-type: none"> - The format is really old-fashioned 	<ul style="list-style-type: none"> - There is almost no tutorial. 	<ul style="list-style-type: none"> - Logos and letters are readable, but colours are plain

⁸ The emotion business: who's cashing in on your emotions is an educational material about facial recognition created by the Financial Times, a good model for the “explainer” approach.

⁹ Beholder is a game in which players are tasked with spying on tenants in a totalitarian state building. They use surveillance cameras, spy on conversations, and report any suspicious activities to the government. The game explores themes of surveillance, ethical dilemmas, and the consequences of technology in authoritarian societies. It was created by Warm Lamp Games in 2016.

¹⁰ These online courses in the platform Udemy, offer a structured approach to understanding the ethical, social, and political implications of AI, surveillance, and technology in modern society.

¹¹ Democracy 3 is a political simulation game that allows you to take control of a country and manage various aspects, including the influence of the media and public opinion on politics. It was created by Positech Games in 2013.

	<ul style="list-style-type: none"> - Is not really playable 	<ul style="list-style-type: none"> - I don't understand this type of game. - Is not intuitive. - Difficult to use and i not dynamic to configure - Easier logos and visuals. - Lack of instructions 	<p>and is not visually attractive</p> <ul style="list-style-type: none"> - Is not intuitive - Confusing icons
--	--	--	---

Table 5 Group 3 Participant Feedback

GROUP 4	FORMAT	UX	UI
<p>Educational material:</p> <p>The emotion business: who's cashing in on your emotions</p>	<ul style="list-style-type: none"> - A little bit weird - The objective is not clear - A lot of potential for the criminology or advertising field. - Is not fluid. 	<ul style="list-style-type: none"> - Easy to understand - It is positive to ask for consent to collect your data - It doesn't detect well your facial expressions - Is not precise at all - Very dynamic - Not attractive - If it is done in an intrusive way, people will necessarily be more inexpressive 	<ul style="list-style-type: none"> - Does not have a language option. - Not very fluid - Intuitive interface but improvable. - Lacks content adapted for people with learning disabilities - Lack of a variety of visual content.
<p>Game:</p> <p>Orwell</p>	<ul style="list-style-type: none"> - The idea of the game is interesting - Creates moral conflicts - Sinister concept 	<ul style="list-style-type: none"> - Difficult to understand how the game works - Good aesthetics - Entertaining, complex - There is no need for such an intrusive tutorial 	<ul style="list-style-type: none"> - Improvable interface. - Dinamic - Difficult to find the settings

Table 6 Group 4 Participant Feedback

5.2.4.4 Co-creation exercise results

Divided in the same 4 groups that the participants took part in through the whole session, the participants were asked to engage in a co-creation exercise, based on the materials they have tried in the validation phase of the workshop.

The aim was to obtain a prototype or mockup, of similar games or educational materials, about AI. The facilitators highlighted the importance of creating good quality materials, since some of them would be designed in the future.

Two groups were prompted to create educational material while the other two were asked to create a game.. Participants were given big white cardboards, markers and post-its to develop their ideas.

We gave the participants the following instructions so that they could have some guidance on where to start designing the materials:

1. Brainstorm topics that are of interest to you for the games/materials.
2. Consider how you can make these resources attractive, informative and interactive.
3. What key messages do you want to include in these co-created materials?
4. How do you imagine these materials being used to educate and engage others in your community?

Figure 8 (below) represents the first game that was sketched. The objective of the game is to become familiar with AI, to apply it properly, know its limits and prevent its overuse or misuse.

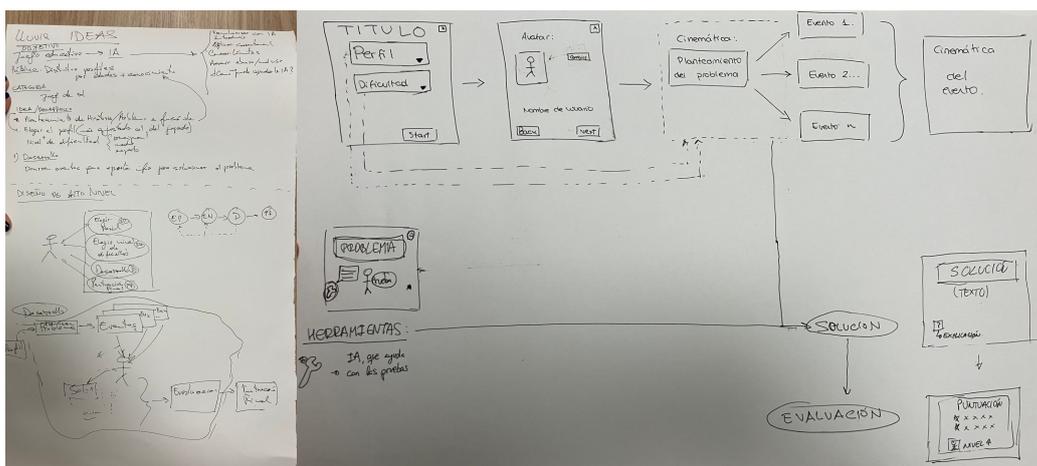


Figure 8 First Co-creation Exercise

The target group of this game would be open to a variety of profiles, regarding level of knowledge, that will determine the difficulty of the game, and targeted to different age groups. Steps in the game:

1. Choose an avatar, which will be your character in the story.
2. You face a certain challenge, which could be a crime, a robbery, a dilemma that happens with a friend, becoming broke and so on.
3. With this challenge, you start to experience some events that will occur during the game, like information that you start to receive, you read in the newspaper... These will all be clues to help you solve the challenge you are facing.
4. To solve the challenge, helped by the events or clues, you will have to use all **AI tools** available. It would be determined if you could look for them online, or they would be integrated in the game, as smaller versions.

5. There will be an evaluation part, focused on two aspects: firstly, evaluating the accuracy of the solution provided, and how many events you had to experience to get to the solution, and on the other hand, how well you implemented AI tools.
You will receive feedback on the use of AI tools, with the aim of using them as efficiently as possible in the future.
6. Through evaluation you will receive certain points that will allow you to reach different levels and compete with other participants, progressively learning more.

Figure 9 refers to the second prototype of game created during the session, called “Virtual insanity”

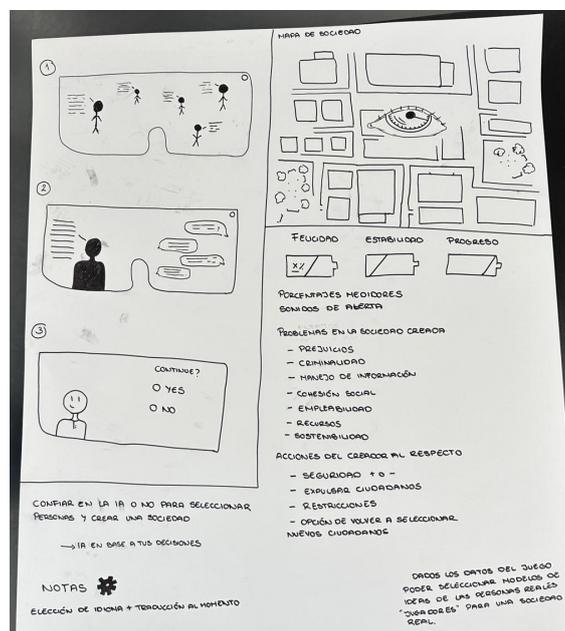


Figure 9 Virtual Insanity

- The ideas the participants suggested evolved around the topic of surveillance, very influenced by the content of Orwell’s game.
- It ended up being a game in which you have to decide whether or not to put people into the society that you are creating, and these people you are bringing in are artificial intelligence bots. You can't see their faces, you can only know information about them, they are fictitious people, created by artificial intelligence, and you are trying to decide whether or not you want to bring them into society. There is an artificial intelligence that is trained based on your decisions about the type of person you want to have in society. And you can leave it working, adding people to your civilization, while you manage that civilization.
- The second part is that with these bots that you are putting in a civilization, you are managing the technological problems that a civilization may have, such as fake news, terrorism, lack of work and so on
- It is a game like many of creating worlds, such as those of farms or ancient civilizations, but in this case, instead of food and houses, you are trying to create democratic stability and an advanced

Figure 11 represents a draft of the second educational material created, called “Learning with Fanti”

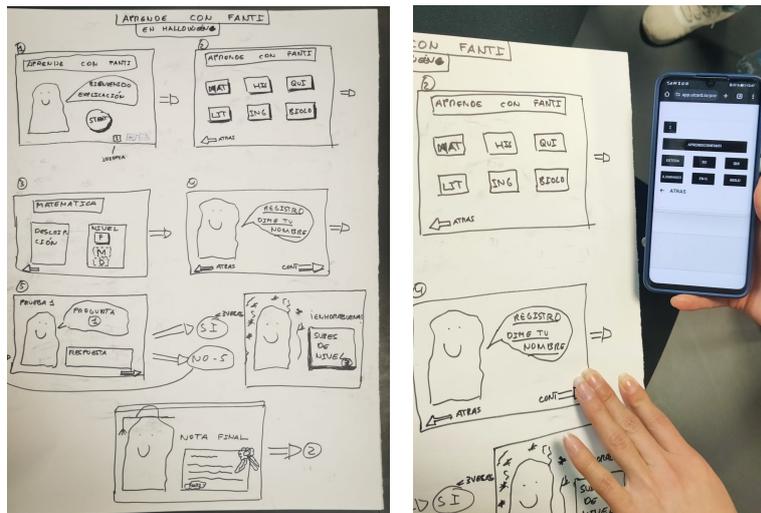


Figure 11 Learning with Fanti

- Resources aimed at children.
- There is a main character, Fanti, which is a ghost created with AI, that will follow the user along the game, calling you by your name to make it more personalised, and helping to complete the tasks.
- The user have to choose a topic, and start answering related questions, to get marks and achieve different levels
- In the end, the user gets a certificate greeting the person and explaining how much they have learnt.

5.2.5 Plans for future interactions

The Conclusion of this document includes future interactions and deliverables stemming from Meeting 1 of UC 2 and the results of Meetings 2 and 3 in Madrid will be included in future versions of this Deliverable.

5.2.6 Results of Use Case 3 Meeting 1: Warsaw

5.2.6.1 Participants and their demographics

There are 20 participants taking part in the Warsaw Use Case. Among them, 11 identify themselves as male and 9 as female. In terms of their occupation status, 12 are students, 6 are employed persons, and 2 are retired. The majority of participants are Polish and two were immigrants from another country. The youngest person is 18 years old and the oldest among the participants is 78 years old.

5.2.6.2 Attitudes towards AI at the start of the session

19 participants took part in this exercise. 12 participants have positive perception of AI, 1 - negative, and 6 - neutral.

5.2.6.3 General discussion about the AI and what it is

Participants associated the following words to describe AI: speed, innovation, replacement of human work and thinking. Although the participants noted that AI cannot replace manual labour yet. They also noted that the more reliable data AI has, the better results it produces. Moreover, participants think that AI cannot yet

distinguish between the false and true information, and is not affected by emotions and feelings as are humans.

5.2.6.4 Choosing the AI technologies/uses that will have the greatest impact on democracy and civil rights in the future

Each participant could give up to five votes and each person could give more than one vote for each item.

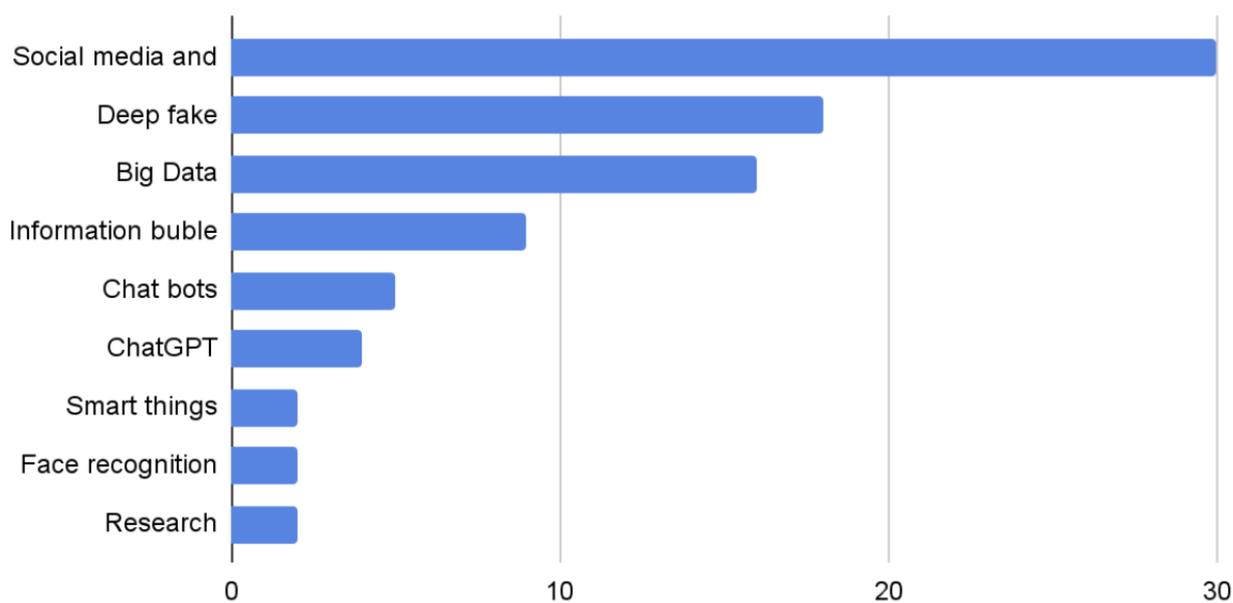


Figure 12 Votes on AI Technology with the Greatest Impact on Democracy

Inviting participants from diverse backgrounds proved to be very valuable. After the ice-breaker part of the session, we observed lively discussions, especially across the inter-generational dimension. The achieved **adversarial collaboration** meant that opposing views and experiences from participants were confronted, the younger participants had a chance to put themselves in the shoes of the elderly, and vice versa. That resulted in a constructive dialogue and enhanced creative thinking and collaboration.

5.2.6.5 In group discussions about different areas AI will have an impact in 10 years

The participants come up with the list of challenges when discussing the impact of AI on “Education” in 10 years (see Figure 13 below). The main observations from their discussion are:

- the amount of knowledge will increase,
- educational processes will start earlier,
- there will be more specialisation in education,
- automation will increase,
- inequalities and divergence in the access to education will also increase.

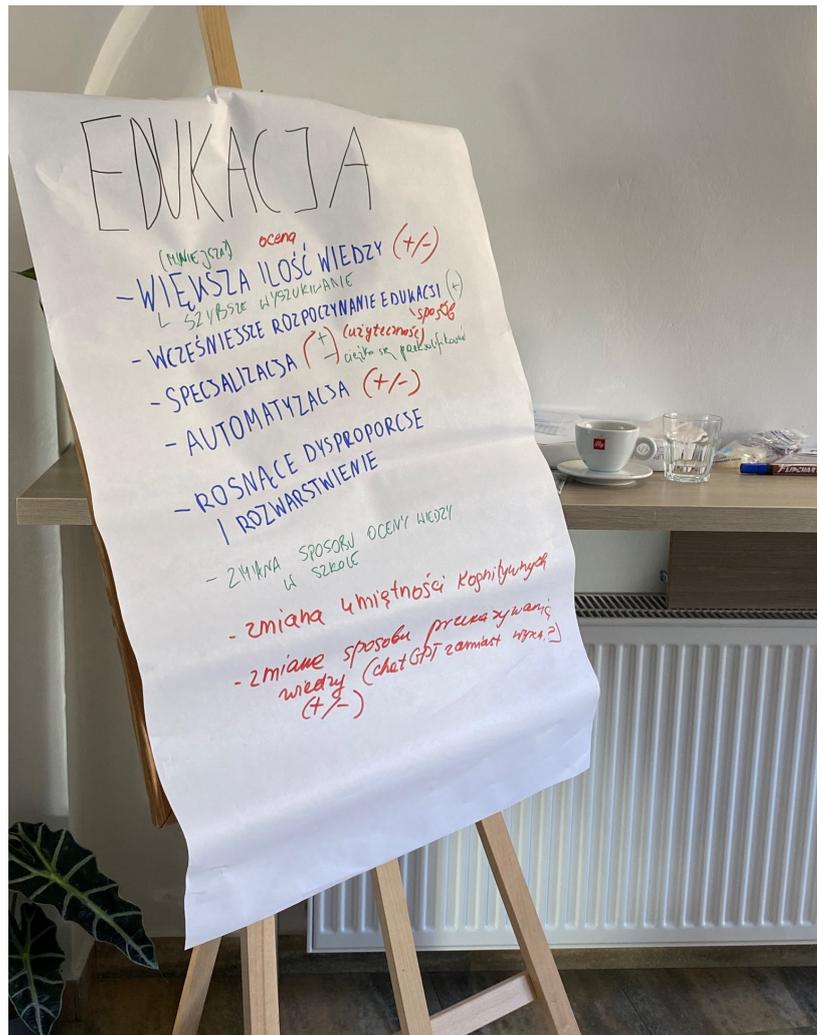


Figure 13 In-group discussion results of the impact of AI on “Education” in 10 years

The participants come up with the list of challenges when discussing the impact of AI on “Entertainment” in 10 years (see Figure 14 below). The main observations from their discussion are:

- More digital books will be available instantly in all languages,
- Remote and multi-language museum visits using more advanced technology,
- Expansion of streaming services enabling more animated art,
- Expansion of social media,
- Art is going to be made with AI more often.

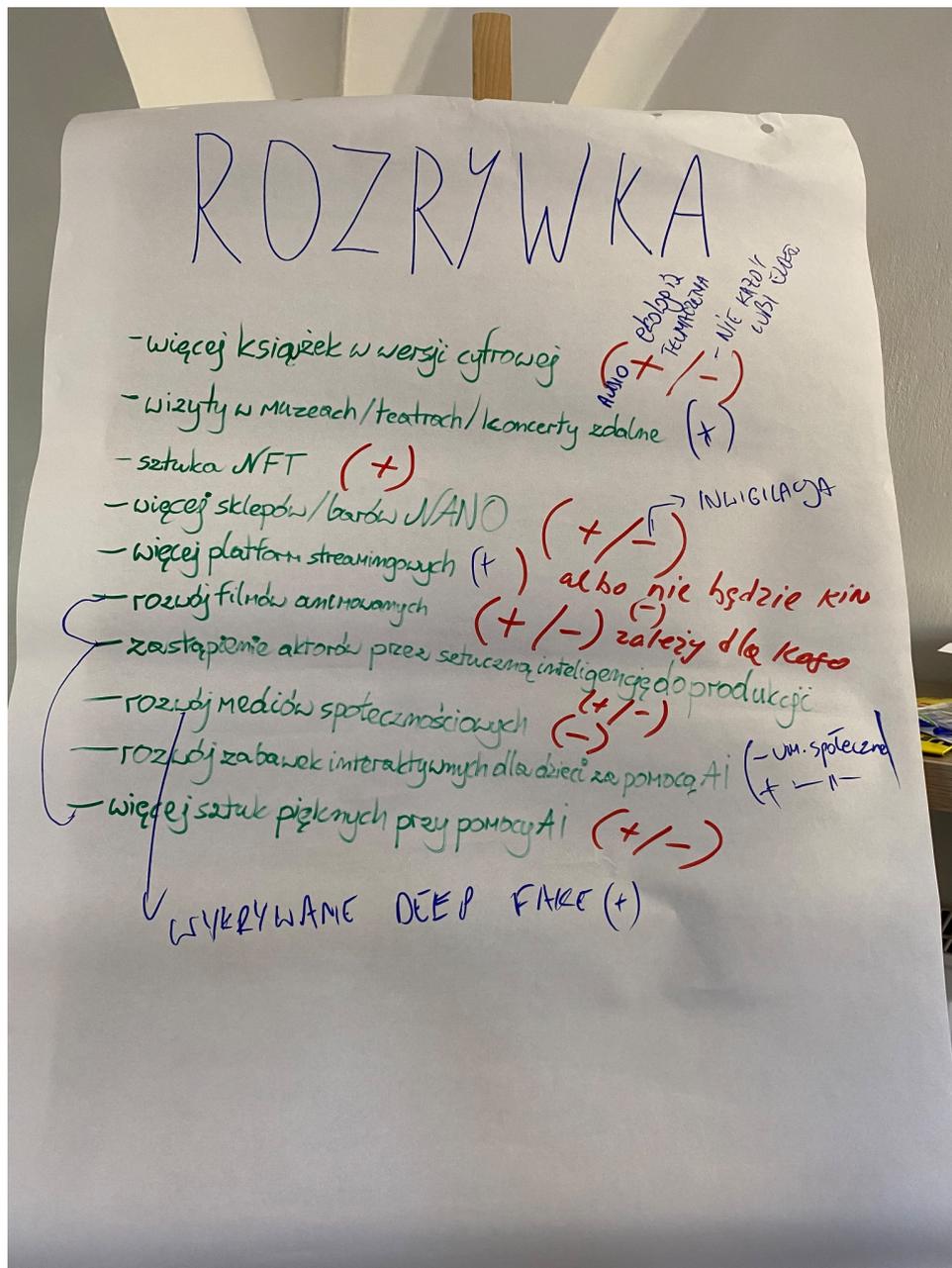


Figure 14 In-group discussion results of the impact of AI on “Entertainment” in 10 years

The participants came up with the list of challenges when discussing the impact of AI on “Health” in 10 years (see Figure 15 below). The main observations from their discussion are:

- AI can improve medical diagnostics through better analysis of, for example, imaging like X-rays.
- However, there will be more privacy and data security concerns.
- AI enables faster development of new medicines and treatments.
- Remote surgeries will become more advanced with AI assistance.
- AI may help extend human lifespan, but more research is needed to realise this potential.

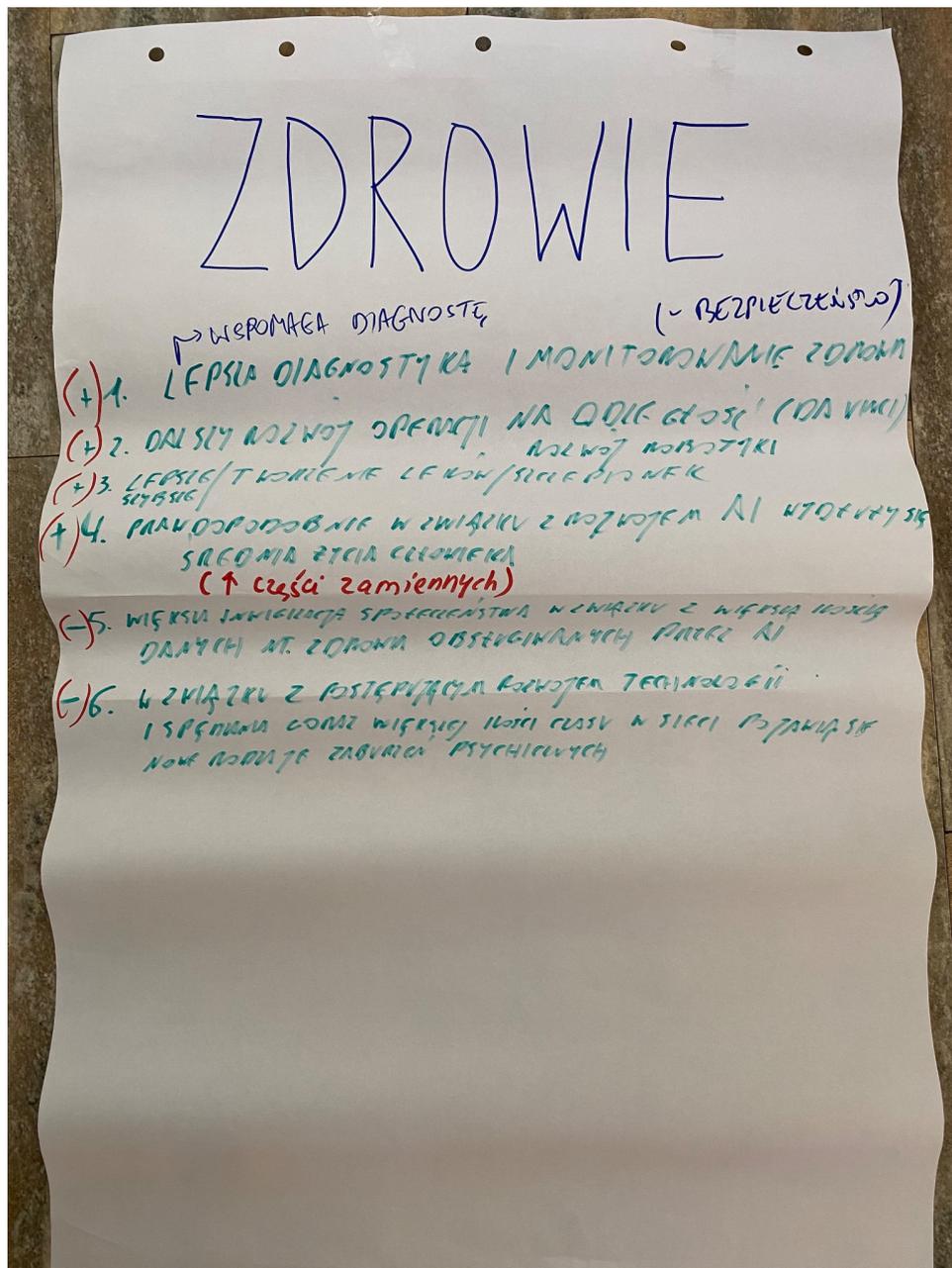


Figure 15 In-group discussion results of the impact of AI on “Health” in 10 years

The participants came up with the list of challenges when discussing the impact of AI on “Work” in 10 years (see Figure 16 below). The main observations from their discussion are:

- AI will replace humans in some sectors of the economy (e.g. in transportation can improve safety but may lack alternative job options for displaced workers, especially those more advanced in age).
- AI will make it easier to summarise large amounts of data.
- Shorter work weeks that will require labour laws to be updated.
- increase in the importance of jobs that require direct human contact.
- Improvements in the speed of automated processes and their safety.

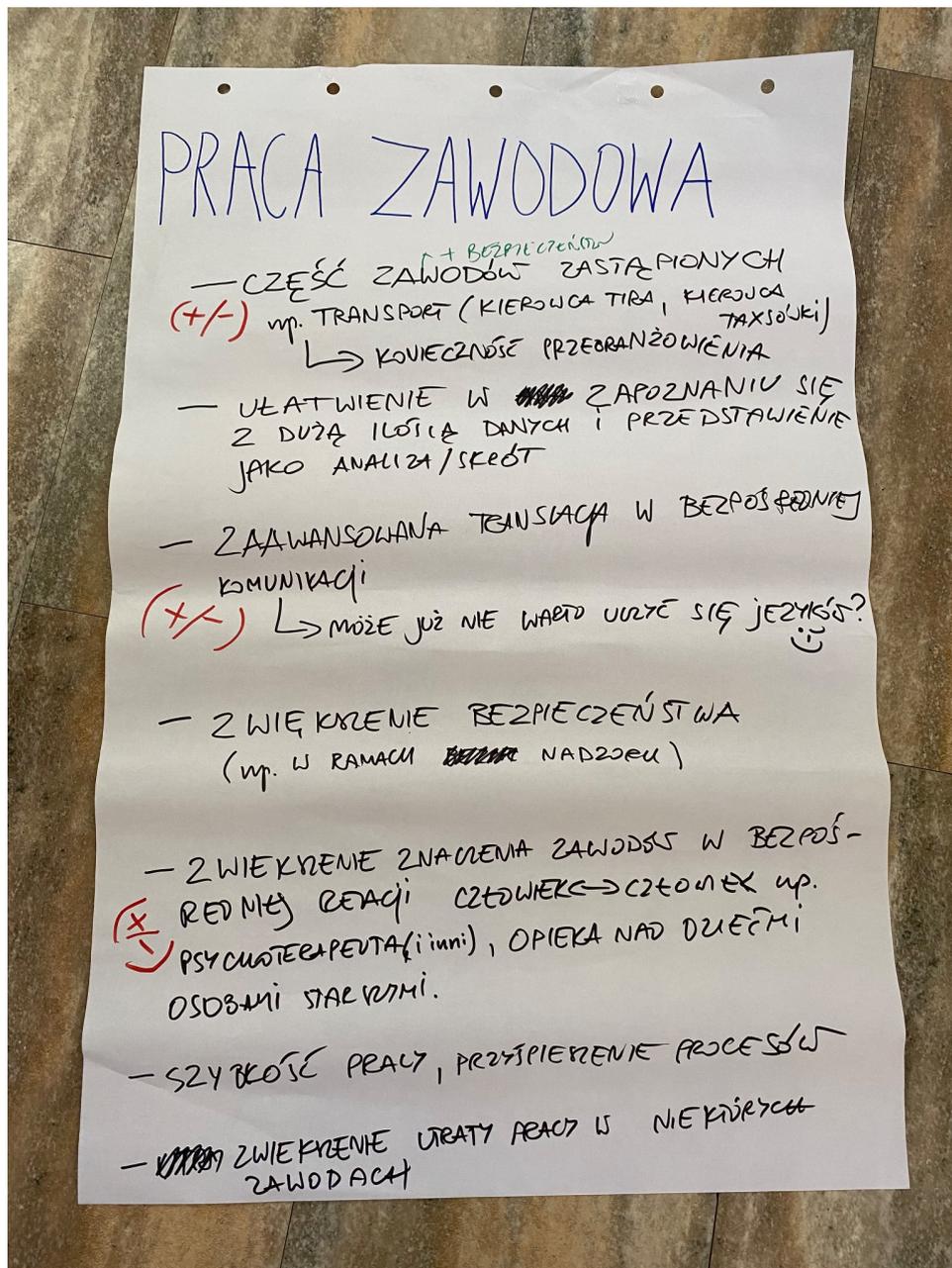


Figure 16 In-group discussion results of the impact of AI on “Work” in 10 years

The participants come up with the list of challenges when discussing the impact of AI on “Civil society, politics, and democracy” in 10 years (see Figure 17 below). The main observations from their discussion are:

- Online voting enables more fluid democracy but raises election security risks.
- Increased access to information and ease of expressing views leads to greater pluralism. However, anonymity also enables greater manipulation.
- More frequent elections and voting (referendums) in the future but needs oversight against fraud/hacking.
- Corporations are likely to find ways around regulations.
- More fear of government abuse of power than corporations.

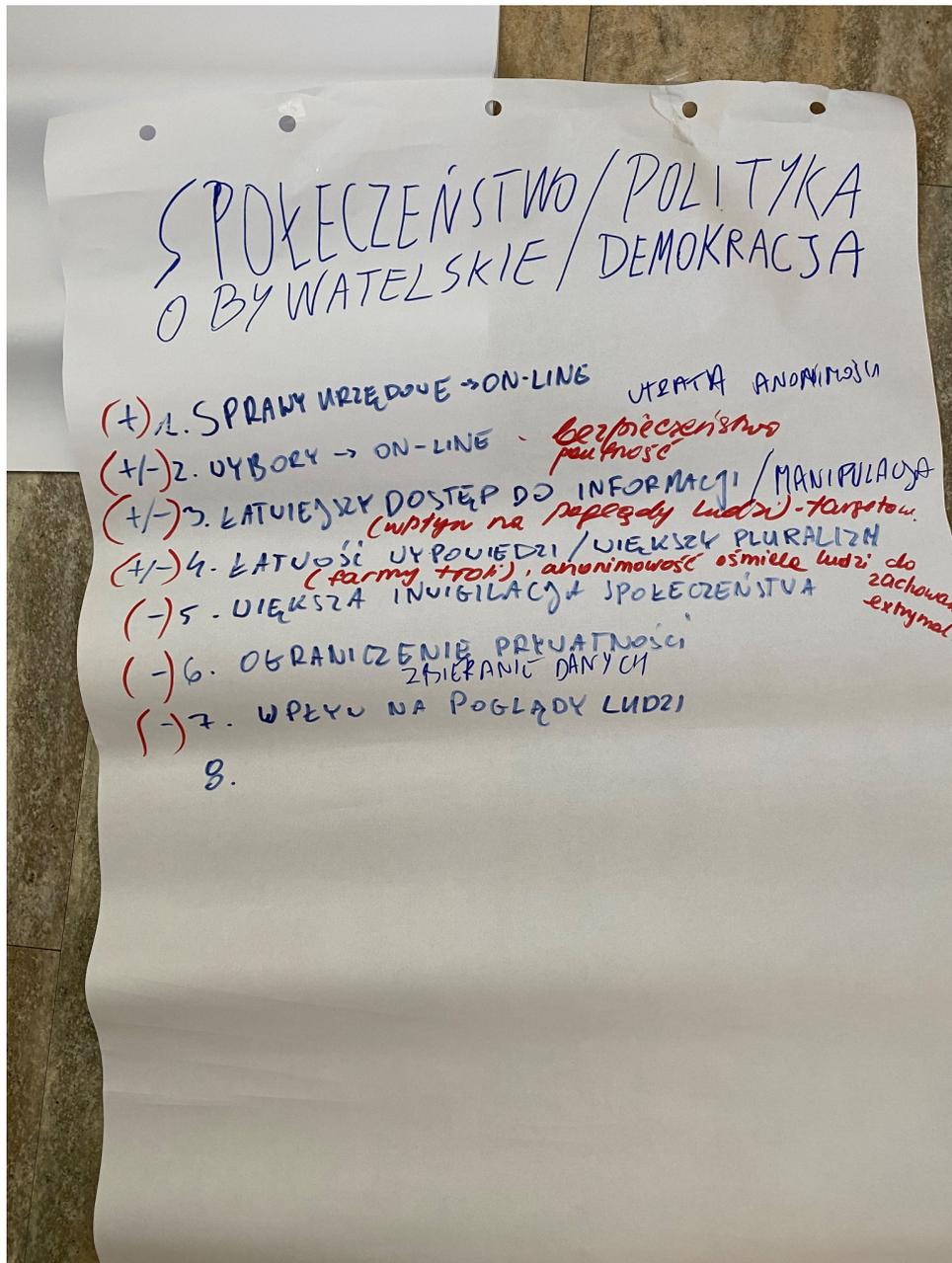


Figure 17 In-group discussion results of the impact of AI on “Civil society, politics, and democracy” in 10 years

5.2.6.6 Materials about AI exercise results

Based on the discussion, here are some key points for **effective** education materials about AI:

- Use a stepped approach - define AI, explain risks, then discuss development and regulation. Emphasise that technology is still evolving.
- Include concrete examples and scenarios, not just technical definitions. Stories and case studies work well.
- Consult older adults when designing materials to ensure intuitiveness and use of accessible language. Avoid complex tech jargon.

- Tailor depth/complexity of information to the target audience - more basic for older adults, nuanced for younger.
- Present materials in engaging formats - videos, podcasts, interactive elements - not just text.
- Leverage trusted experts and thought leaders to communicate, not just celebrities. Match messengers to audiences.
- Recognize no one format will work for all. Multi-modal approach needed to reach diverse groups across generations.

The key is making materials relatable, easy to understand, and digestible for the specific audience. Testing with target users and co-designing with them is advisable.

Funding of the education materials and points raised by the participants in relation to that:

- Transparency about funding sources is important so people can assess potential bias. Materials should disclose who sponsored them.
- Government funding from a relevant ministry or agency could lend credibility, as it's ostensibly for public benefit. But some may distrust government-sponsored messages.
- EU-level financing may seem more neutral and trustworthy to promote common understanding across member states. But those critical of the EU may perceive the content as biased.
- Corporate sponsorship, e.g., by a tech firm, could be viewed sceptically as promoting a particular agenda vs. impartial education.
- A mix of public and private funding from diverse sources could help signal neutrality and balance bias concerns. But full transparency on sponsors is still needed.
- Regardless of financing, materials should emphasise accuracy, objectivity and inclusion of diverse viewpoints. Proactively addressing potential conflicts of interest can build trust.
- If possible, providing ways to verify or fact-check information can give people confidence in assessing credibility themselves. But this may not always be feasible.

In summary, funders should aim for transparency while materials themselves should focus on balanced, non-partisan education accessible to all.

5.2.6.7 Co-creation exercise results

The participants were divided into five groups, each tasked with a creating or development of an idea or a concept of a puzzle, quiz or an element of a game in a convention of an escape room. Each group received a large sheet of paper, colourful markers and around 30 minutes for the task. From the possible themes described above, the groups drew at random their specific assignments. Two groups drew a deserted island as their theme, then the remaining groups drew a submarine, a Jurassic Park and a haunted house each. The participants demonstrated a high degree of creativity by coming up with entire scenarios of a game, putting together questions and puzzles that complete the games. The content of the games also reflects the diverse demographics of the participants.

Figure 18 shows the scenario set in Jurassic Park. A player is trapped with dinosaurs in Jurassic Park. The shortest (and safest) escape route is blocked by a fallen tree. This means that the player must turn their car around and look for another, more perilous way. This means that the player is repeatedly arriving at cross-

roads where needs to answer a question or solve a problem. A correct answer ensures safe passage ahead, an incorrect answer means that the player gets eaten by a tyrannosaurus. The aim is to reach a helicopter that takes the player to safety. The questions and tasks encountered by the player may include:

- AI has a conscience. True or false?
- AI is constantly learning by processing new data. True or false?
- AI is not able to imitate human voices. True or false?
- Assess the truthfulness of some sample information given by AI.
- Can you always trust what AI tells you?
- AI helps to learn. True or false?



Figure 18 Jurassic Park

Figure 19 shows a scenario in which a player is trapped in a submarine. It is a Yellow Submarine to be specific and the game's content focuses on culture (AI in Culture). The player must complete tasks to be able to move to the next room, eventually reaching the exit. Each task is educating the player about a different use and capabilities of AI. Thus, the player passes rooms (or sections) and each of them received a task:

Room / Task 1: Using AI, play the Beatles' song "Yellow Submarine" using the Morse code.

Room / Task 2: Using AI and its functions, generate a cover of any song by Metallica but in the Beatles style.

Room / Task 3: Using AI, write up one or more additional stanzas to the Octopus's Garden song by the Beatles.

Room / Task 4: Using AI, swap John Lennon and Yoko Ono. This could be editing an image of the two persons that assigns gender or race specific attributes to the other person, but also could mean re-rendering a song by swapping their voices.

Room / Task 5: To escape from the submarine, find out what John Lennon and Yoko Ono think about AI.



Figure 19 - Submarine

Figure 20 shows the first version of a desert island. A player is stuck at a deserted island. A rescue plane drops an AI robot that asks the player 3 questions. For each correct answer, the player gets a hint helping them to escape from the island.

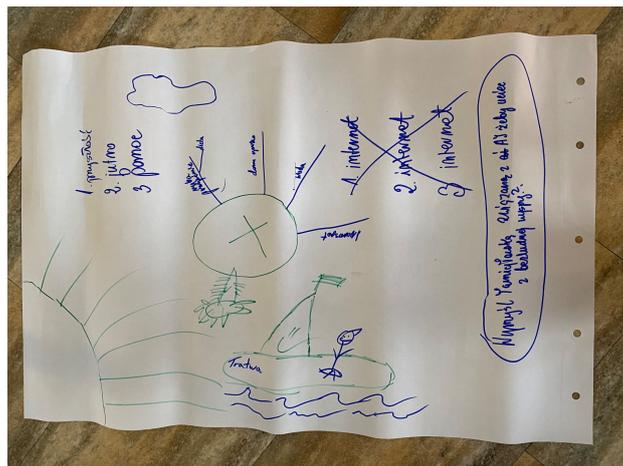


Figure 20 Desert Island - First Version

In Figure 21 a player is stuck at a deserted island. To escape from the island, four tasks must be completed:

1. A neural network on the beach. There is a hidden control panel on the island. The player must find it to activate a virtual network of neurons which shows hidden paths criss-crossing the island. Having discovered the paths, the player can find a mysterious but coded map.
2. The map of algorithms. The mysterious map enables the player to identify key locations on the island. However, it is coded using some algorithm. The player must decode it.
3. The Robot Guardians are defending the access to the next stage. The player must use the control panel to re-program the guardians so that they will the player pass.
4. The labyrinth of algorithms is the final stage of the game. Each path in the labyrinth is a different algorithm. A correct reading/solution of an algorithm shows the way ahead. An incorrect reading/solution of an algorithm leads the player to a dead end. After finding a way out of the labyrinth, the player finds a boat that takes them safely home.

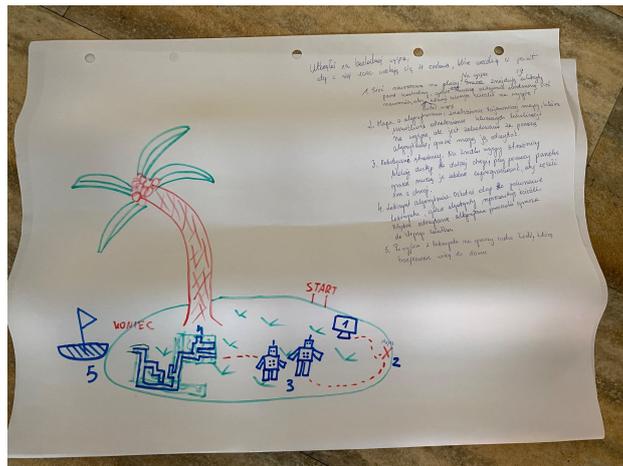


Figure 21 Deserted Island - Second Version

Figure 22 illustrates a game proposed by the participants that is set in a haunting house in an amusement park. A player rides a cart and must complete the course by solving quizzes that directly involve an AI engine. The player must obtain information from the AI that enables them to answer a question. In the first quiz (or station), the AI is disguised as a ghost and a player has a brief chat with it to obtain required hints and tips. At the second station, the puzzle is constructed to make the player aware of the AI’s capability of creating images, including deep fakes, and their quality. The AI is displaying images, and the player must guess which are real and which are AI-generated. At the third station, the player sees various messages and symbols displayed on the walls which must be used to answer a question to pass to the next stage. Finally, at the fourth station the player gets out of the cart and plays hide and seek with the AI. Once the player finds AI, they are allowed to leave the haunted house.

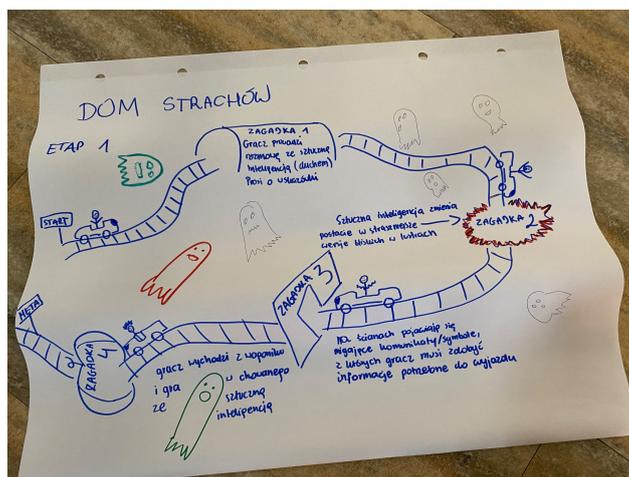


Figure 22 Haunted House

5.2.7 Plans for future interactions

The Conclusion of this document includes future interactions and deliverables stemming from Meeting 1 of UC 3 and the results of Meetings 2 and 3 in Warsaw will be included in future versions of this Deliverable.

5.3 Use Case 4 (Dublin): Social Computing Compass for self-assessment.

5.3.1 Goal

Use Case 4 is designed to interact with software developers working in the field of AI and big data in industry and academia and the general goal is to improve their awareness of the impact of their design tools, systems, and processes on democratic principles and support human flourishing and civic participation. Although the term is not precise enough to be comfortably operationalised in internal KT4D discussions, we use the term ‘ethics’ as a known and accepted shorthand for this debate, so as to allow this stakeholder group to connect their current and past experiences with questions being asked in the project.

The call for a human centric and democratic AI and the need for software developers to follow and implement ethical guidelines and regulations is a highly debated topic. It has become even more pressing in the last year since the release of GPT3 and 4, and their easy-to-use public-facing manifestations such as Chat-GPT, which raised public concern around the ethical repercussions of large language models and AI-powered systems more in general.

However, regardless of media and public attention and regardless of the efforts from governments and regulatory bodies, it is still difficult to “communicate the need for professional understanding of computer ethics”.¹² And this is true for multiple reasons:

1. First, this has to do with **education. Teaching computer ethics at the university level is not a top priority nor a generalised practice.** A survey conducted in 2022 found out that only 2/3 of institutions surveyed teach computer ethics as part of the Computer Science curriculum, which means that a full third does not, at all.¹³ Also, computer ethics is often taught as a standalone subject, and the number of hours dedicated to it is often very limited (10 hours or less per semester).¹⁴
2. Second, this has to do with **implementation. Implementing ethical guidelines is difficult:**
 - a. **because of the nature of those guidelines and principles.** These tools have multiplied in the last few years, making it hard for software developers to identify a shared and universally recognised code of conduct. Also, AI ethical guidelines often promote isolated principles that are perceived as an add-on to the ‘real’ work required of developers. Finally, failing to implement AI ethical principles often lacks real consequences other than public outcry at a significant distance from the work of building software tools and platforms, which reduces the incentive to adopt them.
 - b. **because of the specific environment and workflow in which software developers operate.** A central question that is hard to answer is: at what stage do we implement these guidelines, at the level of the person, the project, or the field? Answering this

¹² Heron, M.J. 2016. Ethics in computer science. In Laplante, P.A. (ed.) *Encyclopedia of computer science and technology*. Second edition. Boca Raton: CRC Press [online], Available at: <https://www.crcpress.com/Encyclopedia-of-Computer-Science-and-Technology-Second-Edition-Print/Laplante/p/book/9781482208191>.

¹³ Stavrakakis, I., Gordon, D., Tierney, B. *et al.* The teaching of computer ethics on computer science and related degree programmes. a European survey. *International Journal of Ethics Education* 7, 101–129 (2022). <https://doi.org/10.1007/s40889-021-00135-1>.

¹⁴ *Ibidem*, p. 116.

means to break down the entire workflow and identify who and at what step is confronted with ethical issues so as to provide them with tools tailored to their specific tasks and roles.

However, it is also true that there is a growing awareness of these issues and appetite among software developers for implementing ethical principles in their work. In 2022 the Institute for Human-Centered AI at Stanford University reported that topics like algorithmic fairness and bias are no longer exclusive to the academic debate, but that “Researchers with industry affiliations contributed 71% more publications [...] at ethics-focused conferences in recent years.”¹⁵

The goal of Use Case 4 is thus to take advantage of this growing interest and leverage the perspectives of software developers to identify shortcomings in existing practices and to tackle concrete, specific issues. In particular we are interested in the extent to which the cultural dimensions of ethical software, be that understood in terms of languages and discourses, national or regional identities, religions, beliefs and practices, values and tolerances, etc., which is often not understood as a part of traditional ethics.

5.3.2 Outline of Use Case 4 Meeting 1

Within this framework, the specific goal for the first meeting of Use Case 4 was to invite a group of software developers and people in positions of overseeing software development (such as industry CTOs and investment firms and agencies) to explore the limits of current approaches to ethical AI development and to envision new, more effective ones that could take into consideration the cultural dimension of these software and systems.

The workshop was entitled "Re-Imagining Ethical AI in Academic and Industrial Software Development," and took place in the Neill Lecture Theatre, Trinity Long Room Hub, Trinity College Dublin, Ireland on 27 October 2023 from 10.00 - 12.30. It focused on fostering a deeper understanding of the ethical challenges, tools, and cultural considerations in software development.

Stakeholder mapping and participants recruitment

The goal was to involve up to 12 software developers working both in industry and academia. We chose this focus because we believe it is important to consider the different approaches to AI ethics across different sectors. Also, since we believe that higher education could and should play a leadership role in advancing ethical self-awareness among software developers (as even those who do not work in this context would likely have been trained there), we designed our recruitment so as to enable an extension of the conversation between computer science programs and the software industry on the topic of AI ethics.

Participant recruitment was effectuated by ADAPT, a Research Centre for AI-Driven Technologies coordinated by Trinity College Dublin and part of the KT4D consortium, and specifically to ADAPT’s Education & Public Engagement Officer, Dr Emma Clarck. Direct invitations were sent to people via email (using email addresses publicly available on institutional websites of companies and universities); we advertised the

¹⁵ Daniel Zhang, Nestor Maslej, Erik Brynjolfsson, John Etchemendy, Terah Lyons, James Manyika, Helen Ngo, Juan Carlos Niebles, Michael Sellitto, Ellie Sakhaee, Yoav Shoham, Jack Clark, and Raymond Perrault, “The AI Index 2022 Annual Report,” AI Index Steering Committee, Stanford Institute for Human-Centered AI, Stanford University, March 2022, p. 11.

workshop on KT4D and Adapt social media accounts and mailing lists, and we distributed flyer at ADAPT's Machine Learning MeetUp and in various places in Dublin attended by developers. It was also distributed via European Commission channels such as the [AI Alliance](#) and the National Contact Point for Ireland for Socio-economic Sciences and Humanities (SSH) within Horizon Europe was informed and promoted it as well within that network.

Designing the event

The Dublin-based team put significant thought into how best to balance the need to frame and inform the discussion, but still allow it to be participant-led. In the spirit of participatory design, we began with the concept of a presentation that would lead the audience through an ever more specific set of conversations, first about their attitudes toward software ethics (and wider attitudes and understanding of software ethics, then turning to tools that have been developed to guide software ethics, and finally looking at the tensions between culture and AI/big data (norms, values, beliefs, discourses, identities, community toes, etc.) to determine whether this aspect is adequately represented in wider AI ethics conversations. This presentation and discussion were to be followed by a 1 hour co-creation session, in which participants would design their own tool for cultural ethics. For this purpose, we were guided by the following primary research questions:

1. At what level to implement AI ethical guidelines: the person, the project, or the field?
2. Are the existing tools and network of support helping you considering the cultural dimensions of your work?
3. How could different kinds of ethical support help you...
 - a. ...translating complex cultural issues into technical recommendations?
 - b. ...capturing the complexity and heterogeneity of identities?
 - c. ...anticipating potential future damages and nefarious use?
 - d. ...thinking differently in complex contexts?

Our original model opening with presentation and discussion on these issues was felt to bring with it a significant risk of passivity and primed responses among the participants, however. We therefore pivoted to the 'conversation station' methodology,¹⁶ supported by prompts and voting stations on large, visually engaging posters (3-4 per station) which is presented in more detail below. The morning session therefore

¹⁶ The 'conversation stations' method, also called 'chat stations' (Bastkowski, M. (2022). The EFL Transition from Primary to Secondary School. In: Summer, T., Böttger, H. (eds): English in Primary Education: Concepts, Research, Practice. University of Bamberg Pres. 91-115: 108. <http://dx.doi.org/10.20378/irb-58793>) is a kind of activity that aims at fostering active conversation on a set of topics without imposing any particular structure or talking points. This technique is mostly used to enhance students' language and conversational abilities (Bond, M. A., & Wasik, B. A. (2009). Conversation stations: Promoting language development in young children. *Early Childhood Education Journal*, 36(6), 467–473. <https://doi.org/10.1007/s10643-009-0310-7>). The use of visual props – often posters – at each station and the need to move around and change topics and activities ensure better engagement (Gonzales, J. (2013). Students Sitting Around Too Much? Try Chat Stations. *Cult of Pedagogy*. <https://www.cultofpedagogy.com/chat-stations/>). A similar technique to 'conversation stations' from which we borrowed is 'gallery walk', often used with more mature students and for professional development (McCafferty, A. S., Beaudry, J. (2017). The Gallery Walk: Educators Step up to Build Assessment Literacy. *Learning Professional*, 38(6): 48-53). According to McCafferty and Beaudry, a gallery walk is "a discussion technique that gets learners out of their seats and invites them to become active participants", which is what we created for the UC4 workshop. The gallery "consists of images, graphics, and text on posters of various sizes with embedded hot links, QR codes, and stations for hands-on activities. Participants move through the gallery to interact with, reflect on, and discuss ideas with peers" (McCafferty and Beaudry: 49). This is exactly how we designed our stations in terms of material and flexibility.

had to begin with the physical setup of the venue, including the provision for attendees to receive name badges and sign the informed consent forms while helping themselves to coffee, tea, and light refreshment.

The workshop began at 10:00 with a very short introductory presentation on the KT4D project followed by some information about our photography policy, giving attendees the option to opt-out if they did not want to appear in photographs.

The first hour of the workshop (10:15-11:15) utilised the ‘conversation stations’ to open opportunities for discussing questions around **agency** (who is in charge of implementing these ethical principles and to what extent this is effective), around **support** (what people working in different roles think that is missing at the moment and what tools they would like to see), and around **translation** of cultural issues into technical recommendations.

During the second half of the workshop (11:30-12:30) participants were involved in a co-creation activity during which they engaged with hypothetical scenarios to develop a prototype for a tool to assess the ethical impact of their work. This gave them the opportunity to consider their own ethical practices and the state of their field in very concrete terms and through co-creation.

The workshop concluded with summary remarks, expressions of gratitude to the attendees, and information on what to expect next. Participants were invited to stay for a light lunch during which they would have the chance to continue the discussions with the project team members. This was intended to foster an informal atmosphere for networking and knowledge sharing, as well as for further relationship building on behalf of the project.

5.3.3 Methodology and Activities of Use Case 4 Meeting 1

5.3.3.1 Activities

As stated above, the event’s first activity (10:15-11:00) revolved around three "Conversation Stations". Participants were divided into two groups of 3 and 4, although initially we planned for 3 groups of 4-5 people (see section 5.3.4). Each group had 15 minutes to engage with each one of the three stations. At each station a facilitator explained the activity, answered any questions or concerns, and facilitated the conversation without steering it or nudging people. The facilitators took notes while a timekeeper ensured that groups rotate stations every 15 minutes.

A number of posters for each station specifically created for the workshop functioned as talking points. In each case, there was one main poster (printed in A1) and several additional ones (printed in A3). Each station’s facilitator is tasked with ensuring a productive and insightful exchange of ideas at each station and with taking notes. These posters were designed to be visually interesting and intellectually stimulating, including popular memes, excerpts from news articles, findings from research, links to ethical research support tools, and other relevant content related to the themes of the stations. They also integrated points at which participants were asked to ‘vote’ on particular issues or questions (using coloured adhesive dots provided to them) or place post-its on the posters to extend or respond to issues stated there. A selection

of these posters appear integrated into the text below, the full set of posters can be accessed in the KT4D Zenodo repository at the following links given is the text describing each station below.

The three stations were organised as follows:

Station 1: The challenges of ethical software development.

Overarching question: What are the professional / social / ethical challenges with ethical software development?

Activity: Using post-it notes, participants outlined the challenges they saw with ethical software development. There were post-its available for participants to add comments, if they wanted to.

Material: Figure 23 depicts the main conversation poster for Station 1. The remaining posters for Station 1 can be viewed on [Zenodo's KT4D Community](#).



Figure 23 Use Case 4 Meeting 1 Main Poster

Station 2: Tools for ethical software development.

Overarching question: "What kinds of support do you rely on to help you maintain the ethical dimension in your work?"

Activity: Participants received 5 green and 5 red sticky dots. They were asked to vote on which ethical tools they found most (green) / least (red) useful. There were also post-it notes available for participants to use to add comments, if they wanted to.

Material: Figure 24 depicts the main conversation poster for Station 1. The remaining posters for Station 1 can be viewed on [Zenodo's KT4D Community](#).



Figure 24 Use Case 4 Meeting 1 Station 2 Main Poster

Station 3: Cultural issues / challenges posed by ethical software development.

Overarching question: "Do you feel prepared to consider the identities and cultures of people who may use your tools?"

Activity: Participants were given sticky dots of different colours, and asked to 1) vote on the issue of whether they felt they received sufficient support; 2) decide if the cultural issues presented in the secondary posters were not relevant, relevant but difficult to implement, or already implemented. There were post-its available as well for participants to add comments, if they want to.

Material: Figure 25 depicts the main conversation poster for Station 1. The remaining posters for Station 1 can be viewed on [Zenodo's KT4D Community](#).



Figure 25 Use Case 4 Meeting 1 Station 3 Main Poster

The workshop continued with feedback from each of the stations. Each facilitator presented a summary of the discussions and outcomes (5 minutes per station), providing valuable insights into the challenges and opportunities discussed by participants.

Following a short comfort break, the participants were divided into new groups for the co-creation activity. In this session, participants were tasked with re-imagining ethical software design. Participants were given 45 minutes to brainstorm and create a paper prototype for a tool that could help them tackle the cultural issues in AI software development previously discussed in the three stations.

Each group then presented their prototype (5 minutes per group), with the most innovative idea receiving a prize.

5.3.3.2 Methodology

For both activities our goal was to gather participants' insights in the field of interest to us, without imposing our own ideas and beliefs so that we can get a picture of developers' everyday challenges and needs and so that we do not introduce confirmation biases.

However, for the second activity to be productive, we needed participants to consider in a critical and informed way a set of issues and challenges that they then need to tackle in practical terms through the creation of the prototype for an AI ethics tool.

The decision to organise the first activities around the three 'conversation stations' stemmed directly from this. By avoiding a traditional frontal presentation and plenary discussion, we were able to give the participants the freedom to shape the conversation. However, having a series of posters and voting/ranking activities allowed us to gently inform and nudge them towards the main issues and challenges that are the core of UC4 as well as at the centre of Activity 2.

This decision to proceed in this manner was influenced by the report from IRMIR about their pilot workshop for UC3. Indeed, they found that the most interesting discussions and input came from less structured activities, in which their participants were freer to stir the conversation. This is testament to the success of the collaborative practice adopted by the KT4D consortium.

For the second activity we design the co-creation exercise to allow for:

- participants with different background and working in different role to come together and consider AI ethics in software development from a multiplicity of point of view;
- participants to challenge and improve existing ethics tools in a practical and hands-on way;
- participants to think creatively and imaginatively, using physical objects (paper, post-its, posters, glue, etc.) to create their paper prototypes;
- participants to make an instant and clear connection between cultural issues (discussed in Activity 1) and practical guidelines and tools (developed in Activity 2).

5.3.4 Results Use Case 4 Meeting 1

During the recruitment phase, 19 people registered for the event, however only 7 actually took part in the workshop. This was initially disappointing, but it actually turned out to be advantageous as the smaller group allowed for better conversations at the three stations and gave us more time for the plenary discussion.

The participants were very engaged and took part in the activities with great enthusiasm and sense of collaboration. The different age groups and backgrounds represented made for a diverse group that brought to the table different perspectives and points of view.

Most of the participants worked in the academic sector, with one working for a government agency and one working in industry. This meant that all participants were aware of the issues of ethics in AI development and sensitive to the issues discussed. While this made for good engagement, it meant that we did not have the chance to hear from developers who might be more reluctant – if not averse – to adopting ethical tools and guidelines. Reaching out and speaking to this group will definitely be a major goal for the ethical tool that we are building as an outcome of this first workshop.

Although the group participating in Use Case 4 Meeting 1 was smaller than expected the level of interest and engagement with the project as a whole and the issues presented was very high, and very encouraging for the next two meetings of the cohort in Dublin. With this cohort, the use of poster-based, interest-led investigation was very effective for reaching software designers and development managers (though it would not necessarily be equally effective for every group). The input received was very helpful in framing an advanced understanding of where the real gaps in ethical AI in the context of democratic processes and civic participation may lie and will be very instructive in framing the development of the Cultural AI compass.

5.3.4.1 Activity 1: The Conversation stations

The 7 participants were divided into two groups respectively of 3 and 4 people and each group engaged as planned with one conversation station for 15 minutes before rotating to the next one. Some of the key points of discussion from the stations can be summarised as follows:

Conversation Station 1: Challenges of Ethical Software Development (Facilitator: Emma Clarke)

Summary: Over the course of the conversations at Station 1, participants agreed that implementing ethical software development is very challenging, but absolutely *should* be implemented. At the station, there was some discussion about **when** ethical frameworks / guidelines should be implemented. There was a suggestion that these should be in place from the inception of a project, and throughout development, but participants acknowledged that this is challenging in itself. It was also suggested that academia “**may**” be more ethical than organisations outside of the academic setting.

Some specific challenges that were raised by participants during the discussion:

- Many developers just want to code and sometimes they just want to build cool things that will test their development skills;

- The issues concerning AI are very new, and this is a fast-moving space;
- There is a perception that ethics guidelines will stifle innovation;
- The features that users want can often be unethical;
- How can we make people want to learn ethics?

Challenges for companies include:

- The emphasis is often on delivering a product (time and cost were mentioned here as barriers to ethical dev)
- Company culture differs across organisations, thus companies can attract different types of workers
- A positive example arose around the companies that take ethics as their unique selling point

Enforcement

The EU AI Act was discussed, but there are considerations including:

- The AI Act is a good step, but figuring out how to implement it on a day-to-day basis will be challenging
- What will the impact be on businesses
- Unless it is enforced, it won't be treated as a priority

Conversation Station 2: Tools for ethical software development (Facilitator: Eleonora Lima)

Summary: Participants discussed the different types of tools but mostly focussed on 'Checklists and assessment tools', 'Games', and to a lesser extent on 'Colleagues and network' (nobody discussed 'Training and events'). All the four categories received green and red dots (positive and negative reviews) and, in some cases, participants decided to combine two stickers to express their ambivalent opinion.

Points raised about *Checklists and assessment tools*:

- The general opinion about the usefulness of these tools was negative;
- they are often too general, as they pose yes or no questions and ignore complexities (e.g. GDPR checklist);
- are rarely targeted towards AI specifically and, even when they are, they are not reliable as not many people have tested them for a long enough time;
- often developed by people in positions of power and privilege and this leads to blind spots and biases, even when the intentions are good.

Points raised about *Games*: The general opinion about the usefulness of these tools was that games could be useful, but not the ones that they have used/know about.

- Existing games to detect biases are sometimes biased themselves;
- Good models for ethical games come from outside the field of AI ethics (e.g. role-playing games like Dungeons & Dragons allowing for collaboration)

- Need for non-competitive games that recognise people’s different roles, power, and abilities (suggestion of taking inspiration from Golf [handicaps](#));
- Games are good to introduce ethics in computer development without sounding preachy or judgemental.

Participants did not know any of the games and tools presented in the posters and they were very interested. This clearly demonstrates that there is a problem of communication: many useful tools out there, but they do not reach their intended audience.

Points raised about *Colleagues and networks*: The general opinion about reaching out to colleagues and network was that while this is a somehow useful and widespread practice, it can also lead to complacency (if my colleague does it, I can do it too) and power imbalances.

Conversation Station 3: Cultural issues / challenges posed by ethical software development (Facilitator: Jennifer Edmond)

Summary: Most of the participants felt that this was a central set of concerns for them, though some did not feel they had adequate tools to do so. The greatest hesitation was not around the importance of culture and identity, but the importance of staying humble in the face of these aspects of who their users might be, and not overestimating the breadth of their perspective. You can’t ever say yes to the question of feeling able to deal with intersectionality, as this would imply that the problem was sorted, and there were no possible perspectives one was not able to adequately appreciate and address. “Saying yes,” said one participant “feels like I know it all.”

Some participants offered examples of approaches they felt helped them engage with cultural complexity, including:

- non-algorithmic XAI
- universal design for learning,
- accessibility guidelines (though these need continuous reconsideration and updating),
- the ‘never complete’ status of design processes,
- red teaming.

Some specific challenges that were raised:

- The challenge of integrating identity is not a new one, as it is related to age-old issues of power structures, in particular corporate power structures driven by commercial imperatives: how can we learn from history in this respect?
- Big tech operates on too many assumptions, as emblematised by Apple’s famous statements about the consumer not knowing what they want.
- The myth of a truly meritocratic society was seen as a social problem that became incorporated into technology platforms.

- More contemporary barriers had to do with the limits of technological systems themselves, such as the nature and selection of (biased) AI training data unable to recognise underrepresented identities.
- Of all the problems proposed, the issues of the local versus global were felt to be the biggest challenge, and therefore raised issues that tended to fall by the wayside. How could the moment of recognising differences and biases be instigated?

5.3.4.2 Activity 2: Co-creation activity

The 7 participants for two groups, different from the one in Activity one, of respectively 3 and 4 people. They each created a prototype (described below) for an AI ethics tool by using the material provided (post-its, sheets of paper, markers, etc.).

The plan was to give a prize in the form of a gift token to each of the members of the winning team. In our initial plan, all participants were supposed to vote for the best tool. However, in consideration of the participants' high level of commitment and enthusiasm, and the presence of only 2 teams due to the lower than expected participation, we decided to give a prize to every participant and not to have them vote.

Group 1: CULTURaLi

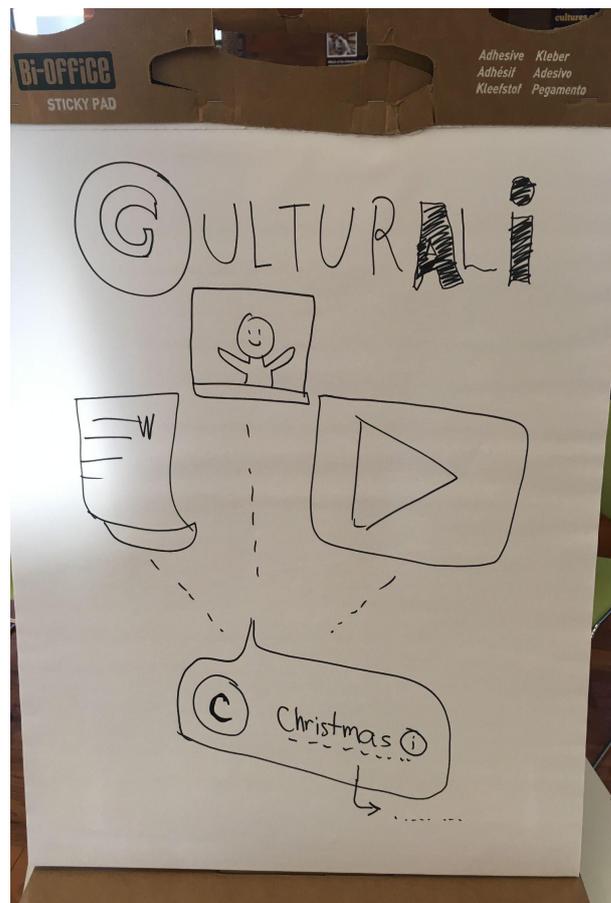


Figure 26 Co-Creation Exercise CULTURaLi Tool

- The group envisaged a tool called CULTURaLi. It would be similar to Microsoft Accessibility Tool for Vision, which is made for visually impaired people.
- The tool was not meant to change anything, but instead to highlight potentially problematic and discriminatory aspects and elements in LLMs (similarly to what Grammarly does).
- The primary goal of the tool was to target cultural differences (e.g., scientist initially identified as He in a hypothetical - why was the scientist automatically a He instead of She?)
- The tool also pointed out what can potentially be culturally specific and how to change it depending on the culture / context.
- One example is the use of cultural and idiomatic expressions that are not understandable by everyone and might cause incomprehension and barriers. For instance, the name 'Joe Bloggs' broadly refers to an unidentified male, which is used in Ireland but not in other countries. In this case the tool will suggest different names for different cultures. This small change will help address cultural differences, making the content more accessible to more people.
- The tool would be open source (added onto Microsoft Office) and multi modal.

Group 2: ZombAI

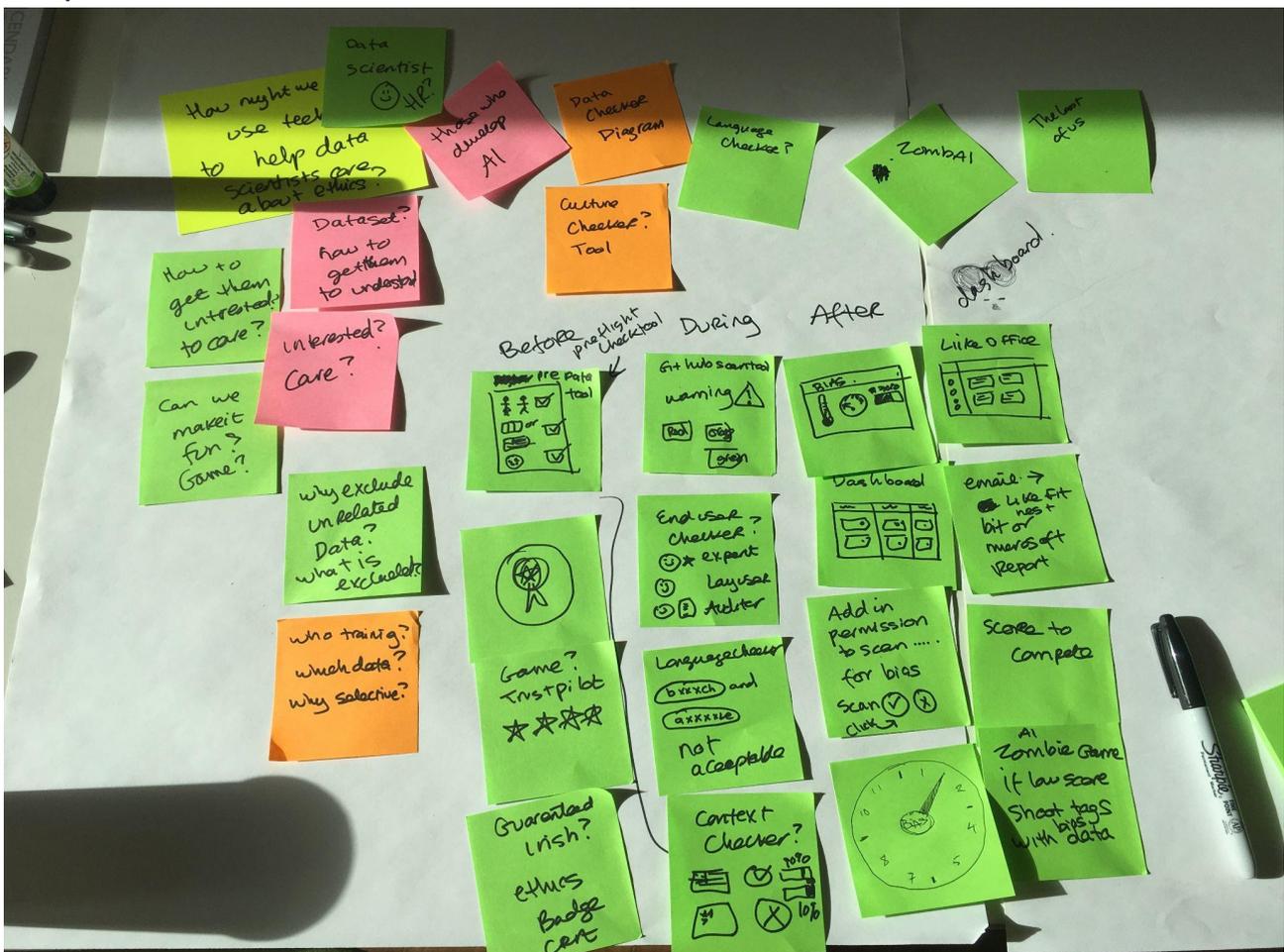


Figure 27 Co-creation Exercise ZombAI

- The initial question guiding this prototype is: from data scientists’ point of view, how can we use technology to make computer scientists care about ethics?
- The group imagined developing a suite of tools to be used at different stages of the software development:
 - Pre-check, before processing data
 - During the development of any software or system
 - At the end, as a final check

At each stage developers using this tool successfully would get a certificate showing that necessary ethical standards have been met. The label could be something like “Guaranteed Global”, meaning that the standards are not only tested on and respectful of Western cultures.

Before Development: the ‘before tool’ would tackle any form of bias in the dataset (e.g. gender bias, if the data is US centric or EU-centric, etc.).

During Development

- The tool would deliver warnings about potential ethical issues during the software development stage. This would use a traffic light system that gives feedback about the compliance of a software with ethical principles and guidelines.
- The tool would have a ‘language checker’ that can be used, for instance, to detect sarcasm and distinguish it from harmful language.
- The tool would identify bias and issues in the context of a specific cultural group (e.g. something is culturally acceptable in the USA, but not China).

After Development

At this stage, each team of developers would play a game against each other. The team that created the software that complied the best with the ethical standards and tool into proper consideration of cultural issues (assessed during the two previous stages), would receive the highest score and win. This would mean that the winning team would have created an **ethical AI** while the other team would have created a **zombie AI**, deprived of culture and human value.

5.3.5 Plans for future interactions

The Conclusion of this document includes future interactions and deliverables stemming from Meeting 1 of UC 4 and the results of Meetings 2 and 3 in Dublin will be included in future versions of this Deliverable.

6 Conclusion

As illustrated in the above sections, the first interaction of each the Use Cases was based upon the model of Participatory Design, in which participants were guided in the course of a workshop to consider the goals of the project and previous work related to their potential concerns about AI, big data and democracy, and

prompted via a series of co-creation exercises to offer structured and unstructured input to the project KERs (Key Exploitable Results).

Indeed, through the four Use Cases we successfully reached out to our key target groups and engaged substantively with them so as to embed their experiences into the project’s results, specifically KER 2, KER 3, and KER 4, on which the KT4D team is currently working.

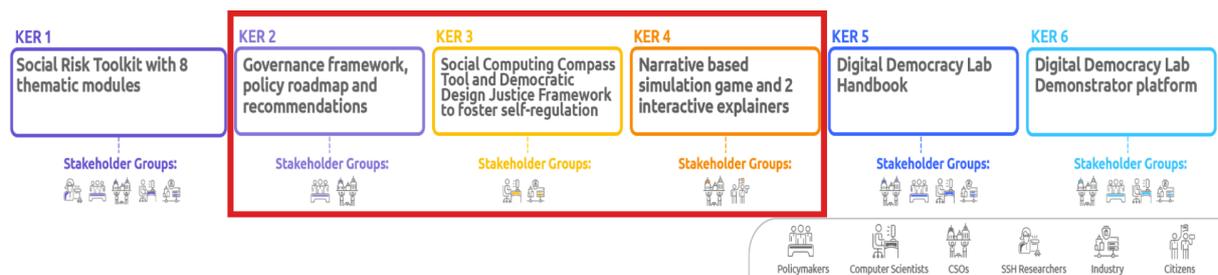


Figure 28 KT4D KERs (highlight KERs 2, 3 & 4)

KER 2, Governance Framework for Democratic Use of Knowledge Technologies will stem from the work and insights of the **Meeting 1 for Use Case 1**. The goal is to provide an easy-to-use tool for policymakers and policy-facing CSOs, guiding assessment of democratic risks when considering regulation within the context of broader European values systems. This KER will propose governance frameworks and other values-based solutions to protect citizens from the possible negative impact of these technologies on fundamental rights and democracy, taking steps towards the protection of citizens, upholding European values, and ensuring public trust in AI and the processing of big data. The governance framework will consider both the potential negative impacts of AI solutions on fundamental rights, such as rise of algorithmic bias and disinformation as well as the positive potential of big data to reinforce democratic governance. It will draw from the first workshops of Use Case 1 and from Delphi study with policymakers and will be developed by DEMOS with the support of ICTLC, tasked with drafting concrete and implementable guidelines for the lawful and ethical application of AI and big data in the context of technology development.

KER 3, Social Computing Compass Researcher Self-Assessment Tool for Design Justice will stem from the work and insights of the **Meeting 1 for Use Case 4**. The goal is to expand on ethical standards and software design assessment, and build upon them with a new tool more attuned to the understanding of cultural sensitivities, relational and holistic identity paradigms, and awareness of individual biases. Software platforms are enactments not only of technology, but of software developers, their perceptions, values and biases. Therefore, the Social Computing Compass will offer tools to foster bottom-up self-regulation and self-assessment by software developers. This work will produce a democratic design justice framework to ensure that whoever designs, develops, and uses AI can do so in a just manner inspired by democratic principles and public good. The Social Computing Compass will be developed by TCD, who has undertaken a review of existing tools to guide software developers toward human-centred design for meeting 1 of UC4, and by Demos, who will contribute expertise to participatory design.

KER 4, Narrative based simulation game and two interactive explainers will stem from the work and insights of the **Meeting 1 for Use Case 2 and Use Case 3**. The goal is to create interactive educational/infotainment materials targeting citizens and citizen-facing CSOs to realise the concept of culturally sensitive critical digital literacy. This KER will produce a framework as well as pedagogical materials (namely a narrative based simulation game and 2 interactive explainers) with which to operationalise it. These tools will be designed to protect citizens from the possible negative impacts of advanced knowledge technologies on fundamental rights and democracy and they are being designed by taking into account the insights (both discussion and prototypes) of the first meeting for UC2 and UC3. They will empower users, and thereby leverage enhanced civic participation and democracy. HYB will develop and deploy the technical framework for the games, while CIB and IRMIR will support the development of KER 4 through the experience gained during their UCs first interaction. The Development process will include: gamification dynamics; immediate and individualised feedback formulations; game simulations, characters, graphic designs and sound effects.

These four KERs are now in active development, with the input from the Use Cases central to the project’s thinking about how to optimally shape these important outputs.

The second interaction of the four KT4D Use Cases will occur in the project’s second year (ca. M 20) and be organised around a Digital Democracy Lab (KR5 and 6), a one-day experience of working within a transparently constructed, open and explainable (for each of its components will be explained and engaged with critically by the users in the lab) platform for civic interaction.



Figure 29 KT4D KERs (Highlight KERs 5 & 6)

DemSoc will lead the development, ensuring that all modules and technical solutions are implemented in ways that allow for citizens to be engaged in practical ways, while mitigating ethical risks inherent in the use of advanced knowledge technologies to foster citizen engagement and democratic debate. HYB will be involved to continuously improve and align the components developed in Task 7.2 to the diverse user needs related to the 4 Use Cases. IRMIR, TCD, CIB and Demos will participate in the refinement of the lab components to meet the specific needs (in terms of language and culture, but also professional or knowledge based) of the use cases they lead.

At the moment, the four Use Cases are actively sharing insights and information and collaborating with each other and with the other partners of the consortium so as to successfully design and deliver four parallel and yet different workshops for the second meeting of each Use Case, like the ones described in this document. Given the positive responses and active participation for the first round of interactions, we are optimistic about the next steps in the UCs, while also reflecting on these past experiences so as to grow and improve.

The feedback loop on which the four UCs are predicated demands that, in order to ensure the best results for our project, we must offer our participants the highest quality of engagement, so as to receive their feedback and insights in exchange. It is a difficult challenge, but one that we are prepared and excited to face.

Appendix I. Consent Form Template

Tailored consent forms for each Use Case are available upon request.



KT4D
Knowledge Technologies
for Democracy

KT4D CONSENT FORM

LEAD RESEARCHERS: Jennifer Edmond, Trinity College Dublin

I volunteer to participate in this research conducted by the KT4D consortium, coordinated by Trinity College Dublin. The KT4D consortium consists of the following 12 organisations:

- 1 Trinity College Dublin (Coordinator, TCD) Ireland
- 2 Trust-IT (Trust) Italy
- 3 Commpa SRL (CmP) Italy
- 4 Institute of Urban and Regional Development (IRMiR) Poland
- 5 Demos Research Institute (DEMOS) Finland
- 6 Fundación Cibervoluntarios (CIB) Spain
- 7 University of Warwick (UW) UK
- 8 Strane Innovation (Strane) France
- 9 Beyond the Horizon ISSG (BtH) Belgium
- 10 Hybridcore BV (HYB) Belgium
- 11 ICT Legal Consulting (ICTLC) Italy
- 12 Democratic Society (DemSoc) Belgium

The project is funded by the European Commission under the Horizon Europe funding programme, grant agreement number 101094302. The project began on 1st February 2023 and will come to an end on 31st January 2026.

By signing this form, I agree to take part in the KT4D research. The nature of the research, my involvement in it and my rights regarding my participation are explained in the Information Sheet accompanying this form.

DECLARATION:

- I have read, or had read to me, a document providing information about this research and this consent form. I have had the opportunity to ask questions and all my questions have been answered to my satisfaction and understand the description of the research that is being provided to me.
- I agree that my data is used for scientific purposes and I have no objection that my data is published in scientific publications in a way that does not reveal my identity.
- I understand that if I make illicit activities known, these will be reported to appropriate authorities.

- I consent to being audio and video recorded, having my personal data processed, and having notes of my activities taken for:
 - A. Ongoing research to improve KT4D user requirements, revise design, develop KT4D technologies and KT4D policy management;
 - B. Dissemination activities (e.g. articles for peer-reviewed journals, presentations at conferences);
 - C. Promotion of KT4D in general.
- I understand that I may stop electronic recordings at any time, and that I may at any time, even subsequent to my participation, have such recordings destroyed (except in situations such as above).
- I understand that, subject to the constraints above, no recordings will be replayed in any public forum or made available to any audience other than the current researchers/research team.
- I freely and voluntarily agree to be part of this research study, though without prejudice to my legal and ethical rights.
- I understand that I may refuse to answer any question and that I may withdraw at any time without penalty.
- I understand that my opinions and views arising from my participation are fully anonymised and that no personal details about me will be recorded other than my name and email which will be used solely for administration purposes that strictly follow the GDPR.
- I am aware of the potential risks and benefits of this research study.
- I consent to the processing of my data. My personal data can be gathered to be used, stored and shared in the ways described on the accompanying Information Sheet. The personal data collected will be processed following the GDPR and pseudonymised/anonymised to the greatest extent possible.
- I understand that the KT4D consortium intends to retain my personal details (e.g. name, contact details) for a period of up to 12 months following the completion of the project. Information from research activities (e.g. observation notes and information from feedback forms) will be permanently and irrevocably deleted after a maximum of 5 years after the end of the project.
- *<If the research involves viewing materials via a computer monitor>* I understand that if I or anyone in my family has a history of epilepsy then I am proceeding at my own risk.
- I would like to receive updates on the progress and findings of the project.
- I have received a copy of this agreement.

PARTICIPANT'S NAME:

PARTICIPANT'S SIGNATURE:

Date:

Statement of investigator’s responsibility: I have explained the nature and purpose of this research study, the procedures to be undertaken and any risks that may be involved. I have offered to answer any questions and fully answered such questions. I believe that the participant understands my explanation and has freely given informed consent.

RESEARCHERS CONTACT DETAILS:

INVESTIGATOR’S SIGNATURE:

Date:

Appendix II. Information Sheet Template

Tailored information sheets for each Use Case are available upon request.



INFORMATION SHEET

You have been invited to take part in the European Commission funded KT4D project coordinated by Jennifer Edmond from Trinity College Dublin. The KT4D consortium will be conducting this research.

Your participation is voluntary, and you are free to withdraw your participation at any time without suffering any negative consequences. Before you decide whether or not to take part, you should understand why the research is being done and what it will involve. Please take time to read the following information carefully and feel free to ask the researcher any questions you may have. If you have questions after participating in the research session, please contact [researcher name], whose information can be found at the end of this information sheet.

The project

The purposes of this project are: (1) to understand the concerns of citizens like you regarding how new information technologies are affecting democracy (for example deepfakes, filter bubbles, algorithms, surveillance... etc.) and (2) to build together tools, games and educational materials that can help us either counter the effects of these technologies or make them help for democracy and participation. The project began on the 1st February 2023 and will come to an end on 31st January 2026.

The KT4D consortium consists of 12 organisations:

Partner	Short Name	Country
Trinity College Dublin	TCD	Ireland
Trust-IT	Trust	Italy
Institute of Urban and Regional (IRMiR) Development	IRMiR	Poland
Commpla SRL	CmP	Italy
Demos Research Institute	DEMOS	Finland



KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

Fundacion Cibervoluntarios	CIB	Spain
University of Warwick	UW	UK
Strane Innovation	Strane	France
Beyond the Horizon ISSG	BtH	Belgium
Hybridcore BV	HYB	Belgium
ICT Legal Consulting	ICTLC	Italy
Democratic Society	DemSoc	Belgium

As part of the study we would like to achieve

This study aims to gather opinions about technology and democracy. The KT4D solution will be created with the help of citizens like you with the intent to be an accessible tool and set of materials to help European citizens with democracy from a technological point of view.

Your participation is very important since it will allow us to adapt the solution to your real online protection needs and to adjust its design to improve your experience while using it.

If you agree to take part in the research, any personal information (e.g., name, contact details) that will be collected from you is for our internal processing and administrative purposes only, and to enable us to contact you if we require further information. Your personal information will be kept for a maximum period of 24 months following the end of the research project. Information from research activities (e.g. observation notes and information from feedback forms) will be permanently and irrevocably deleted after a maximum of 5 years after the end of the project. Unless you prefer otherwise, we will not publish any information in reports or communications materials that would enable you to be directly or indirectly identified.

What will I be asked to do?

If you attend an event during which a workshop for the creation of the KT4D educational materials takes place, we will ask you to give us your opinions and participate in joint dynamics of creation, roundtables and interviews.

If you participate in the testing of the materials, you will be asked to interact directly with the KT4D solution following the trainer’s instruction. After this instruction you will be requested to complete a questionnaire in order to provide us with your impression upon the solution.

Recording

We will video/audio record certain parts of the sessions that require posterior analysis.

Where will the research take place?

The research will take place at [insert location]



KT4D has received funding from the EU’s Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

What will you use my participation for?

Your participation will be used to provide input for us to:

- Understand the concerns of European citizens regarding democracy and new technologies.
- Help us design and create new educational materials and tools for the project.

Additionally, the information that you provide may be used to write articles for peer-reviewed journals and magazines, for presentations at conferences and workshops, or to promote the KT4D project. Unless indicated otherwise, for all information that could either directly or indirectly identify you will be anonymized.

What are the potential risks of participating in research?

We do not envisage any risks to be caused by the participation.

Are there any costs?

There are no costs for participating.

Will I be paid?

We cannot pay you for participating in this research.

Storage of data

[Partner Name] takes the protection of the privacy and protection of personal data of the research participants very seriously. Therefore, the personal information of the participants is kept securely and treated with the utmost care, fully complying with the provisions of Regulation (EU) 2016/679 of the European Parliament and of the Council, of April 27, 2016, regarding to the protection of natural persons with regard to the processing of personal data and the free circulation of these data (hereinafter, "GDPR") and in Organic Law 3/2018, of December 5, on the Protection of Data and guarantee of digital rights (hereinafter "LOPD-GDD").

In this sense, the legal bases used by [partner name] for the processing of user data are:

- I. The execution of a contractual relationship.
- II. The legitimate interest.
- III. The consent of the Users to the processing of their personal data for the purposes specified in each case.

Observation notes and information from comment forms will be shared only with those consortium members who require access for their work. This information will



KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

be retained for the life of the project. After the end of the research, it will be permanently and irrevocably deleted after a maximum of 5 years.

Data Protection Officer (DPO)

The DPO of our institution is: _____

Please contact him/her for further questions.

Data Administrator

The Data Administrator of our institution is: _____

Please contact him/her for further questions.

Lawfulness of Data Processing

The legal basis of the processing of your personal data is based on your consent, the performance of a task carried out in the public interest, and the legitimate interests of the data controller (GDPR Art. 6a, e, f).

Your rights and confidentiality

We will anonymise the data you provide us to the extent possible. We will only collect and process data that is strictly necessary for running the research, for our internal processing, administrative purposes, and to enable us to contact you if we require further information. The record of your participation will be kept in a file separate from the research data. This data will not be shared with or disclosed to anyone outside the research team.

All information we collect about you will be kept strictly confidential unless we are required to share your information with the European Commission as part of our legal obligations. However, the researcher has a duty of care to report to the relevant authorities possible harm/danger to the participant or others. If this was the case, we would inform you of any decisions that might have an impact on your rights, including data protection rights. All data, including audio files, will be stored on password protected computers, hosted by the ADAPT centre, TCD, using secure servers within the EU. Every effort will be taken to protect your identity. You will not be identified in any report or publication of this study or its results. You can review any recording/notes that concern you should you choose to do so.

You have the right to access, update, correct and erase all personal data. As to the qualitative information that you provide us with, you have a right to withdraw the same up to the point of publishing the information in the relevant deliverable. Your researcher will inform you as to the planned publication date.

You have a right to lodge a complaint, to do so please contact the researcher or project coordinator (details below); they will in turn pass your complaint onto an independent panel.

Right to withdraw



KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.

You may withdraw your participation from this project at any time without giving a reason. You may walk away at any time. Should you wish to stop participating in the research, you may tell the researcher at any time that you would like to stop. You may be asked why you have decided to withdraw, but you are under no obligation to give a reason. If you would like the researcher to delete your personal data collected in the context of the researcher, please inform the researcher or contact them at [insert researcher name].

Sensitive information

Special category or sensitive personal data relates to specific categories of data which are defined as data relating to a person's racial origin; political opinions or religious or other beliefs; physical or mental health; sexual life; criminal convictions or the alleged commission of an offence; trade union membership.

Research with participants in non-EU countries

If you are from outside of the EU we ask you to note that the personal data will be transferred to and stored in the EU/EEA. All transfers of personal data outside of the EEA will be subject to suitable safeguards such as the Standard Contractual Clauses issued by the European Commission. For more information on data transfers, please contact the DPO at [insert email address]

Keeping in touch with the project

As KT4D is a research and innovation action, it is essential to share the high-quality results of the project with stakeholders who are likely to benefit from it. You can choose to be kept informed about the project's progress, and will thus be put on a mailing list, however this is not mandatory.

More information?

For more information on the project, please contact:

Project Coordinator

Name: Jennifer Edmond
Organisation: Trinity College Dublin
E-mail: edmondj@tcd.ie

Researcher

Name:
Organisation:
E-mail:

To file a complaint with the Data Protection Officer:

Name:
E-mail:



KT4D has received funding from the EU's Horizon Europe research and innovation programme under Grant Agreement no. 101094302.