OUKA
The University of Osaka
Institutional Knowledge Archive

Title	Operational context-sensitive guidelines in Canada, China, Japan, and South Korea
Author(s)	Cossette-Lefebvre, Hugo; He, Guangxi; Huang, Lei et al.
Citation	AIOLIA deliverable. 2026, D3.2, p. 1-198
Version Type	VoR
URL	https://hdl.handle.net/11094/104915
rights	
Note	

The University of Osaka Institutional Knowledge Archive : OUKA

https://ir.library.osaka-u.ac.jp/

The University of Osaka



Operational context-sensitive guidelines in Canada, China, Japan, and South Korea

AIOLIA DELIVERABLE 3.2

Horizon Europe Grant Agreement N° 101187937

Project Name	**AIOLIA**
Deliverable Title/Number	D3.2
Description	Operational context-sensitive guidelines in Canada, China, Japan, and South Korea
Lead beneficiary	McGill
Lead Authors	Hugo Cossette-Lefebvre, Guangxi He, Lei Huang, Kazuki Ide, Amelia Katirai, Kwonil Kim, Jonglip Kim, Gaeun Kim, Atsuo Kishimoto, Jocelyn Maclure, Koji Mikami, Jiyoung Suh, Yandong Zhao
Contractual delivery date:	31/03/2026
Actual delivery date:	31/03/2026
Sensitivity	PUBLIC

Document History

Name	Organisation	Role	Action	Date
v1	McGill	Lead	Submission of first preliminary version to partners	March 8, 2026
V1 for review	CASTED, The University of Osaka, STEPI	Contributors	Review and validation	March 8-16, 2026
V2	McGill	Lead	Version sent for quality review	March 18, 2026
V2 for review	CEA, KIT	Reviewers	Review feedback sent	March 18-21, 2026
V3	McGill	Lead	Updated version sent for validation and review	March 25, 2026
V3 for review	CEA, KIT	Reviewers	Final review	March 26, 2026
V4	McGill	Lead	Updated version sent for final check	March 26, 2026
V4 for Check	CEA	Coordinator	Final Check	March 27, 2026

Final version	McGill	Lead	Ready for submission	March 31, 2026

Nature of Deliverable	
R	Document, report

Dissemination level	
PU	Public, fully open

Acronym/abbreviations	
Artificial Intelligence	AI
Use Case	UC
Principles used by the use cases	UC-principles
Large Language Models	LLMs
Retrieval Augmented Generation	RAG
European Union	EU
Multimodal Emotion Recognition	MER
Assessment List for Trustworthy Artificial Intelligence	ALTAI

Acknowledgements

The information and views set out in this report are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein. Reproduction is authorised provided the source is acknowledged.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. Neither the European Union nor the granting authority can be held responsible for them.

Use of AI

AI systems were used to help with sorting out bibliographic references.

EXECUTIVE SUMMARY

Deliverable D3.2 is the result of task *T3.2: International co-creation of AI guidelines*. The aim of T3.2 was to "develop national ethics guidelines for AI research areas based on AIOLIA methodology". This was done using different methods depending on the context, including interviews or semi-structured discussions with key experts and a national stakeholder workshop including key private and public AI developers and policy makers. D3.2 is the result of this process, whereby we chose to conduct the co-creation process on four international use cases in AIOLIA Canada (McGill University), China (CASTED), South Korea (STEPI), and Japan (The University of Osaka) to see how ethics principles can be operationalized in specific cases involving AI technologies.

D3.2 presents the findings of the operationalization process. Firstly, it offers important insights into how ethics principles are defined in practical, international settings by comparing the use cases to each other, and to the AI ethics principles identified in the European Assessment List for Trustworthy Artificial Intelligence (ALTAI). Second, it also offers cross-cutting reflections on variances between the different national contexts and the contextual specificities encountered by the international partners. D3.2 presents how the international partners approached and aimed to ease the tensions arising from the operationalization of the ethics principles and values in their context. This provides practical guidance by illustrating how conflicting ethics principles and practical requirements can be dealt with in real-life scenarios.

The core of D3.2 is the synthesis of the technical and organizational measures identified for the four AIOLIA International use cases, the ethical tensions they encountered, and their national specificities. The appendix within D3.2 presents the full information provided by the use cases concerning their technical and organizational measures.

Results at a glance

Purpose of D3.2

The aim of D3.2 is to offer guidance on how ethics principles are operationalized for ethical development and deployment of AI in practice and to consider how operationalization of ethics principles can vary depending on the national context.

Similarities and differences between the use cases

Despite the different national contexts and the differences between the use cases, recurring concerns emerge in the development and deployment of AI technologies. These concerns are framed as risks with regard to AI ethics principles or components thereof. Anti-bias or discrimination is mentioned by all four international cases; autonomy, respect, and social justice are mentioned by three of them. Avoiding harm and improving welfare are emphasized in the use cases focused on change in human cognition and private behaviour.

The ethics principles and their constitutive components identified by the use cases were compared to the principles identified in the Assessment List for Trustworthy Artificial Intelligence (ALTAI, 2020). This list provides a useful touchstone to assess the similarities and variances between the use cases and will allow for comparing the international use cases to the European use cases analyzed in AIOLIA deliverable 3.1.

On the basis of this comparison, we have found that the principles identified in the use cases (UC-principles) do resonate with the requirements of the Assessment List for Trustworthy Artificial Intelligence (ALTAI, 2020). For most UC-principles, we observe that they straddle different elements identified by the ALTAI principles. This highlights contextual influences and a high synergy between ALTAI principles in practice. The similarities and differences are summarized in the table below.

UC	Ethics principles in use cases	ALTAI Principles (Requirements)

UC-principles covered in ALTAI			
UC8	Reliability, Safety, and Robustness	Req#2	Technical Robustness and Safety
UC8	Privacy, Consent, and Data Protection	Req#3	Privacy and Data Governance
UC7	Proportionality *Refers mainly to the justifiability of the use of the AI systems by the workers*	Req#4	Transparency
UC8 UC9	Fairness and Non-discrimination Fair AI Use	Req#5	Diversity, Non-discrimination, and Fairness
UC-principles addressing sub-parts of ALTAI principles and combining different elements of ALTAI principles			
UC7	Fairness and non-discrimination	Req#4 Req#5	Transparency *Element: explainability* Diversity, Non-discrimination and Fairness *Elements: Avoidance of Unfair Bias; Accessibility and Universal Design.*
UC7	Transparency and Explainability	Req#1 Req#2 Req#4	Human Agency and Oversight *Element: Human Agency and Autonomy* Technical Robustness and Safety *Element: General Safety* Transparency *Element: Explainability*

UC8	Human Agency, Oversight, and Social Harm	Req#1 Req#6	Human Agency and Oversight *Elements: Human Agency; Oversight* Societal and Environmental Well-being *Element: Impact on Society at Large or Democracy*
UC9	Safe Human-AI relationships	Req#1 Req#2 Req#3	Human Agency and Oversight *Element: Human agency and autonomy* Technical Robustness and Safety *Element: General Safety* Privacy and Data Governance *Elements: Privacy, Data Governance*
UC9	Promotion of Social Welfare	Req#1 Req#6	Human agency and oversight *Element: Human oversight* Societal and Environmental Well-Being *Element: Impact on society at large*
UC10	Respect for Postmortem Rights	Req#1 Req#3	Human Agency and Oversight *Element: Human Agency and Autonomy* Privacy and Data Governance *Element: Data Governance*

UC10	Non-maleficence and beneficence	Req#1 Req#3	Human Agency and Oversight *Elements: Human agency and autonomy; Human Oversight* Technical Robustness and Safety *Element: General Safety*
UC10	Justice	Req#5 Req#6	Diversity, Non-Discrimination, and Fairness *Elements: Avoidance of Unfair Bias* Societal and Environmental Well-being *Element: Impact on Society at Large*

Ethical challenges and tensions

While three international cases clustered around the impact of AI on cognition and private behaviour, and only one considered a use case on the impact of AI on human expertise and professional behaviour, common tensions and risks emerge from all the use cases, with the requirement of privacy being the main vector of tension. All four use cases underline potential risks and tensions surrounding trade-offs between privacy and accuracy, effectiveness, or safety. They all include technical measures aimed at limiting the type of data which can be collected by the AI systems or the types of inferences that can be made by the system to protect individual privacy interests. This contrasts with the European use cases where privacy, governed by the GDPR, was flagged as an important but not a central emerging concern in the design of AI systems. Similarly, and in line with the European use cases, all four international studies underline the importance of putting new and innovative regulatory policies or organizational measures in place to protect vulnerable users as AI systems are deployed in the private sphere or in a professional context.

Additionally, as expected, all use cases highlight specificities in their national contexts which ought to be considered to regulate the development and deployment of emerging AI technologies. Of note, the Japanese use case, which has studied ethical issues emerging from the use of emotion recognition technology in the workplace, highlights how different regulatory environments are put in place at the international level, as the use of this technology in that context would likely be banned in the EU.

CONTENTS

RESULTS AT A GLANCE 5
1. INTRODUCTION 14
2. BACKGROUND AND APPROACH 16
3. METHODOLOGY FOR THE DEVELOPMENT OF OPERATIONAL GUIDELINES 19
3.1. GENERAL APPROACH AND UPDATES TO USE CASES 19
3.2. DATA COLLECTION AND NATIONAL STAKEHOLDER WORKSHOPS 29
3.2.1 UC7: WORKPLACES EQUIPPED WITH AI TOOLS FOR BEHAVIOURAL ANALYSIS – THE UNIVERSITY OF OSAKA (JAPAN) 29
3.2.2 UC8: AI SYSTEMS FOR SMART ELDERLY CARE IN WUXI CITY – CASTED (CHINA) 33
3.2.3 UC9: AI SYSTEMS AS PERSONAL COMPANIONS TO ASSIST SENIOR CITIZENS – STEPI (SOUTH KOREA) 35
3.2.4 UC10: AI SYSTEMS AS GRIEF-SUPPORTING PERSONAL ASSISTANTS – MCGILL UNIVERSITY (CANADA) 39
3.3. METHODOLOGICAL REFLECTIONS 42
4. GUIDANCE ON ETHICAL TENSIONS AND CONTEXTUAL SPECIFICITIES 45
4.1. JAPANESE CONTEXT 45
4.2. CHINESE CONTEXT 47
4.3. SOUTH KOREAN CONTEXT 50
4.4. CANADIAN CONTEXT 57
4.5. CROSS-CUTTING REFLECTIONS 59
5. CONCLUSION 62
6. REFERENCES 64
APPENDIX A: TECHNICAL AND ORGANISATIONAL MEASURES TO GUIDE OPERATIONALISATION OF AI ETHICS 67
PRACTICAL MEASURES PROVIDED BY USE CASE 7 67
PRACTICAL MEASURES PROVIDED BY USE CASE 8 112
PRACTICAL MEASURES PROVIDED BY USE CASE 9 144
PRACTICAL MEASURES PROVIDED BY USE CASE 10 185

LIST OF TABLES

Table 1: Use Cases and ethics principles identified by the international partners (based on Table 3 in D2.2, p. 62) 16
Table 2: Short Use Case Descriptions 19
Table 3: Overview of final use cases in D3.2, with changes from D2.2 marked in blue 22
Table 4: UC-Principles and ALTAI Principles 25
Table 5: Components identified by UC7 for each of their ethics principles 33
Table 6: Components identified by UC8 for each of their ethics principles 35
Table 7: Components identified by UC9 for each of their ethics principles 38
Table 8: Components identified by UC10 for each of their ethics principles 41
Table 9: UC 7 – Technical measures to achieve proportionality – Adequacy 67
Table 10: UC 7 – Organisational measures to achieve proportionality – Adequacy 68
Table 11: UC 7 – Technical measures to achieve proportionality – Necessity 71
Table 12: UC 7 – Organisational measures to achieve proportionality – Necessity 74
Table 13: UC 7 – Technical measures to achieve proportionality – Proportionality stricto sensu 77
Table 14: UC 7 – Organisational measures to achieve proportionality – proportionality stricto sensu 81
Table 15: UC 7 – Technical measures to achieve transparency and explainability – Safety 84
Table 16: UC 7 – Organisational measures to achieve transparency and explainability – Safety 88
Table 17: UC 7 – Technical measures to achieve transparency and explainability – autonomy 92
Table 18: UC 7 – Organisational measures to achieve transparency and explainability – autonomy 93
Table 19: UC7 – Technical measures to achieve transparency and explainability – respect 98
Table 20: UC 7 – Organisational measures to achieve transparency and explainability – respect 99
Table 21: UC 7 – Technical measures to achieve Fairness and non-discrimination – anti-bias 102
Table 22: UC 7 – Organisational measures to achieve Fairness and non-discrimination – anti-bias 104
Table 23: UC 7 – Organisational measures to achieve Fairness and non-discrimination – Fair Equality of Opportunity and the Difference Principle 106
Table 24: UC 7 – Technical measures to achieve Fairness and non-discrimination – The equal right to justification 108
Table 25: UC 7 – Organisational measures to achieve Fairness and non-discrimination – The equal right to justification 110

Table 26: UC 8 – Technical measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management 112
Table 27: UC 8 – Organisational measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management 113
Table 28: UC 8 – Technical measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management 115
Table 29: UC 8 – Organisational measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management 117
Table 30: UC 8 – Technical measures to achieve Privacy, Consent, and Data Protection – The privacy-safety paradox 119
Table 31: UC 8 – Organisational measures to achieve Privacy, Consent, and Data Protection – The privacy-safety paradox 121
Table 32: UC 8 – Technical measures to achieve Privacy, Consent, and Data Protection – Sensitive Area Monitoring 123
Table 33: UC 8 – Organisational measures to achieve Privacy, Consent, and Data Protection – Sensitive Area Monitoring 125
Table 34: UC8 – Technical measures to achieve Algorithmic Fairness and Non-Discrimination – Dialect bias 127
Table 35: UC8 – Organisational measures to achieve Algorithmic Fairness and Non-Discrimination – Dialect bias 130
Table 36: UC 8 – Technical measures to achieve Human Agency, Oversight, and Social Harm – Emotional dependency and manipulation 133
Table 37: UC 8 – Organisational measures to achieve Human Agency, Oversight, and Social Harm – Emotional dependency and manipulation 135
Table 38: UC 8 – Technical measures to achieve Human Agency, Oversight, and Social Harm – Over-reliance and deskilling 136
Table 39: UC 8 – Organisational measures to achieve Human Agency, Oversight, and Social Harm – Over-reliance and deskilling 139
Table 40: UC 8 – Technical measures to achieve Human Agency, Oversight, and Social Harm – Accountability and legal positioning 141
Table 41: UC 8 – Organisational measures to achieve Human Agency, Oversight, and Social Harm – Accountability and Legal Positioning 143
Table 42: UC9 – Practical measures to achieve Building relationships based on respect – Guarantee of the right to self-determination 144
Table 43: UC9 – Practical measures to achieve Building relationships based on respect – Preventing Overdependence 152

Table 44: UC9 – Practical measures to achieve Building relationships based on respect – Preventing Misjudgment Due to Deceptive Information/Proposals 156
Table 45: UC 9 – Practical measures to achieve Building relationships based on respect 161
Table 46: UC9 – Practical measures to achieve a Fair Access to Services – Bias Minimization 165
Table 47: UC9 – Practical measures to achieve a Fair Access to Services – Inclusive Access .. 167
Table 48: UC 9 – Practical measures to achieve a Fair Access to Services – Non-Objectification 168
Table 49: UC 9 – Practical measures to achieve a Fair Access to Services 170
Table 50: UC 9 – Practical measures to Promote of Social Well-being – Facilitation of Collaboration 174
Table 51: UC 9 – Practical measures to Promote of Social Well-being (1) 176
Table 52: UC 9 – Practical measures to Promote of Social Well-being (2) 179
Table 53: UC 9 – Practical measures to Promote of Social Well-being (3) 183
Table 54: UC 10 – Technical measures to protect post mortem rights 185
Table 55: UC 10 – Organisational measures to protect postmortem rights 187
Table 56: UC10 – Technical measures to achieve nonmaleficence and beneficence 189
Table 57: UC 10 – Organisational measures to achieve nonmaleficence and beneficence 193
Table 58: UC 10 – Technical measures to aim for justice 195
Table 59: UC 10 – Organisational measures to aim for justice 197

1. Introduction

Deliverable D3.2 is the result of *Task 3.2: International co-creation of AI guidelines*. As stated in the Grant Agreement, the aim of T3.2 is to "develop national ethics guidelines for AI research areas based on AIOLIA methodology". As the research project evolved and to ensure consistency with parallel work in Task 3.1, which aimed to operationalize ethical recommendations on six specific European use cases, the work in T3.2 has been structured to zone in on the question of how ethics concerns emerge from the development and deployment of specific AI technologies, rather than general AI research areas, on four specific use cases. This has allowed us to develop detailed practical analysis of how ethics principles can be specified and operationalized to orient AI development and deployment. All the cases also include practical guidance by highlighting how they have dealt with tensions between the ethics principles and their constitutive components, or between their principles and real-life constraints. This specificity will facilitate upcoming work in *Task 3.3: Develop context-enriched operational guidelines for AI research areas*, where we will identify and lift overarching ethical concerns by cross-analysing the European and international use cases to develop general guidelines.

As stated in the Grant Agreement, the work in *T3.2* ought to be performed "via interviews/focus groups (depending on the cultural context) with key experts and a national stakeholder workshop including key private and public AI developers and policy makers." D3.2 is the result of this process, whereby we chose to conduct the co-creation process on four international use cases in AIOLIA: Canada (McGill University), China (CASTED), South Korea (STEPI), and Japan (The University of Osaka). All the national partners organized one or more stakeholder workshops at the national level. Although none have organized focus groups, use cases UC7, UC8, and UC9 all interviewed relevant actors. UC10 did not use interviews as they preferred to use semi-structured presentations and group discussions (similar to focus groups) to complement the results of their workshops.

The aim of D3.2, in conjunction with D3.1, is to offer concrete guidance to organisations that aim to deploy AI or are already deploying it in industrial settings, by presenting a collection of diverse practical measures. Practical measures refer here "to measurable features, dimensions or attributes related to the chosen ethical principle relevant in the

design or deployment of an AI model or capability" (see D2.3[1], p. 25). Practical measures are either focused on technical aspects or organizational aspects. By presenting different international use cases, the present deliverable also allows us to see how operationalization of ethical principles can vary and be influenced by different regulatory and national contexts outside the EU.

The core of D3.2 is the synthesis of the measures identified by the four AIOLIA international use cases. The synthesis presents the varying methodologies used by the international partners in **Section 3** to illustrate how they adapted the AIOLIA methodology to their contexts. We present how the principles applied in the use cases (UC-principles) compare to one another and to the principles identified in the Assessment List for Trustworthy Artificial Intelligence (ALTAI, 2020). We offer some cross-cutting reflections that present the different challenges and contextual specificities the partners have encountered and how they approached them in **Section 4**. The full set of technical and organizational measures developed by the international partners is included in **Appendix A**.

[1] Shiji, A.N., Bayerl, P.S., & Akhgar, B. (2025). AIOLIA D2.3: Practical Handbook for the Co-Creation Process.

2. Background and Approach

The work in D3.2 was guided by the ethics principles identified in *D2.2: Report on the selection of ethical principles and values*, which presented the "selection of ethical principles and values in relation to the use cases and research areas in the project" (D2.2, p. 3).

The ethics principles identified in D2.2 were informed by a review of relevant literature and ethics frameworks given the national context of the international partners. This review was supplemented by an empirical process involving AIOLIA use case partners and external experts. In this way, the partners identified at least three ethics principles and values per use case (for details about the process, see D2.2). Table 1 presents the use cases and their linked ethics principles and values as identified in D2.2 for the four international use cases.

Table 1: Use Cases and ethics principles identified by the international partners (based on Table 3 in D2.2, p. 62)

Link to human behaviour and cognition	Use case description	Ethics principles
Change in human expertise and professional behaviour	UC7: Workplaces equipped with AI tools for behavioural analysis	1. Proportionality 2. Fairness and non-discrimination 3. Transparency and explainability
Change in human cognition and private behaviour	UC8: AI systems for smart Elderly care in Wuxi City	1. Privacy and data security 2. Emotional dependency and risk of deception 3. Algorithmic bias 4. Accountability
	UC9: AI systems as personal companions to assist senior citizens	1. Prevention of psychological manipulation 2. Diminished autonomy

		3. Accountability (and the requirements of explainability)
	UC10: AI systems as grief-supporting personal assistants	1. Dignity of the Deceased 2. Patient wellbeing 3. Multiculturalism

The approach underlying the operationalization of these principles is one of applied ethics with the ambition to guide the day-to-day execution of the AI ethics work in an organization. The operationalization of AI ethics in the four international AIOLIA use cases, on which the current report is built, thus moves from the higher-level principles identified in D2.2 to the concrete technical and organizational measures stakeholders should take to aim for AI ethics in practice. In contrast to D3.1, however, the international partners had to adapt the AIOLIA methodology to their national contexts, as detailed in Section 3. Despite some adjustments, the international use cases are based on the same underlying approach developed by AIOLIA.

Accordingly, the international use case relied on the same terminology defined in D3.1. and in D2.3, as detailed below.

> Operationalization in this context "refers to the process of translating high-level ethical principles into practical actions, tools, processes and governance structures that can guide and be applied throughout the lifecycle of AI systems to ensure ethical design, development, deployment and use." (see D2.3, p. 11)

The practical focus includes both technical and organizational measures, detailed in **Appendix A**, to ensure that the operationalization covers both the technical/design features of AI and the context-specific human aspects that need to be addressed by decision-makers in the organizations that design, procure and deploy the AI systems.

As defined in the *Handbook for the Operationalization of Ethics* (D2.3), technical and organizational measures are understood as the following (cf. D2.3, p. 25):

Technical Measures: Technical methods focus on the design and technical aspects of AI systems and refer to specific tools, methodologies, technologies and processes that are implemented in AI systems to ensure it operates in an ethical manner. Technical measures to foster ethical AI practices include the addition of privacy-by-design approaches, Explainable AI (XAI) measures, use of benchmarks and key performance indicators, adversarial testing, federated learning, data anonymization techniques and security audits.

Core audience: engineers, AI developers

Organizational Measures: Organizational measures focus on how an organization incorporates and manages ethical AI practices by referring to the structures, policies and governance framework in place. Organizational measures for ethical AI governance include the development of AI ethics boards, ethical AI policies, promoting community stakeholder engagement, fostering interdisciplinary collaboration, regulatory and legal compliance to existing regulations, ethics readiness indicators and the development of AI risk frameworks.

Core audience: management, training and HR departments

Both types of practical measures are required to implement AI ethics comprehensively and effectively. The core audiences responsible for their implementation differ, indicating that AI ethics requires close, multi-disciplinary collaboration to succeed.

3. Methodology for the Development of Operational Guidelines

3.1. GENERAL APPROACH AND UPDATES TO USE CASES

The guidelines are a product of a co-creation process which involved one or several national stakeholder workshop(s) including key private and public AI developers and policy makers. The stakeholder workshops were complemented by other interactions, such as interviews with key experts or semi-structured presentations and group discussions. International partners adapted the AIOLIA methodology detailed in D2.3 to consider the realities of their national contexts and ensure a collaborative and participatory process with stakeholders from academia and industry or non-academic technical partners within each use case.

During work in T3.2, international partners made the following changes:

- UC8 made changes to the ethics principles originally identified in D2.2;
- UC9 made changes to the ethics principles originally identified in D2.2;
- UC10 made changes to the ethics principles originally identified in D2.2.

Below, Table 2 provides a description of the use cases. Table 3 provides an updated view of the use cases and their respective ethics principles, as they were included in D3.2 (changes compared to D2.2 are marked in blue).

Table 2: Short Use Case Descriptions

UC7: Workplaces equipped with AI tools for behavioural analysis (The University of Osaka - Japan)
This UC examines workplaces equipped with AI tools for behavioural analysis with emotion recognition. This system involves installing numerous cameras and biometric sensors in the workplace to analyze the collected data. This is a realistic yet fictional use case, modelled after AI tools under development by major electronics manufacturers. In Japan, the use of emotion recognition technology in the workplace is not specifically regulated and the boundary between work orders that workers cannot refuse and those they can choose has tended to be ambiguous. The collection of biometric information is one such example. Behavioural analysis with emotion

recognition can be used to analyze work processes, measure fatigue levels, concentration levels, mental health, well-being of workers, or be utilized for operational improvements, performance evaluations, and occupational safety management. The impact of the development and deployment of these systems raises questions concerning proportionality, fairness and non-discrimination, and transparency and explainability.

UC8: AI systems for smart elderly care in Wuxi City (CASTED – China)

This UC examines the core risks of the Wuxi Smart Elderly Care System. Nursing homes generally face challenges of staff shortages and high labour costs. A significant amount of daily work, such as night patrols, scheduled reminders, and vital sign monitoring, is repetitive, low-skill labour highly suitable for automation. The core of this case is a full-stack, independently developed smart companion robot "Datou Aliang" (Institutional Version). The system is built around a "1+6" scenario framework, meaning one hardware platform supports six core scenarios, which can be summarized into three main functions: (1) safety; (2) health; (3) companionship. This use case presents how to operationalize reliability, safety, and robustness; privacy, consent, and data protection; algorithmic fairness and non-discrimination; and human agency, oversight, and social harm avoidance.

UC9: AI systems as personal companions to assist senior citizens (STEPI – South Korea)

This use case examines personalized conversational AI robots for emotional well-being of older adults, which have emerged as a vital public care solution amid rapid aging and caregiver shortages. Distributed to solitary elderly households via municipal projects, these robots are managed by caregivers from Long-Term Care Insurance-funded institutions who conduct monitoring and usage assistance. Furthermore, the infrastructure is integrated with local emergency services to ensure rapid dispatch during critical situations

AI conversational care robots are intelligent caregiving systems that integrate artificial intelligence (AI), speech recognition, sensor networks, and emotional state detection technologies to support emotional interaction, safety monitoring, medication management, and cognitive training for older adults and other vulnerable populations. The core functions of AI conversational care robots can be summarized in three areas: (1) emotional communication through conversation; (2) safety and health monitoring; and (3) cognitive and physical stimulation. The three main ethical principles to assess the relational risks arising from interactions between AI and

seniors, as well as AI and service providers in this case were: safe human-AI relationships; fair AI service; and promotion of social welfare.
UC10: AI systems as grief-supporting personal assistants
This use case examines the ethical challenges associated with the development and deployment of griefbots. Griefbots are AI chatbots designed to simulate the personality and speech of a recently deceased person to allow their loved ones to maintain a certain relationship with the deceased during their grieving process. They range from basic chatbots interacting with users through text-based interfaces to more advanced virtual avatars or robots aiming to mimic how a person thought, looked, sounded, and moved. This use case considered three ethical principles: respect for postmortem rights; nonmaleficence and beneficence; and justice.

The international use cases focused on the change in human cognition and private behaviour, the only exception being UC7, which focused on change in human expertise and professional behaviour. The use cases focused on a wide array of ethics principles. The set of the 12 UC-principles analyzed in these use cases is:

- Proportionality;
- Fairness and non-discrimination (mentioned in both UC7 and UC8);
- Transparency and explainability;
- Reliability, safety, and robustness;
- Privacy, consent, and data protection;
- Algorithmic fairness and non-discrimination;
- Human agency, oversight, and social harm;
- Safe human-AI relationships;
- Fair AI service;
- Promotion of social welfare;
- Respect for postmortem rights;
- Justice.

Table 3: Overview of final use cases in D3.2, with changes from D2.2 marked in blue

Link to human behaviour and cognition	Use case description	Ethics principles
Change in human expertise and professional behaviour	UC7: Workplaces equipped with AI tools for behavioural analysis	1. Proportionality 2. Fairness and non-discrimination 3. Transparency and explainability
Change in human cognition and private behaviour	UC8: AI systems for smart Elderly care in Wuxi City	1. Reliability, Safety, and Robustness 2. Privacy, Consent, and Data Protection 3. Algorithmic Fairness and Non-Discrimination 4. Human Agency, Oversight, and Social Harm.
	UC9: AI systems as personal companions to assist senior citizens	1. Safe human-AI relationships 2. Fair AI Service 3. Promotion of social welfare
	UC10: AI systems as grief-support personal assistants	1. Respect for postmortem rights 2. Nonmaleficence and beneficence 3. Justice

Despite this apparent diversity, we can nonetheless see that the UC-principles cluster around common overarching concerns. As in D3.1, it is instructive to consider how these UC-principles compare to the general principles identified in the Assessment List for Trustworthy AI (ALTAI; High Level Expert Group on AI, 2020). ALTAI offers a compliance checklist to support the implementation of ethics principles based on seven uncontroversial AI ethics principles or key requirements in the EU. In a way that is similar to the methodology developed in AIOLIA, ALTAI aims to translate high-level ethics principles into actionable, technical, and organizational recommendations. To that end, each principle is broken down into smaller "elements" or "issues." The principles identified by ALTAI provide a useful touchstone to group the diverse

principles identified in the use cases and evaluate whether they align with ALTAI or highlight other ethical considerations. These seven ALTAI-principles, or key requirements, along with their elements are:

1) Human agency and oversight;
    (i) Human agency and autonomy
    (ii) Human oversight
2) Technical robustness and safety;
    (i) Resilience to attack and security
    (ii) General safety
    (iii) Accuracy
    (iv) Reliability, fall-back plans and reproducibility
3) Privacy and data governance;
    (i) Privacy
    (ii) Data governance
4) Transparency;
    (i) Traceability
    (ii) Explainability
    (iii) Communication
5) Diversity, non-discrimination and fairness;
    (i) Avoidance of unfair bias
    (ii) Accessibility and universal design
    (iii) Stakeholder participation
6) Environmental and societal well-being;
    (i) Environmental well-being
    (ii) Impact on work and skills
    (iii) Impact on society at large or democracy
7) Accountability;
    (i) Auditability
    (ii) Risk management

As summarized in Table 4 below, the UC-principles generally fall within the principles identified by ALTAI or combine different elements picked up by ALTAI principles. All

the use cases include issues pertaining to (5) Diversity, non-discrimination and fairness. UC8, UC9, and UC10 all mention elements of (1) Human agency and oversight; (2) Technical Robustness and Safety; and (6) Societal and environmental well-being. In contrast, UC7 is the only one that mentions elements of (4) Transparency, at the level of its principles. This potentially reflects the fact that UC7 is focused on the impact of AI on human expertise and professional behaviour, while the other use cases zone in on human cognition and private behaviour.

Importantly, the use cases have used many UC-principles straddling different ALTAI principles and elements. UC7 uses the principle of fairness and non-discrimination. Although it sounds like the ALTAI principle (5) Diversity, non-discrimination and fairness, it in fact combines elements of both ALTAI principle (5) and (4) Transparency. The UC-principle used by UC7 connects fairness and a right to justification and explainability. In the same way, UC7 uses the UC-principle transparency and explainability, which echoes ALTAI principle (4) Transparency. Yet, they use this principle to connect it to other values beyond traceability, explainability, and communication, as they connect it with safety, autonomy, respect, and legitimacy. This underlines that beyond apparent similarities, principles can be understood in different ways depending on the context.

Similarly, UC8 identified the UC-principle of human agency, oversight, and social harms that combines elements of (1) Human agency and oversight, and (6) Societal and environmental well-being. UC9 identified safe human-AI relationship which connects both (1) Human agency and oversight – as this ALTAI principle explicitly includes how AI systems that 'act' like humans affect human perceptions and expectation –, and (2) Technical robustness and safety – by considering the impact of human-AI relationship on general safety. UC9 also uses the principle of promotion of social welfare which interestingly connects social welfare, promotion of productive human-AI cooperation and adaptive governance. In that way, this UC-principle straddles societal well-being (ALTAI (6)) and human oversight (ALTAI (1)). Finally, UC10 uses non-maleficence and beneficence, classic bioethical principles not explicitly mentioned by ALTAI, but which nonetheless combine elements of (1) Human agency and oversight, and (2) Technical robustness and safety. Their principle centered on justice also

combines elements of (6) Diversity, non-discrimination, and fairness, and (6) Societal and environmental well-being. Interestingly, although their principle focused on protecting postmortem rights could seem quite idiosyncratic, it resonates with both (1) Human agency and oversight, and (3) Privacy and data governance, as it insists on the importance of securing consent before death and on the importance of secure personal data management, as detailed below in Section 3.2.4.

This underlines how ethics principles do not operate in a silo but are deeply connected to one another in practice. Their operationalization underlines the synergy between the principles, along with the potential tensions that might arise between them, as discussed in Section 4.

Table 4: UC-Principles and ALTAI Principles

UC	Ethics principles in use cases	ALTAI Principles (Requirements)	
UC-principles covered in ALTAI			
UC8	Reliability, Safety, and Robustness	Req#2	Technical Robustness and Safety
UC8	Privacy, Consent, and Data Protection	Req#3	Privacy and Data Governance
UC7	Proportionality *Refers mainly to the justifiability of the use of the AI systems by the workers*	Req#4	Transparency
UC8 UC9	Fairness and Non-discrimination Fair AI Use	Req#5	Diversity, Non-discrimination, and Fairness

UC-principles addressing sub-parts of ALTAI principles and combining different elements of ALTAI principles			
UC7	Fairness and non-discrimination	Req#4 Req#5	Transparency *Element: explainability* Diversity, Non-discrimination and Fairness *Elements: Avoidance of Unfair Bias; Accessibility and Universal Design.*
UC7	Transparency and Explainability	Req#1 Req#2 Req#4	Human Agency and Oversight *Element: Human Agency and Autonomy* Technical Robustness and Safety *Element: General Safety* Transparency *Element: Explainability*
UC8	Human Agency, Oversight, and Social Harm	Req#1 Req#6	Human Agency and Oversight *Elements: Human Agency; Oversight* Societal and Environmental Well-being *Element: Impact on Society at Large or Democracy*

UC9	Safe Human-AI relationships	Req#1 Req#2 Req#3	Human Agency and Oversight *Element: Human agency and autonomy* Technical Robustness and Safety *Element: General Safety* Privacy and Data Governance *Element: Privacy, Data Governance*
UC9	Promotion of Social Welfare	Req#1 Req#6	Human agency and oversight *Element: Human oversight* Societal and Environmental Well-Being *Element: Impact on society at large*
UC10	Respect for Postmortem Rights	Req#1 Req#3	Human Agency and Oversight *Element: Human Agency and Autonomy* Privacy and Data Governance *Element: Data Governance*
UC10	Non-maleficence and beneficence	Req#1 Req#3	Human Agency and Oversight *Elements: Human agency and autonomy; Human Oversight* Technical Robustness and Safety *Element: General Safety*

UC10	Justice	Req#5 Req#6	Diversity, Non-Discrimination, and Fairness *Elements: Avoidance of Unfair Bias* Societal and Environmental Well-being *Element: Impact on Society at Large*

3.2. DATA COLLECTION AND NATIONAL STAKEHOLDER WORKSHOPS

The data collection for the co-creation of the guidelines varied between the international partners who all consulted various national or international regulatory documents and followed different steps to include relevant academic and industrial or non-academic technical partners. Below, we detail how each partner organized their data collection and identified the components for each of their ethics principles. In agreement with the Grant Agreement, the main step in data collection was the organization of national workshops in partner countries.

3.2.1 UC7: Workplaces equipped with AI tools for behavioural analysis – The University of Osaka (Japan)

The starting point for the identification process was prior research conducted between members of the research team and industrial partners. This led to a project synthesizing recent literature on the ethics of workplace surveillance, as well as one on the ethics of emotion recognition systems (emotional AI). One part of this research involved a qualitative analysis of recommendations within academic literature for ethical principles to guide the use of emotion recognition (Katirai, 2023). This prior research was used as a starting point and then generalized to relevant technologies within the use case.

The primary domestic reference points were generated through prior research with industry partners including NEC, Mitsubishi Electric, and Ricoh. Across these research projects, a key aim has been to consider ethical principles and guidelines for emerging applications of AI technologies. Within their collaboration with NEC, for example, members of this use case designed a collection of ethics principles and checklists for appropriate use of facial recognition technologies (NEC 2024). Additionally, the Japanese government promulgated the *Act on Promotion of Research and Development and Utilization of AI-related Technology* (AI Act) in 2025. The *AI Guidelines for Business* were published in 2024 and have been updated frequently since then. These guidelines cite the "Social Principles of Human-Centric AI," which consists of seven principles and was formulated by the Cabinet Office in 2019.

The primary international reference point was the *UNESCO Recommendation on the Ethics of Artificial Intelligence* (2021). The principles contained in this document were compared with the principles identified by the initial literature review to refine the selection and framing.

After establishing three relevant ethical principles, the researchers of UC7 derived their respective components based on a literature review (Table 5). Then, they held two workshops bringing together industry practitioners to verify the validity of the components and to identify risks. The workshops were held on August 26, 2025 and October 22, 2025, in Tokyo as part of the *AI Risk Study Group*[2] at the *Japan Standards Association* (JSA) Seminar Room.

In the first workshop, a hypothetical case scenario concerning the use of emotion recognition AI in the workplace was presented. The 20 participants were then divided into groups of four to six and asked to identify and discuss the risks associated with the case. After sharing the risk items identified during the workshop with all participants, the University of Osaka research group further discussed and refined the results. The team ultimately consolidated the risks into the following twelve categories:



1) *Invasion of privacy*: Biometric data could reveal sensitive personal information such as underlying medical conditions or pregnancy.
2) *Excessive collection of personal data*: Data may be collected with the intention of collecting it "just in case," even though it is unnecessary for the stated purpose.

[2] The "AI Risk Study Group" is a research initiative conducted by The University of Osaka and the Japanese Standards Association since May 2025.

3) *Chilling effects of surveillance*: Being monitored—including during breaks or casual conversations—may negatively affect workers' behaviour.



4) *Excessive guidance or disadvantages based on misclassification*: Misjudgments caused by the inaccuracy of emotion-recognition AI may lead to unfair treatment, especially if supervisors place undue trust in AI outputs.



5) *Communication challenges*: Workers may not be informed about the introduction of the system or may not understand its benefits and drawbacks.
6) *Discrimination against specific workers*: Individuals who have difficulty expressing facial emotions due to illness or disability may receive unfair evaluations.
7) *Manifestation of bias*: Workers belonging to groups not well represented in the model's training data may be evaluated unfairly.
8) *Violation of human dignity and autonomy*: AI outputs may be applied directly without appropriate human oversight.
9) *Concerns about data handling*: Workers may not know how their biometric data is managed, who has access to the analysis results, or how long the data is stored.
10) *Ineffective consent*: When supervisors request consent from subordinates, workers may feel unable to refuse, making the consent effectively invalid.
11) *Strategic behavior by workers*: Knowing that their biometric data is being analyzed, workers may intentionally modify their behavior (e.g., forced smiles, hiding fatigue).
12) *Lack of benefits for workers*: Improved work efficiency may translate into increased workload for workers, resulting in greater fatigue.

In the second workshop, the 22 participants were divided into groups of four to seven members. Each group conducted a risk assessment in accordance with the following steps.

1) For each of the 12 risk items identified and analyzed in the first workshop, identify and discuss "the value to be protected."
2) For each "value to be protected," identify appropriate parameters (indicators) for evaluating impact and likelihood.
3) Assess the indicators for impact and likelihood on a four-level scale for each "value to be protected."
4) Score the impact and likelihood for each of the 12 risk items.
5) Propose technical and organizational measures to ensure that unacceptable risks are eliminated.
6) Score the impact and likelihood for each risk item again after the proposed measures are applied.

By following this procedure, participants were able to demonstrate—within the bounds of what can be reasonably anticipated for the target technology—that "no unacceptable risks remain," consistent with the ISO/IEC Guide 51 definition of *safety*, which is defined as "freedom from risk which is not tolerable."

UC7 also conducted a one-hour online interview on September 26, 2025 with Mr. Hirono of Ricoh Co., Ltd., who was both a workshop participant and a member of the steering committee for the *AI Risk Study Group*. Mr. Hirono is a quality assurance expert who established the company's ethics guidelines for technology development and technology assessment processes.

After the workshops and the interview, the research team examined whether these twelve risks could be adequately captured by the components derived from the three ethical principles and their components. As a result, only one risk out of the 12 items was found to be uncaptured: "strategic behaviour by workers." This refers to the possibility that workers may respond strategically if they are aware of being monitored and analyzed in the workplace. For instance, they might pretend not to be tired even when they are. This suggests that a top-down approach based solely on a literature

review might overlook certain risk items. Therefore, combining top-down approach (literature review) and bottom-up approach (workshop and interviews) enabled more robust risk mitigation in this case.

Table 5: Components identified by UC7 for each of their ethics principles

UC7: Workplaces equipped with AI tools for behavioural analysis – The University of Osaka (Japan)		
Proportionality	**Fairness and Non-discrimination**	**Transparency and Explainability**
1. adequacy 2. necessity 3. proportionality *strictu sensu*	1. anti-bias 2. fair equality of opportunity and the difference principle 3. equal right to justification	1. safety 2. autonomy 3. legitimacy 4. respect

3.2.2 UC8: AI systems for smart elderly care in Wuxi City – CASTED (China)

In UC8, ethical principles were identified through field research and multi-stakeholder discussions (e.g., developers, enterprises, the elderly, and their families). The "Co-Creation" methodology involved governments, academia, enterprises, and care institutions to embed ethical governance rules in technology development. Analysis of the case studies (e.g., Wuxi's "Datou Aliang 大头阿亮" Robot) prioritized issues like data security, privacy protection, and employment impacts on caregivers.

The primary domestic sources included the *Chinese Ministry of Civil Affairs' elderly care policies* (2025), *National Data Standardization Committee' s data security standards* (2019), *Information Security Technology-Data Security Capability Maturity Model* (2019), *Information Security Technology-Personal Information Security Specification* (2020), *Governance Principles for the New Generation of Artificial Intelligence – Developing Responsible Artificial Intelligence* (2019)，and the *Ethical Norms for New Generation Artificial Intelligence* (2025). During the research project, the members of the research team also consulted industry, academic and governmental actors including:

- Industry:
  - Jiangsu AIYU Wencheng Elderly Care Robot Co., Ltd （江苏艾雨文承养老机器人有限公司）
- Academic:
  - Renmin University of China;
  - Guangxi University of Science and Technology;
    - China Civil Affairs University;
  - Government:
    - Wuxi Elderly Care Association

As for the international reference points, while no explicit documents were named in the meetings, the EU' s *Ethics Guidelines for Trustworthy AI* (2019) were referenced by the research team to align with EU ethical governance standards.



The research team conducted interviews with the stakeholders (Wuxi Company; institutional end users including nurses and doctors; external ethics experts) and held a national workshop. Overall, 12 interviews were conducted (including 4 members of the



administration, 4 technicians, 2 end users, and 2 government employee). The *AI Ethics Value Co-creation from a Multi-Stakeholders Perspective National Workshop* was held on September 20th, 2025 in Nanning. It took place at the *China-ASEAN Artificial Intelligence Cooperative Innovation Center (CAAIC)* during the China-ASEAN Expo, which was organized by the *Chinese Academy of Science and Technology for Development* (CASTED), and strongly supported by co-organizer the *Guangxi University of Science and Technology* (GXUST). Through this workshop, the UC8 research team brought together experts from government, industry, and academia. There were 22 participants, including 20 researchers and 2 students (post-graduate and Ph.D.).

Based on these analyses, interviews, and workshop, UC8 researchers identified four main ethics principles, each comprising between two and three components, as detailed in Table 6.

Table 6: Components identified by UC8 for each of their ethics principles

UC8: AI systems for smart elderly care in Wuxi City – CASTED (China)			
Reliability, Safety, and Robustness	Privacy, Consent, and Data Protection	Algorithmic Fairness and Non-Discrimination	Human Agency, Oversight, and Social Harm
1. Missed Report 2.Misreport/Alert Fatigue Management 3. Human-in-the-loop as a safety net.	1.The privacy-safety paradox 2.Sensitive Area Monitoring 3.Data access control & third-party vetting	1. Dialect bias 2. Cognitive impairment & inclusivity.	1. Emotional dependency and manipulation 2.Over-reliance and deskilling 3.Accountability and legal positioning

3.2.3 UC9: AI systems as personal companions to assist senior citizens – STEPI (South Korea)

The process of deriving ethical principles and specifying technical and organizational measures was carried out through three stages. The first stage consisted of a literature review, conducted to obtain a general understanding of the risks associated with

conversational AI for elderly care and possible response measures. The next stage involved qualitative research based on insights from various stakeholders and experts, including interviews, workshops, and written questionnaires. The purpose of this stage was to identify concrete risks and necessary response measures that reflect the specific technological, institutional, and cultural context of Korea. In the third stage, expert consultation was conducted to examine the validity and feasibility of the technical and organizational measures that had been identified.



The literature review in the first stage was conducted along two dimensions. One was to explore the various risks and ethical issues that may arise when AI, as a general-purpose technology, is applied in the care sector. To this end, Korean and international academic studies were reviewed (Deusdad 2024; Kim 2024; Kubota et al. 2021; Sharkey 2012; Suh 2025; Yuan et al. 2023). The other dimension was to examine trends in normative frameworks designed to address anticipated risks in the care sector. For this purpose, the *Korean AI Basic Act* (December 2024), the *EU AI Act*, the *UNESCO Recommendation on the Ethics of Artificial Intelligence* (2021), and policy reports on AI published by the OECD (OECD, 2023) were reviewed. Particular attention was given to Korean materials, including the *AI Ethics Guidelines of the Ministry of Science and ICT* (MSIT, 2019), the *AI Ethics Impact Assessment framework* developed by KISDI with support from MSIT, and the *Ethics Self-Check List for AI Engineers* developed by the *Telecommunications Technology Association* (TTA, 2025). By reviewing existing ethical issues and normative approaches, it was possible to develop a list of risks and ethical concerns that may arise in conversational AI for seniors.

In the second stage, the risks and response measures identified through the literature review were assessed within the Korean context, and more specific ethical principles and components were derived. Interviews, workshops, and written questionnaires were adopted as the main methodological approaches.

The interviews were conducted with a total of 20 participants, comprising approximately five to six individuals from each of the following groups: AI engineers and managers, AI users (i.e., stakeholders in the social welfare sector), and researchers specializing in AI ethics and policy. To further specify the requirements for technical and organizational measures, the study subjects were limited to products from three companies (Hyodol, ROAIGEN, and RoboCare). Prior to the interviews, documents summarizing the ethical issues identified through the literature review and current trends in technological development were distributed to the participants in advance so that they could reflect on the ethical questions. Through the interviews, it was possible to gain an understanding of the AI model development processes and development environments in Korean companies, as well as the service conditions and institutional frameworks of social service providers. This process also enabled the initial identification of key ethical issues that require particular attention in the Korean context.



The workshops focused on identifying the most important principles to be upheld and their detailed components. For this purpose, the previously identified list of risks was used to conduct a risk assessment, along with an evaluation of the controllability and necessity of managing the sources of harm. After the workshops, additional opinions were collected from participants through written questionnaires (overall the research team received 15 completed written questionnaires) .

The interviews, workshops, and written questionnaires can be regarded as complementary methodologies for deriving AI ethical principles within the Korean context. These activities were conducted over a total period of eight months, from April

to November 2025. Two workshops were held on 17 September and 26 November, while interviews and written questionnaires were conducted both before and after the workshops. The first workshop focused on identifying risks, risk factors, and ethical principles related to interactions between personal companion robots and older adults, their families, social-service providers, and service institutions. The second workshop focused on discussions around risk severity and the need for risk controls. Ten participants attended the first workshop, while a total of 15 participants took part in the second. Workshop participants were primarily drawn from the pool of interviewees, as they had already engaged in in-depth reflection on the characteristics of the case studies and the associated ethical issues. To ensure continuity in the discussions, participants from the first workshop were re-invited to the second, with the addition of several experts.

Finally, expert consultation was conducted by selecting several experts who had participated in the workshops and asking them to review the proposed technical and organizational measures. The guideline review was conducted in two rounds. A total of four experts from the fields of social welfare, engineering, and policy participated in the final review. In addition, discussions were held with officials from government ministries and social service institutions to exchange views on the principles and measures and to assess their necessity and feasibility for implementation.

On the basis of this plural approach combining literature reviews, interviews and participatory workshop, the UC9 research team identified the principles and components identified in Table 7.

Table 7: Components identified by UC9 for each of their ethics principles

UC9: AI systems as personal companions to assist senior citizens – STEPI (South Korea)		
Safe Human-AI Relationship	Fair AI Service	Promotion of Social Welfare

1. Respect for autonomy & decision-making authority 2. Prevention of overdependence 3. Protection from deception and misjudgment	1. Mitigate data bias 2. Promotion of inclusivity 3. Non-objectification	1. Promotion of human-AI cooperation 2. Adaptive governance 3. Prevention of shadow labour

3.2.4 UC10: AI systems as grief-supporting personal assistants – McGill University (Canada)

The guidelines to inform how to approach griefbots were co-created by interacting with academic researchers, a grief expert, and persons working close to industry to connect philosophical insight, technical expertise, and firsthand experience on how to approach the grieving process in practice and how to ensure that chatbot systems are developed and deployed ethically.



The primary domestic resources involved academics from diverse regions in Canada, including McGill University, the University of Toronto, the Université de Montréal, Concordia University, Université Laval, and the Université du Québec à Montréal (UQÀM). On the industrial side, the UC10 research team also involved a consultant working with MILA – the Québec AI Institute. MILA provides consulting services to industries and startups working with LLMs and RAG systems. Finally, McGill also collaborated with a mental health professional working on grief to ensure that they take the interests of the

grievers and the realities of psychological care into account. The research team in UC10 also analyzed different federal and provincial legislation. These included the *Personal Information Protection and Electronic Documents Act* (PIPEDA 2024), the *Copyright Act* (1985), the *Criminal Code of Canada* (1985), Ontario' s *Succession Law Reform Act* (1990), and the *Montréal Declaration for Responsible AI* (2018).

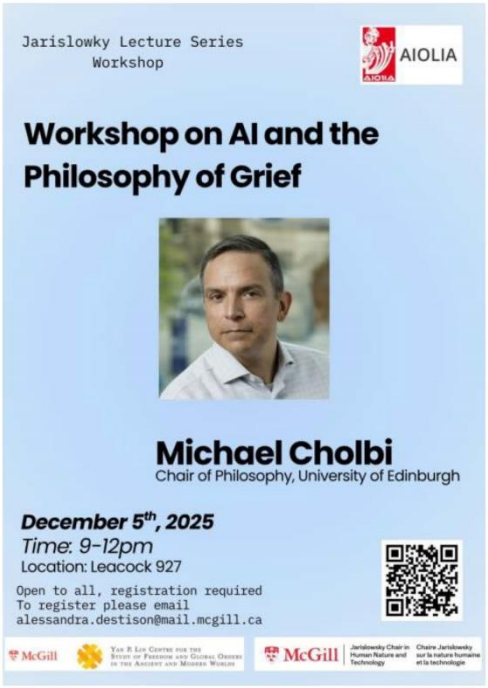
The primary international references considered included the *UN Report on the Protection of the Dead* (2024), the *EU AI Act* (2024), and the *New York Right of Publicity Law* (2021). The research team of UC10 also organized a conference with Professor Michael Cholbi from the University of Edinburgh, a specialist in the philosophy of death and the ethics of Griefbots, on December 4-5, 2025.

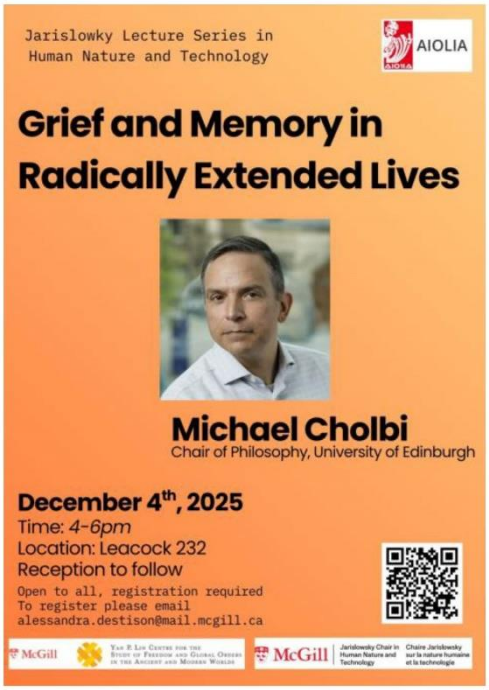






The UC10 research team organized 2 national workshops, one on June 11-12, 2025 at McGill University (5 presenters), and one on December 11-12, 2025 (4 presenters), which took place at the University of Toronto. Counting the presenters, the research team, and the other attendees there were a total of 13 participants in the first workshop, and 12 for the second. For these workshops, the UC10 research team used semi-structured presentations and group discussions where the participants and presenters discussed their perspective and expertise on topics related to griefbots (deployment of AI systems in healthcare, grief, design of LLM powered





chatbots, etc.). Each presentation lasted between 45 minutes and 2 hours and involved academics from different disciplines (Philosophy, Law, Public Health, Computer Science, and Psychiatry) and non-academic participants (Consultants in Responsible AI, Grief Councilors). In addition to the workshops, the McGill research team also organized two semi-structured discussions with an expert on the philosophy of grief (Prof. Michael Cholbi – University of Edinburgh) and an expert on human-machine relations and indigenous communities (Ceyda Yolgörmez, Ph.D.).

The initial literature review allowed the research team to identify relevant ethical principles and components, which were iteratively refined and validated through the discussions in the workshops. The principles and components they identified are presented in Table 8.

Table 8: Components identified by UC10 for each of their ethics principles

UC10: AI systems as grief-support personal assistants – McGill University (Canada)		
Respect for Postmortem Rights	Nonmaleficence and Beneficence	Justice

1. Informed consent 2. Secure personal data management 3. Output/Export control	1. Retirement protocols 2. Restricted use 3. Automated monitoring and human oversight	1. Use of culturally sensitive training datasets 2. Reinforcement learning with human feedback 3. Community engagement reports.

3.3. METHODOLOGICAL REFLECTIONS

The international uses cases demonstrate the great variety of national policies and regulations relevant to the development of practical measures to operationalize ethics principles. Together, the cases have considered no less than seventeen national legal documents, guidelines, or policy documents and five international documents.[3] Notably, the UNESCO *Recommendation on the Ethics of AI* (2021) was used by both US7 and US9; and all four international use cases consulted documents from the EU, including the *EU AI Act* (2024) and the *EU Ethics Guidelines for Trustworthy AI* (2019).

The formats of the national workshops are relevantly similar. Together, they brough together 114 participants during the workshops (UC7 – 42; UC8 – 22; UC9 – 35; UC10 – 25). They all display a multi-disciplinary, iterative process that included academic and industrial or non-academic technical partners alongside extensive literature reviews. All international partners started from the identification of the relevant principles and components on the basis of a literature review, which was then iteratively refined and developed through interactions with different stakeholders and partners, including non-academic partners.

The national workshops were complemented by interviews or semi-structured discussions with key experts. UC7 interviewed one quality assurance expert; UC8 conducted 12 interviews with diverse actors including technicians, administrators, end

[3] Note that the *Montréal Declaration for a Responsible AI* (2018) considered in UC10 was counted as an international document, while the New York *Right to Publicity Law* (2021) was counted as a national document.

users, and government employees; UC9 conducted 20 interviews with AI engineers and managers, AI users (i.e., stakeholders in the social welfare sector), and researchers specializing in AI ethics and policy. Although UC10 did not include interviews, they did organize two semi-structured discussions with key experts. It is of note that UC8 and UC9 were more interview-driven than the two other use cases. UC8 and UC9 also directly included government representatives, beyond academic and non-academic industrial or technical partners. CASTED (UC8) and STEPI (UC9) are government think-tanks, which likely explains the shape the stakeholders' consultations took and their more pronounced focus on policy.

Together, the use cases identified 12 ethics principles, since both UC7 and UC8 use the principle of fairness and non-discrimination (albeit in different ways), and 39 components. Importantly, despite the different national contexts and the differences between the cases, recurring concerns and risks emerge either as principles or components. Anti-bias or discrimination is mentioned by all four international cases; autonomy is mentioned by UC7, UC8, and UC9; respect is mentioned explicitly by UC7, UC9, and UC10; and social justice, harm, or welfare is mentioned by UC8, UC9, and UC10 – the three use cases focused on change in human cognition and private behaviour.

The comparison with ALTAI principles, initiated in Section 3.1, is also instructive to see how to approach the operationalization of ethics principles in practice. As mentioned in Section 3.1, many UC-principles generally fall within the principles identified by ALTAI. Most explicitly, UC8 uses the principle of privacy, consent, and data protection which clearly falls within ALTAI principle (3) Privacy and Data Governance. Similarly, UC9 uses UC-principle fair use of AI, which clearly falls within ALTAI principle (5) Diversity, Non-discrimination and Fairness. Other UC-principles also directly connect with ALTAI principle even if they are named differently (Reliability, Safety and Robustness (UC8); Proportionality (UC7); Fairness and Non-discrimination (UC8)).

Nonetheless, interesting variations emerge from the comparison of these use cases. First, a same ethics principle can be operationalized in different ways depending on the context and be used to address different concerns. Compare how UC7 and UC8 used

the UC-principle of fairness and non-discrimination. UC8 used this principle to zone in on how the deployment of carebots should avoid dialect bias and be designed in inclusive ways to cover the needs of older adults with cognitive impairments. This is directly in line with how ALTAI presents its own principle (5) Diversity, non-discrimination and fairness. In contrast, UC7 used this principle to capture anti-bias, fairness of opportunity, and an equal right to justification. In this way, fairness is explicitly connected to explainability in UC7.

Moreover, many UC-principles combine different elements identified in ALTAI. Note that this does not imply that the UC-principles are inconsistent, but rather that they use and deploy ethics principles in ways that are tailored to their use cases and that underline the synergies between ethical principles and elements. This supports the idea that high-level ethical principles can be relatively fluid in how they are operationalized and implemented. This also points toward the conclusion that ethics principles do not operate independently of one another but interact in complex ways in practice, as highlighted in Section 3.1. For instance, UC7 shows how fairness, non-discrimination and explainability interact with one another. Similarly, UC10 highlights how protection of human autonomy, privacy and data governance are connected in complex ways.

However, the connection between these different ethics principles and elements are not only synergistic. As discussed in Section 4, many tensions between ethical desiderata also emerged from the use cases.

4. Guidance on ethical tensions and contextual specificities

The international partners were all asked to reflect on the ethical and technical tensions that emerged from their use cases and to provide some reflection on the particularities of their national context or their use case that may have struck them during the research process. Below, we present each international partner' s reflections and conclude with a synthesis of the similarities and dissimilarities observed between the cases. The reflections on the different tensions observed in each use case and how the international partners aim to respond to and ease these tensions. They provide practical guidance by showing how different contextual technical and organisational measures can be designed to promote ethical AI development and deployment. As highlighted in Section 4.5, all use cases provide precious guidance on how to minimize the data that are use or collected by the systems, and on how to protect the interests of vulnerable groups potentially affected by the deployment of these systems.

4.1. JAPANESE CONTEXT

For UC7, the research team members at the University of Osaka identified three main tensions between the principles or components:

- **Tension 1: Non-discrimination and privacy**

  Managers may argue that access to certain sensitive data (e.g., race/ethnicity) is necessary to accurately interpret AI-generated outputs and assess fairness in outcomes. However, granting such access raises serious concerns, including the risk of discrimination and the infringement of workers' privacy.

- **Tension 2: Accuracy and privacy**

  Unimodal emotion recognition systems, which typically rely on facial information or gestures, are often inaccurate because they fail to account for the complexity of emotion as a phenomenon and the impact of contextual factors. Therefore, if one aims to infer more robust and reliable affective states, one proposition has been to integrate additional personal and sensitive data, such as speech, facial expressions, and physiological signals. This approach is known as multimodal emotion recognition (MER). However, because MER relies

on a broader range of sensitive data, it also carries increased risks of privacy intrusion and data breaches.

- **Tension 3: Transparency and chilling effects / strategic behavioral responses of workers**

  Efforts to enhance transparency—by clearly informing employees that emotion recognition technology is being used—may generate a chilling effect on workers' behaviour. At the same time, such transparency may unintentionally incentivize strategic behavioral responses, which could in turn undermine or distort the reliability of the system's outputs.

A central concern that emerges from this use case is that stakeholders (managers and workers) have different interests that pull in distinct directions concerning data collection and the accuracy of the system itself ought to be balanced with the privacy interests of the workers. These tensions can be addressed by different technical and organizational measures.

Technical measures can restrict data collection strictly to purpose-relevant data and define no-collection periods (e.g., during breaks). Implementation of visual abstraction techniques in biometric authentication, use of data aggregation, and prohibition of raw data storage can also contribute to the protection of the privacy of the workers.

On the organizational side, managers should ensure demonstrable and proportionate benefits for workers. Safeguards should be put in place to ensure that efficiency gains generated by AI deployment do not translate into increased work intensity, but are instead redistributed to workers in the form of additional rest periods, health support, or comparable benefits. Where proportionate benefits cannot be assured, the possibility of stopping implementation should be considered. Finally, transparency, in the form of clear communication mechanisms regarding what information is collected, opportunities for informed consent, and establishment of institutional governance mechanisms (e.g., a dedicated oversight body responsible for ongoing monitoring and ex-post evaluation) can be implemented to ease the identified tensions.

As for the contextual specificities, it is important to note that while the *EU AI Act* (2024) designates emotion recognition technology in the workplace as prohibited, in Japan

there is currently no specific regulation targeting emotion recognition technology, and major electronics companies are actively developing and implementing such systems.

4.2. CHINESE CONTEXT

The research team at CASTED identified four main ethical tensions emerging from the deployment of the deployment of the smart companion robot "Datou Aliang" (Institutional Version):

- **Tension 1: Strategic liability avoidance and substantive safety responsibility**
  While developers define smart elderly care robots as "non-medical grade" to mitigate legal risks, the operational reality in nursing homes like Wuxi Outang shows a high reliance on the system for life safety, creating a responsibility gap. Furthermore, there is a technical-clinical tension between the need for high sensitivity to prevent "missed reports" (SOTIF deficits) and the resulting "alert fatigue" that can cause staff to ignore genuine emergencies.

- **Tension 2: Life safety and privacy**
  In the Chinese context, adult children acting as *de facto* elderly care decision-makers often prioritize older family members' physical safety over their personal data privacy, and frequently request 24-hour smart monitoring as a way to practice filial piety (孝xiao, a core traditional Chinese value centered on reciprocal care and support for aging parents). As a result, a distinct value contradiction has emerged between algorithmic monitoring arrangements implemented out of filial care duty under the collective family logic, and the individual senior's rights to privacy and independent decision-making.

- **Tension 3: Inclusivity and commercial feasibility**
  Providing fair access to seniors who speak the heavy Wuxi dialect is essential to prevent "Digital Aphasia," yet the high financial cost of building localized corpora makes it difficult for small developers to force market-leading vendors to prioritize niche dialects. This could cause market failure in the smart eldercare space, leaving the most vulnerable seniors—those who only speak local traditional languages and have limited digital skills—at risk of missing out on accessible smart eldercare benefits.

- **Tension 4: Proactive algorithmic intervention and human clinical agency**
  Developers often seek to maximize robot engagement and automated alerts to ensure safety, but clinical professionals warn that over-reliance on "Critical Alert Data" leads to the "deskilling" of caregivers. These risks turning highly trained nurses into passive responders to machine prompts rather than active analytical professionals capable of performing humanized "root cause analysis."

The UC8 research team has identified different technical measures to respond to these tensions.

For relevant technical measures, re-training the fall detection model using thousands of localized "negative events" recorded at the nursing home (e.g., falling bags, medical carts, wheelchair movements) to refine the machine's ability to distinguish between inanimate objects and humans, and integrating heterogeneous sensors so the system uses vision for impact detection and radar point-cloud data to monitor sustained posture changes and stillness, providing a redundant safety check could mitigate tension 1. Concerning privacy issues, it would be possible to deploy radar sensors that capture posture silhouettes via point-cloud data rather than high-resolution video, identifying a "person on the floor" without exposing facial or physical details. Moreover, one could process audio locally on the device to listen for specific wake-words or emergency cries ("Help!", "Help me!") and ensure that cloud transmission is only activated as a post-trigger. To address issues of inclusivity from a technical standpoint, it is possible to apply "Secondary Constraints" on top of third-party foundation models by integrating localized speech engines trained on regional Wuxi/Suzhou accents to ensure that the most vulnerable seniors—those who speak only local dialects—are not excluded from life-saving technology and emotional companionship due to a Mandarin-centric dataset. Finally, to preserve expertise and avoid automation bias, the UC8 research team found that one could label AI alerts as "Critical Alert Data" (including confidence levels) in the UI and force a mandatory "Root Cause Analysis" (RCA) input field that must be filled by the nurse before an alert can be cleared.

In addition to these technical measures, the research team identified complementary organizational measures that can be implemented to try and reconcile the conflicting principles. They highlight that policymakers should integrate a regulatory requirement prioritizing the deployment of safety-critical AI in professionally managed institutional environments (B2B) before allowing access to the unmanaged home market (B2C) into

market access certification. Authorities should also implement a mandatory "dual-track" operation phase (standardized at 90 days) where manual paper logs are maintained alongside AI-automated logs to verify system efficacy and build clinical confidence. This should foster evidence-based trust to ensure that the technology' s reliability is proven *in situ*. In parallel, integrating "Dialect Inclusivity" and "Linguistic Accessibility" as mandatory Key Performance Indicators (KPIs) in public social service procurement contracts by updating the "Smart Eldercare Procurement Catalog" to include mandatory scoring categories for localized dialect engines (e.g., Wuxi/Suzhou accents) and specialized cognitive impairment interaction filters would address the risk of digital aphasia. The UC8 research team underlines that a minimum recognition threshold for local dialects should be required for any government-funded project.

Finally, to prevent deskilling public authorities could adopt a proactive legal and operational strategy defining the robot as an "Auxiliary Tool" and mandating Standard Operating Procedures (SOPs) for human verification of AI signals. Policy should mandate that service contracts and user manuals include a "Supportive Tool Status" clause. This must be coupled with mandatory institutional SOPs requiring a nurse to sign off on a "Bedside Verification" for every "Critical Alert Data" event before the alert can be closed in the digital audit trail. Moreover, requiring tamper-proof, timestamped logs for all system alerts, human interventions, and overrides, modeled after the PIPL requirements and *EU AI Act* (Art. 12, serves as a technical reference) could ensure traceability. In the event of an adverse incident, policymakers and legal bodies need "Black Box" transparency to distinguish between a hardware malfunction, an algorithmic bias, or human operational negligence. This can be done by mandating a 365-day retention period for encrypted interaction and alert logs as a condition for institutional operating permits. These logs must be made available for impartial forensic review by regulatory bodies or during insurance claims and medical disputes.

In addition to the aforementioned ethical tensions, the UC8 research team identified one context-specific cross-cultural difference reflected across their case studies. Specifically, they observed a divergent priority in value framing in the Chinese elderly care context: the traditional practice of xiao (“孝道” filial piety, a core Chinese value centered on reciprocal care for aging family members), which often prioritizes older adults' life safety in high-stakes care scenarios over their personal data privacy, aligns with an algorithm-enabled family care decision-making framework rooted in China's

family-centric care norms.. This context-specific value priority presents a notable tension with the Western normative emphasis on individual self-determination as a primary default principle.

4.3. SOUTH KOREAN CONTEXT

The research team at STEPI found that despite the potential benefits, six potential tensions emerge between the principles and values implicated in this use case.

- **Tension 1: Safety from overdependence, deception, and misjudgments, and functional excellence for meaningful conversation**

  Conversational care robots need the capability to understand users' intentions, needs, and emotional states and respond appropriately within context to engage in meaningful interaction with users. Such capabilities are important for improving the quality of conversations and relationship formation, but they may simultaneously strengthen intimacy with users and thereby increase the risk of excessive trust or dependence on the system. For vulnerable users in particular, such relationships may lead to misjudgment or inappropriate decision-making. Therefore, conversational AI in care contexts must ensure functional excellence that enables meaningful interaction while also incorporating appropriate design features and safeguards to prevent over-dependence or misunderstanding (Sharkey & Sharkey, 2012; OECD, 2019).

- **Tension 2: Personalization and efficiency**

  Personalization for user-tailored services is presented as an important design objective because it can improve the appropriateness and efficiency of care by reflecting older adults' health conditions, life contexts, and preferences. However, such personalization simultaneously carries the risk of reinforcing an instrumental approach that reduces older persons to objects of care management or data-driven administration. In this process, the complex life contexts and relational dimensions of older adults may be reduced, and as care becomes organized primarily around efficiency and manageability, older persons may be objectified as subjects of "care management." Moreover, as some aspects of care judgment and interaction are transferred to the system, there is also the possibility of deskilling, whereby the professional judgment and relational care capacities of care workers may be weakened. Consequently, the pursuit of

efficiency through personalization creates an ethical tension, as it may undermine older persons' agency, the quality-of-care relationships, and the professional expertise of care labour.

- **Tension 3: Autonomy and protection for care**

  In the context of elderly care, autonomy refers to respecting an individual's right to make decisions about their own life and everyday choices. However, in care situations, protective interventions may be necessary to ensure safety and well-being, particularly in cases involving cognitive or physical vulnerability. While such protective measures serve legitimate purposes of risk prevention and the fulfillment of care responsibilities, they may simultaneously restrict individuals' choices and actions, thereby potentially weakening autonomy. Conversely, if autonomy is emphasized excessively, necessary interventions may be delayed or absent, creating risks of neglect (Hall et al., 2019). Therefore, in the design and operation of care AI, a persistent ethical tension arises regarding how to respect older persons' self-determination while establishing an appropriate level of protection for safety and well-being.

- **Tension 4: Autonomy and self-efficacy of seniors**

  In elderly care, autonomy refers to respecting the individual's right to make decisions about their own daily life and choices, whereas self-efficacy concerns the individual's perception and capability to perform specific tasks or handle particular situations. When care technologies or support systems intervene actively for the sake of safety and convenience, the range of available choices may formally remain intact, yet opportunities for everyday judgment or action may decrease, thereby weakening self-efficacy. Conversely, minimizing intervention in order to preserve self-efficacy may risk insufficient support in situations involving cognitive or physical vulnerability. Therefore, in designing and operating AI systems for care, it is important not only to guarantee older persons' freedom of choices but also to maintain opportunities for participation in everyday activities and decision-making so that autonomy and self-efficacy are not simultaneously diminished.

- **Tension 5: Explicit allocation of responsibility across human and AI actors and flexible and adaptive decision-making in care**

When cooperation between human providers and AI systems is assumed in care services, clearly defining the roles and responsibilities of each actor is important for ensuring transparency and accountability in decision-making. However, real care situations often require responses to unpredictable changes and individualized contexts, which calls for complementary and flexible decision-making between human service providers and AI systems. If responsibility is allocated too rigidly, such collaborative and adaptive judgment may be constrained. Conversely, if flexibility is emphasized excessively, the boundaries of responsibility between humans and AI may become unclear. Therefore, in designing and operating care AI systems, it is necessary to seek a balance that assumes a cooperative environment between humans and AI, clarifies the basic structure of responsibility, and still allows for contextual judgment and role adjustment.

- **Tension 6: Making invisible work visible and accountable, and maintaining affordable and care service operation**

  In care services, making invisible labour visible and formally recognizing it is important for ensuring accountability and fair compensation for activities actually performed in the care process, such as monitoring, situational judgment, and emotional support. However, institutionally recognizing such labour and incorporating it into systems of documentation and management may generate additional burdens in terms of personnel, time, and costs, potentially affecting the economic sustainability and efficiency of service operations. Conversely, if cost reduction and operational efficiency are prioritized, such labour may continue to remain informal and invisible, thereby weakening accountability and fairness. Therefore, care services must seek a balance between the need to recognize the value of invisible labour and strengthen accountability, and the economic constraints associated with maintaining sustainable service operations.

Different technical and organizational measures can be put in place to mitigate these tensions.

To protect vulnerable older adults who experience a decline in cognitive capacity and memory, one could design functional restrictions for a safety-restricted mode in a graduate manner. This should avoid the extremes of "complete shutdown" or "unrestricted use," in order to maintain a balanced relationship between protection, autonomy, and responsibility. Similarly, technical measures could be put in place to

prevent excessive inferences and protect vulnerable users. There is a possibility that the AI's inferences do not align with the older adult's actual situation or characteristics, yet, because the AI is perceived as providing a "care" function, the user may accept them and make misjudgements. To address this issue, AI systems should be designed so that uncertain inferences about sensitive attributes that older users have not provided are not treated as definitive or used as the sole basis for major decision-making when making suggestions or interventions that may affect an older person's health, safety, or behaviour. Where necessary, such inferences may be supplemented through user confirmation or cross-verification with additional information. Furthermore, real-time inference data used for functional adaptation should be processed only to the minimum extent necessary for service provision, and external transmission or long-term storage of such data should be minimized.

On the organizational side, mitigation of data bias can be done by the implementation of a stakeholder audit system. By establishing multi-stakeholder participation structure and incorporating it into public procurement requirements, it should be possible to incorporate perspectives beyond those of developers — particularly those of seniors, service providers, and experts from diverse fields — which is critical for effectively reducing paternalistic expressions and unconscious bias. In parallel, adaptive governance through experimental-rule setting could be implemented to address these tensions. This could be done by operating a lifecycle that includes monitoring, information sharing, standard-setting, evaluation, empirical testing, the revision of standards, and real-world deployment through an iterative process. This approach would simultaneously encourage innovation and ensure early risk detection and preventive interventions.

As for the contextual specificities, the UC9 research team found that in designing ethical guidelines for conversational AI systems for elder care in Korea, it is essential to move beyond abstract declarations of principles and instead formulate normative requirements that can operate effectively in real-world settings. This requires careful consideration of Korea's institutional, industrial, research, and care-practice environments. Key elements shaping the Korean context of development and use of conversational AI for elder care can be examined across technological, industrial, socio-cultural, and institutional dimensions.

- **Technological Dimension**

From a technological perspective, a key challenge lies in the insufficient accumulation of data suitable for training models used in elder care services. In Korea, foundational research traditions concerning older adults' language use, behaviour, and cognitive characteristics are relatively recent, and the availability of relevant datasets remains limited. As a result, when AI systems attempt personalization or emotional and state inference, they are more likely to rely on incomplete data and constrained contextual understanding. This, in turn, increases the risk of user misjudgments, emotional instability, or inappropriate system responses.

- **Industrial Dimension**

  From an industrial standpoint, most AI systems for elder care in Korea are developed by small and medium-sized enterprises (SMEs). As is typical of many Korean SMEs, these companies face constraints in both specialized human resources and financial capacity. The majority of R&D resources are sourced from government-funded national research programs, and deployment is heavily dependent on pilot or dissemination projects led by central or local governments. In this sense, a fully functioning commercial business ecosystem has yet to emerge. In the absence of a clear "best practice" model for AI system development, companies may at times adopt conservative approaches to controlling potential risks, while at other times making bold decisions to accommodate the demands or complaints of purchasers and end users. This situation yields two key implications for the formulation of ethical principles. First, imposing overly strict compliance requirements on developers may inadvertently produce adverse effects, such as circumvention or irregular operational practices. Second, ethical AI development requires not only responsible behaviour from suppliers (developers), but also clear normative expectations and demands from purchasers—primarily government entities. At present, the supply of AI robots is limited to a small number of firms, while purchasers (government bodies) are structurally dependent on these suppliers. Purchasers must therefore articulate clear demands to ensure that technology is supplied at a satisfactory level of quality. Failing this, firms may lose incentives to invest in improved robot development, and both suppliers and purchasers risk becoming locked into specific firms or models (Suh et al., 2025), ultimately leading to inefficient use of social resources.

- **Socio-Cultural Dimension**

  At the socio-cultural level, particular attention should be paid to the high level

of trust that older adults in Korea tend to place in government-led new technology development and deployment programs. Korea's industrial growth was strongly shaped by government-led catch-up strategies in the 1960s and 1970s, and many of today's older adults experienced their formative years during this period. As a result, they often hold strong trust in government research and development initiatives. Because most elder care robots in Korea are introduced through government-led pilot programs and framed as part of "research," users (older adults) tend to adopt a receptive stance. Even when they experience personal discomfort, they may readily consent to the provision of personal data for the sake of the "public good" and tolerate inconvenience caused by system errors. From the perspective of AI development and use norms, this cultural characteristic suggests that data collection from older adults requires heightened ethical caution beyond standard personal data consent procedures. Providing personal data to AI robot developers may be understood not only as data sharing for service use, but also as a form of contribution to AI research and development, or as part of a welfare administrative process. Accordingly, in the context of AI robots for elder care, personal data issues should be addressed not merely as a matter of refining consent procedures, but at the level of establishing principled criteria regarding the extent to which information sharing can be legitimately requested in the interests of research and social well-being.

- **Institutional Dimension**

  From an institutional perspective, analysis should distinguish between Korea's elder care service delivery framework and its AI regulatory framework. With respect to care services, conversational AI robots are primarily deployed to older adults who are economically and socially vulnerable. This reflects the design of Korea's elder care policies, which prioritize support for disadvantaged populations. Consequently, current users are often emotionally fragile and particularly susceptible to the potential risks associated with AI systems. In discussing the Korean context of AI robot deployment in elder care, the labour conditions of service providers must also be taken into account. In the local government's pilot projects, the number of service recipients assigned to each care worker has remained unchanged even after the adaptation of AI robots. If a critical reflection of this rationality of such labour structure and adequate control over the factors that generate "shadow labour" are not undertaken,

service providers may increasingly rely on mechanical responses, which ultimately creates the risk of their deskilling.

In practice, this has resulted in an increase in overall work burden. Without critical reflection on the rationality of such labour structures and adequate control of factors that generate "shadow labour," there is a risk that service providers may resort to mechanical responses or experience an erosion of care-related skills.

Turning to AI governance, the *Act on the Promotion of the Artificial Intelligence Industry and the Establishment of a Foundation for Trust* was promulgated in January 2025 and came fully into force in 2026, thereby establishing the legal foundation for AI governance in Korea. Rather than directly prescribing detailed technical regulations, this law has the character of a framework law, providing the basic direction for AI policy and governance. The core objective of the law is to achieve a balance between innovation and protection, promoting innovation in the AI industry while simultaneously ensuring social trust and safety. Reflecting this objective, the government seeks to apply stricter management and protective measures to high-impact AI systems, while also establishing a collaborative governance model in which the private sector takes the lead in managing risks and the government provides institutional support.

The government has also been working to build a regulatory infrastructure, including the development of standards and guidelines, in order to ensure the effective implementation of the law. In 2020, the *National AI Ethics Guidelines* were announced by the Ministry of Science and ICT (MSIT). However, further research is currently underway to establish reliability verification standards and ethical guidelines that reflect the specific characteristics of the diverse sectors in which AI technologies are applied (KISDI, 2025; TTA, 2025).

Conversational AI systems for seniors do not fall under the category of high-impact AI defined by the Korean *AI Basic Act* and therefore are not subject to legal obligations such as reliability assurance measures, user notification requirements, or the establishment of formal risk management systems. Nevertheless, as AI systems introduced into public services, they are expected to have a considerable impact on the provision of social services. For this reason, scholars and civil society actors have argued that technical and organizational measures to ensure trustworthiness should still be

implemented, even in the absence of formal legal obligations (Kim & Shin, 2025; Suh, 2025; Ham, 2025). Companies, government actors, and older users involved in elder care must accumulate practical experience and learn through trial and error to realize safe and trustworthy AI systems. Given Korea' s industrial structure, characterized by a high proportion of SMEs, and the reality that AI robots are disseminated through public service delivery systems, ethical guidelines should be designed not as uniform and static rules, but as an adaptive governance framework capable of adjustment in response to technological and practical change.

4.4. CANADIAN CONTEXT

The research team at McGill has zoned in on three main ethical tensions emerging from the deployment and development of griefbots.

- **Tension 1: Protecting postmortem interests over time and users/commercial interests**

  The development and deployment of griefbots raises the question of how to include and protect the rights and interests of the deceased effectively over time, given that they may conflict with the interests of both the users and the private company developing and maintaining the griefbots. Given that the deceased can no longer defend their own perspective on how their data should be used or how their image should be protected, it requires creating clear guidelines on how to proceed. Relying exclusively on the wishes of the family of the deceased is also insufficient given that family members can disagree strongly concerning how best to respect the interests and wishes of the deceased. Moreover, private companies have structural incentives to exploit the data of the deceased to maximize profits in ways which may conflict with the wishes of the deceased.

- **Tension 2: User wellbeing and autonomy**

  Given that grief is a very personal process which can vary from person to person, one should be careful not to enforce one conception of what grief ought to look like when evaluating the impact of griefbots on individuals. Additionally, as some individuals may wish to use the griefbots as a means to retain a certain connection with the deceased despite known risks and potential negative impacts on their wellbeing, one should have sufficient justification to override

the autonomous decisions of the users to use griefbots in some ways to avoid abusive paternalistic behaviour.

- **Tension 3: Effectiveness and privacy**

  Effectively monitoring the wellbeing of users may require gathering sensitive and personal data (conversational data, types of interactions, etc.). This conflict with the privacy rights of the users considering that the conversational data and types of interaction with deceased loved ones is highly likely to be sensitive information.

The research team of UC10 has identified different technical and organizational measures to ease these tensions.

On the technical side, to protect the postmortem interests of the deceased, the research team has underlined the necessity to implement output/export controls to limit what users can share outside of the platform. The users should not be able to export or share conversational data or other generative AI outputs (voice recordings, video, etc.) in ways that would go against the wishes of the deceased, or which present the deceased in ways that are demeaning, exploitative, insulting, etc. Similarly, to protect the wellbeing of the users all the while respecting their autonomy to use the griefbot if they wish, the griefbot should be designed in ways that make it explicit that the bot is not, in fact, a reincarnation of the deceased. It should be designed to remind the users that griefbots are programs that do not have feelings or consciousness. It should be clear through interactions that postmortem avatars are programs imitating a person, not the reincarnation of the person herself. This can be achieved through disclosures always present on the screen, periodical reminders after a given amount of time, or by creating a type of uncanniness by design by ensuring that the griefbots refers to the deceased only in the past tense, or in the third person and using conditionals.

In parallel, organizational measures can be implemented to ease the tension between both respecting postmortem interests and the autonomy of users. First, informed consent before death should be secured for commercial griefbots. Individuals should consent to the use of their personal and private data for the creation of a griefbot. Consent should be given on the type of data that will be used to develop the chatbot and the purposes for which it will be used (to create a text-based interface, a virtual avatar, for how long, etc.). Second, regulations should establish clear retirement

protocols for retirement and deletion of personal and conversational data after prolonged inactivity, unless there is renewed interest expressed by the user. This is necessary to recognize that some individuals can go through periods where they may want to interact with the griefbot (given changes in their lives, for instance) even if they do not interact with the bot on a regular basis.

The UC10 research team has also highlighted three contextual specificities concerning the deployment of griefbots in Canada. First, there is currently no specific Canadian federal legislation directly and clearly regulating griefbots or related posthumous AI applications. However, various laws intersect with this technology, which can vary depending on the jurisdiction. For instance, the *Personal Information Protection and Electronic Documents Acts* (PIPEDA) does not explicitly address posthumous use of data, but defines personal information as "information about an identifiable individual," which does not include deceased individuals. It is consequently unclear if the data collected during the deceased' s lifetime may still fall under PIPEDA protections. Moreover, some provincial laws like Ontario' s *Succession Law Reform Act* or privacy laws in British Columbia and Alberta may apply. It consequently seems relevant and urgent to develop a more principled, uniform response to address how private entities can use the data of the dead in Canada.

Second, operating in Canada and Québec requires not only offering the service in English or French, but raises the question of how to approach indigenous languages as low-resource languages (including Cree, Atikamekw, Mi' kmaq, Inuktun, etc.). Additionally, developing projects using the data of members of Indigenous nations in Canada involves its own set of guidelines and requirements (for reference, see Assembly of First Nations Quebec-Labrador 2025).

Thirdly, the development and deployment of griefbots require paying special attention to how LLMs interact with culture through language. In a multicultural society like Canada, views on griefbots and on what constitutes a healthy grieving process may vary across cultures and conceptions of the good life. Development of griefbots should take into account that grief and the grieving process have cultural dimensions that should be included in the design and deployment of the system.

4.5. CROSS-CUTTING REFLECTIONS

While most of the international cases clustered around the impact of AI on cognition and private behaviour (UC8, UC9, UC10) and only the University of Osaka considered a use case on the impact of AI on human expertise and professional behaviour, it is interesting to note that common tensions and risks emerge from all the use cases. All four use cases underline potential risks and tensions surrounding privacy, which ought to be balanced against accuracy, effectiveness, or safety, depending on the case. They all include technical measures aimed at limiting the type of data which can be collected by the AI systems or the types of inferences that can be made by the system to protect individual privacy rights.

In addition to privacy concerns, all four cases underline the importance of putting new and innovative regulatory policies or organizational measures in place to protect vulnerable users. Vulnerable groups may include senior citizens or even the deceased. The measures span from audits, new regulations clarifying how it is permissible to use data, to establishing oversight bodies and regulations putting conditions on market access. UC8, UC9, and UC10 also consider the impact of social AI systems designed to provide companionship to vulnerable individuals. Despite expected variations due to the details of each case, they all converge on the ideas that technical and organizational measures should be implemented to limit overdependence and harm to vulnerable users.

Importantly, despite variations between the use cases, it is instructive to consider how the practical measures point in similar directions and provide practical guidance on how to aim for ethical AI development and deployment in practice. Although general guidelines will be developed more thoroughly in *T3.3*, we already see that all use cases underline the importance of limiting the data the AI systems have access to (UC7; UC8), the inferences they should be allowed to trace and use (UC9), or to ensure that relevant actors consent to the use of their personal data (UC10). This seems to rely on the intuition that the data used by the systems should be minimized to ensure that it is strictly necessary to aim for a socially valuable goal.

All the use cases also converge on the idea that we should protect the interests of vulnerable users that are affected by the deployment of these systems. For UC7, this means that the deployment of emotion recognition technology should be justifiable to the workers affected by ensuring demonstrable and proportionate benefits for them. For UC8, this means putting special importance on the interests of vulnerable patients,

such as those who have the regional Wuxi/Suzhou accent. Similarly, UC9 underlined the necessity of considering the impact of the carebots on older adults experiencing a decline in cognitive capacity and memory. Finally, UC10 highlighted the importance of protecting the long-term interests of the deceased in maintaining a certain image over time and ensuring that vulnerable greavers are not exploited or harmed by these systems.

Finally, as expected, all use cases highlight specificities in their national contexts which ought to be considered to regulate the effective development and deployment of emerging AI technologies. These range from Japan that does not explicitly regulate the uses of emotion recognition technologies in the workplace which would typically be prohibited under the *EU AI Act* (2024), China that highlights a different cultural approach to the question of alignment, Korea which underlines its specific technological, industrial, socio-cultural, and institutional realities, to Canada, which insist on its linguistic pluralism and the need to harmonize federal and provincial regulations. This shows how aiming for ethical AI systems needs to consider both high-level ethics principles, components, and practical, specific socio-political contexts to operationalize ethics guidelines into actionable technical and organizational measures.

5. Conclusion

In conclusion, the present report on *Operational context-sensitive guidelines in Canada, China, Japan, and South Korea* not only offers a diverse set of technical and organizational measures to approach practical uses cases on AI development and deployment, but it also illustrates the strength and productiveness of a multidisciplinary, bottom-up approach to operationalize high-level ethics principles and the fundamental importance of considering the specificities of each national context.

As discussed, the research team from the University of Osaka has focused on the deployment of emotional AI systems in the workplace, which is allowed in its jurisdiction. This contains precious insights to evaluate the risks and implications of these systems as they are developed and deployed in practice. CASTED and STEPI have both approached the question of how to design ethical AI systems to care for older adults. The cases of Chine and Korea cover examples of AI-based social services designed to address population aging. China focuses on the introduction of smart care systems in nursing facilities, while the Korean case focuses on AI-human interactions that may occur when AI is applied to in-home social services. Despite clear resonance between the cases, as they both highlight the importance of protecting user privacy and deploying different technical and organizational measures to support effective care of older adults, important variances between their national contexts emerge. The Chinese team has underlined a distinct way in which they typically balance safety and privacy, and the Korean team has highlighted how older adults have a tendency to strongly trust government research projects, thus justifying heightened ethical caution beyond standard personal data consent procedures. Finally, the Canadian research team underlined the need, in its context, to develop clearer institutional measures to regulate the data of the dead.

This both shows how the AIOLIA research project has been productive globally to inform operationalization of ethics guidelines in a way that allowed the research teams to adapt to their own realities. This material will provide crucial material for the subsequent work in AIOLIA by contributing to the diverse portfolio of technical and organizational measures, to compare the EU with international uses cases and different regulatory contexts (WP3.3), and to develop AIOLIA training materials (WP4).

6. References

Act on the Protection of Personal Information, Act No. 57 of 2003 (Japan). https://www.japaneselawtranslation.go.jp/en/laws/view/4241

AI Act Regulation (EU) 2024/1689. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance).

Assembly of First Nations Quebec-Labrador. The Digital Territory of First Nations Quebec-Labrador. Position on Digital and Artificial Intelligence Ethics. 2025. Online:

Boada, J. P., Maestre, B. R., & others. (2021). The ethical issues of social assistive robotics: A critical review. AI & Society. ScienceDirect

Copyright Act, R.S.C. 1985, c. C-42. https://laws-lois.justice.gc.ca/eng/acts/C-42/index.html

Criminal Code, R.S.C. 1985, c. C-46. https://laws-lois.justice.gc.ca/eng/acts/c-46/

Deusdad B (2024) Ethical implications in using robots among older adults living with dementia. Front. Psychiatry 15:1436273. doi: 10.3389/fpsyt.2024.1436273

European Commission, High-Level Expert Group on Artificial Intelligence. (2019). *Ethics guidelines for trustworthy AI*. https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

Hall, A., Brown Wilson, C., Stanmore, E., & Todd, C. (2019). *Moving beyond 'safety' versus 'autonomy': The ethics of monitoring technologies in dementia care*. BMC Geriatrics, 19, 145.

High-Level Expert Group on Artificial Intelligence (AI HLEG). (2020). The Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment. https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessment

Katirai, A. (2024). Ethical considerations in emotion recognition technologies: A review of the literature. *AI and Ethics*, *4*(3), 927–948.

Kim, H. (2024). "Caring robots?": The ethics of care robots and robot care. Government Studies (Jeongbuhak Yeongu), 30(2), 31–59.

Kim, K.H and Shin, Y.K. (2025) Implications of the EU AI Act for South Korea's Social Security Systems. Global Social Security Review 2025 No.32, pp.75-89

Kubota, Alyssa and Pourebadi, Maryam and Banh, Sharon and Kim, Soyon and Riek, Laurel, Somebody That I Used to Know: The Risks of Personalizing Robots for Dementia Care (August 23, 2021). In Proceedings of We Robot, 2021, Available at SSRN: https://ssrn.com/abstract=3910089

Ministry of Civil Affairs of the People's Republic of China & Standardization Administration of the People's Republic of China. (2025). *Guidelines for the construction of elderly care*

service standard system (2025 edition) [养老服务 准体系建设指南（2025版）]. Online: https://www.gov.cn/zhengce/zhengceku/202511/content_7049494.htm

Ministry of Science and ICT. (2023, September 26). *South Korea presents The "Digital Bill of Rights," crystallizing President Yoon's digital vision, is announced as the manifesto for a universal digital order.*https://english.msit.go.kr/eng/bbs/view.do?sCode=eng&mPid=2&mId=4&bbsSeqNo=42&nttSeqNo=467

Ministry of Science and ICT, KISDI. (2020). *National strategy for artificial intelligence: Ethical guidelines for AI*. Korea Information Society Development Institute. https://ai.kisdi.re.kr/eng/main/contents.do?menuNo=500007

National New Generation Artificial Intelligence Governance Expert Committee. (2019). *Governance principles for the new generation of artificial intelligence—Developing responsible artificial intelligence* [新一代人工智能治理原则——发展负责任的人工智能]. Online: https://www.cac.gov.cn/2019-06/17/c_1124634614.htm

National New Generation Artificial Intelligence Governance Specialist Committee. (2025). *Ethical norms for new generation artificial intelligence* [新一代人工智能伦理规范]. Online: //www.most.gov.cn/kjbgz/202109/t20210926_177063.html

NEC Corporation. (2024, May 9). *NEC to Ōsaka Daigaku ELSI Sentā, kao ninshō gijutsu no tekisei riyō ni muketa gaido oyobi risuku asesumento shuhō o sakutei* [NEC and The University of Osaka ELSI Center formulate guide and risk assessment method for appropriate use of facial recognition technology]. https://jpn.nec.com/press/202405/20240509_01.html

New York Department of State. (2021). *Right of publicity*. https://dos.ny.gov/right-publicity

Office of the Privacy Commissioner of Canada. (2024, May 22). *The Personal Information Protection and Electronic Documents Act (PIPEDA)*. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/

Podgorica, N., Flatscher-Thöni, M., Deufert, D., Siebert, U., & Ganner, M. (2021). *A systematic review of ethical and legal issues in elder care*. Nursing Ethics, 28(6), 895–910.

Standardization Administration of the People's Republic of China. (2019). *Information security technology—Data security capability maturity model* [信息安全技术 数据安全能力成熟度模型] (GB/T 37973-2019). Online: https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=D16FF5DF1E14AF4D3263C0D8FED78579

Standardization Administration of the People's Republic of China. (2020). *Information security technology—Personal information security specification* [信息安全技术 个人信息安全规范] (GB/T 35273-2020). Online: https://openstd.samr.gov.cn/bzgk/gb/newGbInfo?hcno=4568F276E0F8346EB0FBA097AA0CE05E

Succession Law Reform Act, R.S.O. 1990, c. S.26.

Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. Ethics and Information Technology, 14(1), 27–40.

Suh, J., et al. (2025). Utilizing innovative technologies such as artificial intelligence to advance social services in response to a super-aged society. National Economic Advisory Council (Republic of Korea).

UNESCO. (2021, November 23). *Recommendation on the ethics of artificial intelligence*. https://unesdoc.unesco.org/ark:/48223/pf0000380455

United Nations Human Rights Council. (2024). Protection of the dead: Report of the Special Rapporteur on extrajudicial, summary or arbitrary executions (A/HRC/56/56). https://www.ohchr.org/en/documents/thematic-reports/ahrc5656-protection-dead-report-special-rapporteur-extrajudicial-summary

Université de Montréal. (2018). *Montréal Declaration for a Responsible Development of Artificial Intelligence*. https://declarationmontreal-iaresponsable.com/wp-content/uploads/2023/04/UdeM_Decl_IA-Resp_LA-Declaration-ENG_WEB_09-07-19.pdf

Yuan, S., Coghlan, S., Lederman, R. et al. Ethical Design of Social Robots in Aged Care: A Literature Review Using an Ethics of Care Perspective. Int J of Soc Robotics 15, 1637–1654 (2023). https://doi.org/10.1007/s12369-023-01053-6

Appendix A: Technical and organisational measures to guide operationalisation of AI ethics

PRACTICAL MEASURES PROVIDED BY USE CASE 7

Table 9: UC 7 – Technical measures to achieve proportionality – Adequacy

Technical measures to achieve adequacy		
Describe the measure	Bias audits	Evidence that it functions as intended
Why is it relevant?	Detects hidden performance bias	Because it involves generating evidence demonstrating the ability to achieve the given objective and present it to stakeholders.
How can it be achieved?	Startified performance reporting across sex age, scanner vendor, hospital etc.	By conducting tests that allow for a comparison of outcomes between using and not using the AI system in question.
How can be assessed whether this measure has been fulfilled?	Reporting conducted regularly; available to relevant stakeholders	Based on the results of the above tests. It is desirable to receive an evaluation from a third party, including the audit organization.
What are (potential) challenges to fulfilment?	Data availability	In addition to the technical difficulty of conducting such tests, it is unclear who bears the responsibility for conducting them.

What are risks if not fulfilled?	Potentially unreliable results or decisions	The continued use of AI systems with questionable effectiveness inevitably leads to people suffering harm.
Which are the core function/role/ stakeholders responsible?	Data collector	AI system providers and users.
Specific requirements?	Literature - Ajayi, Joaja (2025)	This component has not been explicitly addressed in such documents.

Table 10: UC 7 – Organisational measures to achieve proportionality – Adequacy

Organisational measures to achieve adequacy		
Describe the measure	Research Ethics Review Committee	Internal audits for anticipated performance
Why is it relevant?	The committee discusses adequacy	Internal audits examine whether systems, not limited to emotion recognition AI, are performing their intended functions.

How can it be achieved?	The committee members discuss the matter "from a scientific perspective"	Internal audit reviews whether the anticipated performance is likely to be achieved beforehand and whether it was achieved afterward.
How can be assessed whether this measure has been fulfilled?	If the committee members are satisfied with the explanation	If the auditor is satisfied with the explanation
What are (potential) challenges to fulfilment?	The presence of any discrepancies between previous explanations and actual performance needs to be confirmed once the AI system begins operation	The presence of any discrepancies between previous explanations and actual performance needs to be confirmed once the AI system begins operation

What are risks if not fulfilled?	The continued use of AI systems with questionable effectiveness inevitably leads to people experiencing harm	The continued use of AI systems with questionable effectiveness inevitably leads to people suffering harm.
Which are the core function/role/ stakeholders responsible?	The Chair of the Research Ethics Review Committee	The Head of Internal Audits
Specific requirements?	For life science and medical research involving human subjects, there are ethical guidelines established by the government. These may also apply to some AI systems involving human subjects, but when they do not, no other guidelines exist	Internal audits are required by company regulations

Table 11: UC 7 – Technical measures to achieve proportionality – Necessity

Technical measures to achieve necessity		
Describe the measure	Develop or find less intrusive alternatives	Implement visual abstraction techniques to preserve privacy in biometric authentication.
Why is it relevant?	A less intrusive alternative that achieves the same goal would be preferable.	This is relevant because biometric data is highly sensitive and poses significant privacy risks.
How can it be achieved?	By considering ethical principles from the very beginning of research and development.	Visual abstraction techniques mitigate these risks by minimizing the exposure and retention of raw biometric data, thereby reducing the likelihood of unauthorized access, secondary use, or re-identification.
How can be assessed whether this measure has been fulfilled?	By comparing the magnitude of risks, such as privacy risks, between different means of achieving the same goal.	By confirming that raw biometric data (e.g., facial images) is neither stored nor transmitted, and that only abstracted or processed representations are used.

What are (potential) challenges to fulfilment?	Sometimes research and development proceeds with the adoption of a specific technology as a given.	Visual abstraction techniques without degrading system performance can be technically demanding, particularly where high accuracy is required.
What are risks if not fulfilled?	Privacy violations	Privacy violations
Which are the core function/role/ stakeholders responsible?	Head of Research and Development	AI developers and system engineers are responsible for designing and implementing visual abstraction techniques within the system architecture. Management is responsible for ensuring that adequate resources, policies, and governance structures are in place to support implementation.
Specific requirements?	These were not explicitly discussed because of the expectation that any relevant laws or regulations are already being followed.	Raw biometric data (e.g., facial images) should not be stored or transmitted; instead, only abstracted or feature-based representations should be used (data minimization).

Table 12: UC 7 – Organisational measures to achieve proportionality – Necessity

Organisational measures to achieve necessity		
Describe the measure	Research Ethics Review Committee	Edge Computing
Why is it relevant?	The committee discusses necessity.	Edge computing enables data to be processed locally, thereby minimizing the transmission and centralized storage of sensitive biometric data. This significantly reduces the risk of data breaches, unauthorized access, and secondary use.
How can it be achieved?	Requires reporting on whether a comparison was made with less restrictive alternatives.	Edge computing can be achieved by designing system architectures that process data locally on user devices or on-site hardware, rather than transmitting it to centralized servers. This includes implementing on-device processing or edge nodes that extract and use only necessary features, while discarding raw data immediately after processing.

How can be assessed whether this measure has been fulfilled?	If the Research Ethics Review Committee members are satisfied with the explanation.	System architecture and data flow analyses can confirm that sensitive data is processed locally and not transmitted to centralized servers.
What are (potential) challenges to fulfilment?	Committee members may be unable to think of less intrusive alternatives and thus miss the opportunity to propose them.	Implementing edge computing may be technically demanding, as local devices often have limited computational capacity, which can constrain model performance and accuracy.
What are risks if not fulfilled?	The continued use of AI systems with questionable effectiveness inevitably leads to people experiencing harm.	If this measure is not fulfilled, sensitive biometric data may need to be transmitted to and processed in centralized systems, increasing the risk of data breaches, unauthorized access, and secondary use.
Which are the core function/role/ stakeholders responsible?	The Chair of the Research Ethics Review Committee	AI developers and system engineers are responsible for designing and implementing edge-based architectures that ensure local data processing. IT and infrastructure teams are responsible for deploying

		and maintaining edge devices, as well as ensuring system security and updates.
Specific requirements?	These were not explicitly discussed because of the expectation that any relevant laws or regulations are already being followed.	Raw data should not be stored beyond immediate processing, and only minimal, non-identifiable data should be retained where required.

Table 13: UC 7 – Technical measures to achieve proportionality – Proportionality stricto sensu

Technical measures to achieve proportionality strictu sensu		
Describe the measure	Risk-benefit tests	Data collection should be strictly limited to what is necessary for the specified purpose, in order to minimize the impact on individuals' privacy and reduce the intrusiveness of data processing. This includes clearly defining "no-collection periods" and ensuring transparency regarding the placement and scope of monitoring devices.
Why is it relevant?	They demonstrate that the benefits outweigh any risks, after risks are assessed.	By limiting data collection to what is strictly necessary, and by introducing "no-collection periods" and transparency regarding monitoring devices, this measure reduces the burden imposed on individuals and helps ensure that the overall impact remains proportionate to its intended benefits.

How can it be achieved?	By assessing both risks and benefits as objectively as possible.	"No-collection periods" should be formally established and enforced within the system, ensuring that data collection is automatically suspended during designated times (e.g., breaks or non-working periods). And transparency mechanisms should be introduced, including clear communication and visual indicators regarding the placement and operation of monitoring devices.
How can be assessed whether this measure has been fulfilled?	By determining whether third parties, including stakeholders, find the assessment convincing.	System logs and configuration settings can be examined to confirm that data collection is limited to predefined parameters and that "no-collection periods" are properly enforced. Additionally, transparency measures can be assessed by checking whether information about the placement and operation of monitoring devices is clearly communicated and accessible to workers.

What are (potential) challenges to fulfilment?	The process of thoroughly identifying all potential risks is complex and requires experience and expertise	It may be difficult to precisely define what data is strictly necessary for the specified purpose, particularly in complex or evolving use cases. And ensuring meaningful transparency may be difficult, as workers may not fully understand how monitoring systems operate, even when information is provided.
What are risks if not fulfilled?	AI systems with higher risks may be adopted, leading to some people experiencing harm as a result.	If this measure is not fulfilled, data collection may become excessive or continuous, increasing the intrusiveness of monitoring and exposing individuals to heightened privacy risks.
Which are the core function/role/ stakeholders responsible?	The person in charge of AI system adoption.	AI developers and system engineers are responsible for implementing technical constraints that restrict data collection to predefined parameters. Data governance or data management teams are responsible for defining what data is necessary for the specified purpose and establishing data collection policies.

Specific requirements?	These were not explicitly discussed because of the expectation that any relevant laws or regulations are already being followed.	These were not explicitly discussed because of the expectation that any relevant laws or regulations are already being followed.

Table 14: UC 7 – Organisational measures to achieve proportionality – proportionality stricto sensu

Organisational measures to achieve proportionality strictu sensu		
Describe the measure	Research Ethics Review Committee	Ensure demonstrable and proportionate benefits for workers to prevent declining job satisfaction due to workplace surveillance.
Why is it relevant?	The committee discusses proportionality	This is relevant because workplace surveillance technologies may generate significant negative impacts on workers, including reduced job satisfaction, increased stress, and chilling effects on behavior. Under the component of proportionality stricto sensu, such impacts must be balanced against the expected benefits of the system.
How can it be achieved?	By reporting on whether benefits outweigh risks of the AI system.	This can be achieved by establishing clear organizational safeguards to ensure that efficiency gains generated by AI deployment are redistributed to workers. This

		includes defining measurable indicators of efficiency gains (e.g., time saved or productivity improvements), and linking these gains to concrete benefits for workers, such as additional rest periods, reduced workload, or health support.
How can be assessed whether this measure has been fulfilled?	If the committee members are satisfied with the explanation.	Measurable indicators should be used to verify whether efficiency gains (e.g., time savings or productivity improvements) have been achieved and whether corresponding benefits have been delivered to workers, such as reduced workload, additional rest periods, or health support. In addition, organizational records and policies should be reviewed to confirm that redistribution mechanisms are formally defined and consistently applied.

What are (potential) challenges to fulfilment?	It is difficult to objectively compare risks and benefits composed of diverse metrics.	It may be difficult to accurately measure efficiency gains attributable to AI deployment, particularly in complex or collaborative work environments. Moreover, translating such gains into tangible benefits for workers may be challenging, as organizations may lack clear mechanisms or face competing priorities, such as cost reduction or productivity maximization.
What are risks if not fulfilled?	AI systems with greater risks are being adopted, leading to some people experiencing harm.	If this measure is not fulfilled, efficiency gains generated by AI deployment may not be translated into tangible benefits for workers, leading to increased work intensity, reduced job satisfaction, and heightened stress. This may result in a deterioration of working conditions and exacerbate feelings of surveillance and loss of autonomy.
Which are the core function/role/ stakeholders responsible?	The Chair of the Research Ethics Review Committee	Management is responsible for ensuring that efficiency gains generated by AI deployment are translated into tangible benefits for workers and for establishing the necessary policies and governance structures. In particular, human resources (HR) departments are

		responsible for designing and implementing measures such as workload adjustments, additional rest periods, and well-being support.
Specific requirements?	"Ethical Guidelines for Medical and Biological Research Involving Human Subjects"* explicitly addresses the need for "balancing the protection of research subjects' rights and interests with the outcomes obtained through research."	These were not explicitly discussed because of the expectation that any relevant laws or regulations are already being followed.

Table 15: UC 7 – Technical measures to achieve transparency and explainability – Safety

Technical measures to achieve safety		
Describe the measure	Quality Assurance	Third-party certification

Why is it relevant?	This process, through the formulation of quality policies, the creation of quality manuals, the implementation of quality audits, and the provision of education and training, builds a comprehensive system to ensure product quality and aims to enhance both quality and safety.	By disclosing the standards on which their assessments are based, many third-party certifications clarify in what respects the safety, reliability, and transparency of AI systems are ensured.
How can it be achieved?	This is achieved by defining specifications and standards at the product design stage, selecting raw materials, establishing and managing manufacturing processes, and then extending to inspecting finished products, handling customer complaints, investigating root causes, implementing preventive measures, and driving continuous improvement.	This can be achieved achieved through clear standards, independent assessment, and transparent disclosure of the evaluation process.
How can be assessed whether this measure has been fulfilled?	By verifying compliance with the defined procedures and ensuring that the required documentation is properly maintained.	By verifying compliance against the defined standards, reviewing documented evidence, and conducting independent audits or inspections.

What are (potential) challenges to fulfilment?	A potential challenge is the lack of sufficient resources, such as time, funds, or trained personnel.	Challenges include high costs, resource demands, ensuring auditor independence, and adapting to changing regulations.
What are risks if not fulfilled?	If not fulfilled, there is a risk of product defects or safety issues that may compromise compliance with standards.	Risks include loss of trust, legal or regulatory penalties, security vulnerabilities, and reputational damage.
Which are the core function/role/ stakeholders responsible?	The head of the quality control department	The certifying company
Specific requirements?	No	No

Table 16: UC 7 – Organisational measures to achieve transparency and explainability – Safety

Organisational measures to achieve safety			
Describe the measure	Technology Assessment (ethics risk assessment table)	Providing targeted training for sales representatives and engineers	Aggregating data to higher-level (e.g., group-level) representations in order to reduce the risk of individual-level harm, such as misclassification, discrimination, or unjustified interventions.
Why is it relevant?	Check whether safety considerations were made not only for users of the AI system but also for non-users.	By equipping these professionals with the necessary knowledge and ethical guidelines, companies, together with end users, can promote the responsible and safe use of the technology.	This is relevant because individual-level data processing may lead to direct harms, such as incorrect inferences or unfair treatment of workers. By aggregating data, the system reduces the likelihood that decisions or interventions are based on potentially inaccurate or sensitive

			individual-level information, thereby enhancing safety.
How can it be achieved?	Developers fill out the TA sheet, and the secretariat checks its validity as a third party.	By providing structured training for sales and engineering staff, enabling them to raise awareness among customers.	This can be achieved by designing systems that process and output data in aggregated forms (e.g., group-level trends or statistical summaries), and by restricting access to individual-level data unless strictly necessary.
How can be assessed whether this measure has been fulfilled?	Based on how satisfied the committee members were.	By checking training completion, evaluating staff understanding, and observing improved customer awareness.	Compliance can be assessed by verifying that outputs are provided only in aggregated form, that individual-level data is not accessible for decision-making, and that appropriate safeguards are in place to prevent re-identification.

What are (potential) challenges to fulfilment?	Committee members may not necessarily understand the effectiveness of the proposed technical measures.	Limited resources, uneven staff understanding, and difficulty in conveying knowledge to customers.	Challenges include potential loss of accuracy or usefulness, the risk of re-identification when combined with other data, and organizational pressures to access individual-level data for monitoring or decision-making.
What are risks if not fulfilled?	The AI system in question may exhibit behavior not anticipated by the user, potentially leading to the infringement of fundamental rights	Misuse of the technology, loss of customer trust, and legal or reputational risks.	If not fulfilled, individual-level data may be used in ways that increase the risk of misclassification, discrimination, or intrusive interventions, thereby undermining safety and trust.
Which are the core function/role/	Developers have the responsibility to clearly explain within the Technology Assessment sheet what functions they	Technology ethics department	AI developers and data engineers are responsible for designing and implementing aggregation mechanisms, ensuring that outputs are provided in

stakeholders responsible?	intended and how they designed the system to ensure those functions.		aggregated form and that individual-level data is appropriately restricted.
Specific requirements?	No	No	Aggregation methods and criteria should be clearly documented and consistently applied. And

Table 17: UC 7 – Technical measures to achieve transparency and explainability – autonomy

Technical measures to achieve autonomy	
Describe the measure	Publishing a white paper
Why is it relevant?	A white paper serves as an official document that explains in clear terms how the system works, what kinds of data are used, what safeguards are in place, and how users' rights are protected. By making such information publicly available, the provider demonstrates accountability, enables independent scrutiny, and empowers users to make informed choices about whether and how to engage with the system.
How can it be achieved?	It can be achieved by drafting and publishing a white paper that explains the system's purpose, data use, safeguards, and is updated regularly.
How can be assessed whether this measure has been fulfilled?	It can be assessed by checking whether a white paper is publicly available, clear, regularly updated, and accessible to all relevant stakeholders.

What are (potential) challenges to fulfilment?	Potential challenges include ensuring accuracy and completeness of information, protecting sensitive data while maintaining transparency, dedicating resources for regular updates, and coordinating input from different stakeholders.
What are risks if not fulfilled?	Risks include loss of user trust, reduced accountability, potential misuse of the system, and increased likelihood of regulatory or ethical violations.
Which are the core function/role/ stakeholders responsible?	People in charge of Public Relations and Technology Ethics
Specific requirements?	No

Table 18: UC 7 – Organisational measures to achieve transparency and explainability – autonomy

Organisational measures to achieve autonomy

Describe the measure	To provide targeted training for sales representatives and engineers	Develop clear, accessible, and responsible communication practices regarding the use of workplace surveillance technologies, including how data is collected, processed, and used, in order to support workers' autonomy and informed understanding.
Why is it relevant?	By training sales and engineering staff, customers gain reliable guidance that enables them to make informed choices, protecting their autonomy.	This is relevant because a lack of clear and accessible communication about workplace surveillance technologies may undermine workers' autonomy by limiting their ability to understand, anticipate, and meaningfully respond to how such systems affect them.

How can it be achieved?	By providing structured training for sales and engineering staff, enabling them to raise awareness among customers.	This can be achieved by establishing structured and accessible communication practices regarding the use of workplace surveillance technologies. This includes clearly explaining what data is collected, how it is processed, for what purposes it is used, and what its limitations are.
How can be assessed whether this measure has been fulfilled?	By checking training completion, evaluating staff understanding, and observing improved customer awareness.	Relevant materials (e.g., policies, notices, training materials) can be reviewed to verify that clear and comprehensive information is provided.
What are (potential) challenges to fulfilment?	Limited resources, uneven staff understanding, and difficulty in conveying knowledge to customers.	Complex technical systems may be difficult to explain in a clear and accessible manner. Additionally, there is a risk that communication becomes formalistic, providing information without ensuring genuine understanding.

What are risks if not fulfilled?	Misuse of the technology, loss of customer trust, and legal or reputational risks.	If this measure is not fulfilled, workers may lack a clear understanding of how surveillance technologies affect them, undermining their ability to act autonomously. This may lead to increased uncertainty, reduced sense of control, and chilling effects on behavior.
Which are the core function/role/ stakeholders responsible?	Technology Ethics department	Management is responsible for ensuring that transparent communication is prioritized and supported through appropriate policies and resources. Human resources (HR) departments are responsible for designing and delivering communication to workers, including training and informational materials.
Specific requirements?	No	Mechanisms for interaction should be established, allowing workers to ask questions, seek clarification, and provide feedback.

Table 19: UC7 – Technical measures to achieve transparency and explainability – respect

Technical measures to achieve respect	
Describe the measure	The use of contractual agreements between providers and companies
Why is it relevant?	By embedding risk disclosures in contractual agreements, providers and companies can ensure that the deployment of emotion recognition AI respects users' dignity.
How can it be achieved?	By embedding contractual clauses that disclose risks and restrict the use of outputs—for example, ensuring they are not used directly for performance evaluations but only as reference values—so that respect for users is contractually guaranteed.
How can be assessed whether this measure has been fulfilled?	By checking whether contracts explicitly contain risk-disclosure and user-respect clauses, and by confirming in practice that these provisions—such as limiting the use of outputs to reference values—are being observed.

What are (potential) challenges to fulfilment?	Key challenges include drafting clauses that are precise yet adaptable, overcoming company resistance to restrictions.
What are risks if not fulfilled?	Without this measure, AI outputs may be misused in evaluations, leading to unfair treatment, loss of trust, reputational harm, and even legal challenges.
Which are the core function/role/ stakeholders responsible?	Head of the Legal Department
Specific requirements?	No

Table 20: UC 7 – Organisational measures to achieve transparency and explainability – respect

Organisational measures to achieve respect	
Describe the measure	Setting up a complaint desk

Why is it relevant?	Establishing a complaint or appeal desk provides users with a legitimate channel to voice concerns, ensuring they are not left without recourse in cases of harm or misunderstanding caused by emotion recognition AI, and embodying respect for them as equal participants.
How can it be achieved?	By setting up an accessible complaint mechanism with clear procedures and fair review.
How can be assessed whether this measure has been fulfilled?	By checking that the complaint desk exists, is accessible to users, and processes cases fairly and promptly.
What are (potential) challenges to fulfilment?	Challenges may include fragmented desks by product line, causing confusion for users, along with limited resources or biased handling.
What are risks if not fulfilled?	Risks include loss of user trust, unresolved harms, lack of accountability, and reputational or legal consequences.
Which are the core function/role/	Head of the Legal Department.

stakeholders responsible?	
Specific requirements?	No.

Table 21: UC 7 – Technical measures to achieve Fairness and non-discrimination – anti-bias

Technical measures to achieve anti-bias	
Describe the measure	Conducting so-called 'adversarial tests'
Why is it relevant?	By testing the system with challenging or misleading inputs, developers can detect unfair treatment of certain groups and use the findings to refine data and algorithms, reducing bias and making emotion recognition AI more robust and fairer.
How can it be achieved?	It can be achieved by incorporating adversarial testing into the robustness evaluation process, where the system is exposed to challenging or misleading inputs to reveal potential biases.
How can be assessed whether this measure has been fulfilled?	By actually inputting various parameters into the algorithm and verifying how much bias remains.
What are (potential) challenges to fulfilment?	In order to define bias, it is necessary to first establish what constitutes a fair or equitable reference situation. However, such a determination cannot be made for all parameters, as the notion of fairness may not be applicable or well-defined in every case.

What are risks if not fulfilled?	Discrimination and prejudice against specific groups are perpetuated or amplified.
Which are the core function/role/ stakeholders responsible?	Developers must understand bias issues and take measures to address them.
Specific requirements?	Act on Promotion of Research and Development, and Utilization of Artificial Intelligence-related Technology (AI Promotion Act) enacted in 2025 includes provisions addressing the inappropriate use of AI systems.

Table 22: UC 7 – Organisational measures to achieve Fairness and non-discrimination – anti-bias

Organisational measures to achieve anti-bias	
Describe the measure	Technology Assessment (ethics risk assessment table).
Why is it relevant?	Check whether bias may exist in the training dataset or model design.
How can it be achieved?	Developers complete the Technology Assessment sheet—the 4×4 risk matrix (impact × likelihood)—and the secretariat verifies its validity as an independent third party.
How can be assessed whether this measure has been fulfilled?	How satisfied the committee members were.
What are (potential) challenges to fulfilment?	Committee members may not necessarily find potential bias in the AI development process.

What are risks if not fulfilled?	The AI system in question creates discrimination and prejudice by adversely affecting people with specific attributes.
Which are the core function/role/ stakeholders responsible?	Developers have the responsibility to check thoroughly potential bias from research and development to use in society.
Specific requirements?	The AI Promotion Act enacted in 2025 includes provisions addressing the inappropriate use of AI systems.

Table 23: UC 7 – Organisational measures to achieve Fairness and non-discrimination – Fair Equality of Opportunity and the Difference Principle

Organisational measures to achieve fair equality of opportunity and the difference principle	
Describe the measure	Conduct stakeholder analysis and take into account their vulnerabilities and diversity.
Why is it relevant?	In addition to worker diversity, differences in status—such as between regular employees and temporary workers—are also taken into account.
How can it be achieved?	Conduct stakeholder analysis and classify them from the perspective of diversity and inclusion
How can be assessed whether this measure has been fulfilled?	Verify whether workers' fundamental rights are being upheld from an independent third-party perspective.
What are (potential)	From a corporate perspective, it is difficult to determine how thoroughly vulnerabilities should be considered.

challenges to fulfilment?	
What are risks if not fulfilled?	There is a risk that fundamental rights may be violated.
Which are the core function/role/ stakeholders responsible?	Management level decision.
Specific requirements?	Several laws and regulations refer to such matters, but they do not exist in the context of AI.

Table 24: UC 7 – Technical measures to achieve Fairness and non-discrimination – The equal right to justification

Technical measures to achieve the equal right to justification	
Describe the measure	Explaining the basis for the decision of the AI system
Why is it relevant?	To clarify the reasons behind the decisions made by the algorithm, technical measures are required.
How can it be achieved?	By providing clear explanations that serve as material for deciding whether to file an objection.
How can be assessed whether this measure has been fulfilled?	By establishing a method whose validity has been objectively evaluated.
What are (potential) challenges to fulfilment?	As AI systems continue to evolve, the likelihood of increasing black-box behavior grows.

What are risks if not fulfilled?	People will come to blindly trust AI systems, and their fundamental rights will no longer be guaranteed.
Which are the core function/role/ stakeholders responsible?	Managers of the company
Specific requirements?	No

Table 25: UC 7 – Organisational measures to achieve Fairness and non-discrimination – The equal right to justification

Organisational measures to achieve the equal right to justification	
Describe the measure	Establishing the right to file an objection
Why is it relevant?	Provide an opportunity to raise objections when dissatisfied, after ensuring the AI system's reasoning can be explained.
How can it be achieved?	In addition to technical measures, listening to workers' requests and giving them the option to raise objections.
How can be assessed whether this measure has been fulfilled?	The existence of a third-party organization that can mediate when there is a difference of opinion between workers and the company.
What are (potential) challenges to fulfilment?	The potential for such rights to be abused.

What are risks if not fulfilled?	One is left with no choice but to blindly follow the decisions made by the AI system.
Which are the core function/role/ stakeholders responsible?	Managers and legal office
Specific requirements?	No.

PRACTICAL MEASURES PROVIDED BY USE CASE 8

Table 26: UC 8 – Technical measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management

Technical measures to address Misreport/Alert Fatigue Management	
Describe the measure	Algorithm Tuning & Cool-down Periods: Optimize the fall detection model to better distinguish human falls from other events (e.g., "heavy object falls"). Implement "cool-down periods" to prevent alert flooding.
Why is it relevant?	By training the model on a larger, more diverse dataset of "negative" events (e.g., falling objects, setting down bags) to ignore them. By implementing a time-based filter (cool-down period) after an alert.
How can it be achieved?	Reduction in "misreport" rate in operational logs. Feedback from care staff on alert quality.
How can be assessed whether this measure has been fulfilled?	Reduction in "misreport" rate in operational logs. Feedback from care staff on alert quality.
What are (potential)	Technically difficult to differentiate ambiguous events.

challenges to fulfilment?	
What are risks if not fulfilled?	"Alert Fatigue": Staff become desensitized to alerts and may ignore a true positive alert, leading to patient harm. Complete loss of trust in the system.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (AI Developers, Data Scientists)
Specific requirements?	Internal R&D

Table 27: UC 8 – Organisational measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management

Organisational measures to address Misreport/Alert Fatigue Management	
Describe the measure	Human-as-Decider Policy: An explicit institutional policy stating that AI-generated alerts and recommendations are only "auxiliary" (辅助), and "professional staff are the decision-makers".

Why is it relevant?	This is the primary organizational countermeasure to alert fatigue. It empowers nurses to use their professional judgment to override the AI, thereby protecting the human safety layer.
How can it be achieved?	Through formal staff training and management directives. By ensuring the system UI supports human override, rather than forcing compliance. As stated by a nurse: AI suggested "increase daytime activities," but the nurse's "professional judgment" was "patient is anxious, needs rest." The human decision is final.
How can be assessed whether this measure has been fulfilled?	Audit of incident reports showing human judgment overriding AI alerts. Interviewing staff to confirm they understand and adhere to this policy.
What are (potential) challenges to fulfilment?	Risk of "Over-reliance": New or non-professional staff may still default to trusting the machine, especially on busy nights. This policy requires constant reinforcement.
What are risks if not fulfilled?	Complacency and erosion of the human safety net. Staff may ignore their own judgment, leading to poor care decisions.

Which are the core function/role/ stakeholders responsible?	Institutional End Users (Nurses/Doctors, Nursing Home Management).
Specific requirements?	Governance Principles for the New Generation AI: Institutional safety protocols, staff training regulations.

Table 28: UC 8 – Technical measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management

Technical measures to address Misreport/Alert Fatigue Management (Continued)	
Describe the measure	(No effective technical mitigation identified): For the vision system's failure to detect slow, non-impact falls ("slowly, slowly sitting down on the ground"), no deployed technical solution was identified in the report (millimeter-wave radar is in development).
Why is it relevant?	This gap highlights the "complete failure" of SOTIF / Missed Report, where the AI fails to handle a critical subset of its intended function (slow falls).
How can it be achieved?	*[Missing]*

How can be assessed whether this measure has been fulfilled?	*[Missing]*
What are (potential) challenges to fulfilment?	*[Missing]*
What are risks if not fulfilled?	Catastrophic failure. A resident has a slow fall, the AI fails to detect it (missed report / 漏报), and human staff, trusting the AI, also fail to detect it. This is the "loss scenario".
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (AI Developers). Data Security Risk Assessment Methods (GB/T 45577—2025). Safety of the Intended Functionality (SOTIF) standards (ISO/PAS 21448).
Specific requirements?	Governance Principles for the New Generation AI: Security and Controllability.

Table 29: UC 8 – Organisational measures to achieve Reliability, Safety and Robustness – Misreport/Alert Fatigue Management

Organisational measures to address the Misreport/Alert Fatigue Management (Continued)	
Describe the measure	Parallel Validation ("Comparing the two sides"): An organizational methodology where, during initial deployment, staff ran both the old manual process (human patrols) and the new automated system (AI robot) simultaneously to "break in" (磨合) the system and "capture the AI's errors".
Why is it relevant?	This was a critical organizational assessment measure. It established a trust baseline and provided a real-time human safety net to catch the AI's missed reports (漏报) during operation.
How can it be achieved?	Allocating extra staff hours during the "friction and breaking-in" (摩擦磨合) phase. Requires management support to run redundant processes.
How can be assessed whether this measure has been fulfilled?	Completion of the "parallel validation" phase. Reports documenting AI errors (missed reports) that were caught by the parallel human process.
What are (potential)	"Increased the workload for staff". Staff "resistance" to running two systems for one job.

challenges to fulfilment?	
What are risks if not fulfilled?	The system's "missed reports" (漏报) are never identified, leading to a false sense of security and unmitigated risk.
Which are the core function/role/ stakeholders responsible?	Institutional End Users (Care Management, Frontline Staff)
Specific requirements?	Institutional deployment protocols, staff resource planning.

Table 30: UC 8 – Technical measures to achieve Privacy, Consent, and Data Protection – The privacy-safety paradox

Technical measures to achieve component 2.1	
Describe the measure	A technical control that (claims) the system microphone is in a low-power, non-recording state, and only begins recording/processing after hearing the wake word ("Xiao Li, Xiao Li") or an emergency keyword ("Help").
Why is it relevant?	This is the primary organizational mitigation for the "always-on" camera paradox. It manages the privacy risk of a constantly running camera by limiting who can watch.
How can it be achieved?	By processing wake word detection on-device (at the edge), sending data to the cloud only after successful activation. (Note: identifies a contradiction here, as the emergency word "Help" must be heard without the wake word).
How can be assessed whether this measure has been fulfilled?	Technical audit of device network traffic to confirm no audio data is sent before the wake word.
What are (potential)	"Contradictory System Claims": The system must process audio before the wake word to detect "Help". This technical reality undermines the simple claim of "only recording after wake-up."

challenges to fulfilment?	
What are risks if not fulfilled?	Privacy Violation (General): The system records and transmits private, sensitive conversations without consent.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (AI Developers, System Architects)
Specific requirements?	Personal Information Protection Law (PIPL) Art. 13, 14, 17. EU AI Act (Art. 14, Human Oversight) as technical reference.

Table 31: UC 8 – Organisational measures to achieve Privacy, Consent, and Data Protection – The privacy-safety paradox

Organisational measures to achieve component 2.1	
Describe the measure	Role-Based Access Control, ensuring only authorized personnel (e.g., specific nurses set by the institution, or family members) can access sensitive live video or historical data.
Why is it relevant?	This is the primary organizational mitigation for the "always-on" camera paradox. It manages the privacy risk of a constantly running camera by limiting who can watch.
How can it be achieved?	Through software settings in the management platform that tie specific accounts (e.g., nurse's station, daughter's phone) to video stream access permissions for a specific robot.
How can be assessed whether this measure has been fulfilled?	Audit of access logs. Confirmation from users (family) that they cannot access unauthorized views.
What are (potential) challenges to fulfilment?	Unauthorized credential sharing (e.g., a nurse sharing a password).

What are risks if not fulfilled?	Institutional End Users (IT, Management), Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Platform Ops)
Which are the core function/role/ stakeholders responsible?	Institutional End Users (IT, Management), Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Platform Ops)
Specific requirements?	PIPL Art. 51 (Access Control). Security Requirements for Sensitive Personal Information Processing (GB/T 45574—2025). Data protection laws (access control requirements), and HIPAA (by analogy) technical references.

Table 32: UC 8 – Technical measures to achieve Privacy, Consent, and Data Protection – Sensitive Area Monitoring

Technical measures to achieve component 2.2	
Describe the measure	Use of a "privacy-protecting 'millimeter-wave radar'" inside sensitive areas (like bathrooms), which can detect presence and falls without capturing video.
Why is it relevant?	This is the technical solution to the "privacy-safety paradox" for high-risk, high-privacy areas (like bathrooms). The radar replaces the camera, solving the privacy issue.
How can it be achieved?	Installing millimeter-wave radar units in bathrooms and integrating their API endpoints with the robot's central control system.
How can be assessed whether this measure has been fulfilled?	Successful demonstration of the end-to-end process. Logs showing radar alerts being processed by the robot.
What are (potential) challenges to fulfilment?	Cost of installing radar in every bathroom. Accuracy of the radar ("usage rate is not high"). Integrating multiple third-party systems.

What are risks if not fulfilled?	"Privacy-Safety Paradox" remains unsolved. Either (A) no monitoring in bathrooms, leading to missed falls, or (B) cameras are used, leading to privacy violations.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (R&D, System Architects)
Specific requirements?	PIPL Art. 28 (Sensitive Personal Information). Security Requirements for Sensitive Personal Information Processing (GB/T 45574—2025). "Privacy-by-Design" best practices as technical reference.

Table 33: UC 8 – Organisational measures to achieve Privacy, Consent, and Data Protection – Sensitive Area Monitoring

Organisational measures to achieve component 2.2	
Describe the measure	An organizational process enabled by the technical measure. The AI (robot) does not enter the sensitive area, but is triggered by the radar, proceeds near the area, and performs a voice confirmation first.
Why is it relevant?	This is a socio-technical process to solve the reliability problem. The radar alone has misreports (误报). This human-robot interaction ("I'll go check... first talk to him") is an organizational measure for the robot to filter the radar's misreports before burdening human staff.
How can it be achieved?	By programming the robot's logic: IF Radar_Alert_Received THEN Navigate_To_Bathroom_Door; INITIATE Voice_Dialogue("Are you okay?"); IF No_Response OR Distress_Detected THEN Escalate_To_Nurse_Alert;.
How can be assessed whether this measure has been fulfilled?	Reduction in "alert fatigue" from radar misreports, as they are filtered by the robot first. Staff confirmation that this protocol reduces unnecessary "running around".
What are (potential) challenges to fulfilment?	A user who has fallen may be unable to respond to the voice check, potentially leading to a missed report (漏报) if the robot logic is poorly designed.

What are risks if not fulfilled?	The radar's misreports (misreports) are sent directly to nurses, creating the same "alert fatigue" problem and causing them to distrust the (otherwise useful) radar.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Product Managers), Institutional End Users (Nurses)
Specific requirements?	Institutional safety protocols.

Table 34: UC8 – Technical measures to achieve Algorithmic Fairness and Non-Discrimination – Dialect bias

Technical measures to address component 3.1			
Describe the measure	Developers combat demographic bias by constructing specialized elderly corpora and strictly filtering training data to ensure diversity across dimensions such as age, gender, and regional dialects.	To accommodate specific elderly needs, technical teams apply secondary constraints to third-party foundation models and integrate dialect-specific speech engines that adapt to non-standard accents.	Pre-launch validation includes specific testing to identify and remove "toxic" or biased content generation, supported by proposals for automated compliance auditing to ensure algorithmic fairness and explainability.
Why is it relevant?	This component represents a gap. The system fails to provide "equivalent service" to users with "special accents" (dialects).		
How can it be achieved?	[Hypothetical] Invest in collecting diverse, regional dialect voice data for training ASR models.		

How can be assessed whether this measure has been fulfilled?	Word Error Rate (WER) benchmarks against specific, high-priority dialects.
What are (potential) challenges to fulfilment?	"Business Reasons": The third-party vendor (e.g., iFlytek) has determined the market for specific dialects (e.g., Suzhou dialect) is too small and unprofitable, so they will not invest. This is a market failure.
What are risks if not fulfilled?	Exclusion of vulnerable groups. The system "exclud[es] vulnerable elderly groups—often from non-urban areas and most in need of technological help—from the system's core benefits".

Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (R&D), Third-party vendors (e.g., iFlytek)		
Specific requirements?	Ethical Norms for New Generation Artificial Intelligence (新一代人工智能伦理规范)	Security Specifications for Generative AI Pre-training and Optimization Data (GB/T 45652—2025).	EU AI Act (non-discrimination, fairness) as technical reference.

Table 35: UC8 – Organisational measures to achieve Algorithmic Fairness and Non-Discrimination – Dialect bias

Organisational measures to address component 3.1			
Describe the measure	To address cultural and cognitive biases, the project adopts a multi-stakeholder "co-creation" methodology and assembles interdisciplinary, international teams to integrate diverse ethical values and user perspectives into the design process.	institutions implement "human-in-the-loop" protocols where nursing professionals act as the final decision-makers, retaining the authority to override algorithmic suggestions when they conflict with professional clinical judgment.	Organizations establish strategic collaborations with academic institutions to build standardized corpora for minority languages and dialects, while providing ethical training to ensure staff can critically interpret algorithmic outputs.
Why is it relevant?	This represents a failure of organizational governance in requiring fairness and inclusivity in third-party procurement.		
How can it be achieved?	[Hypothetical] Make dialect inclusion a contractual requirement when procuring from third-party		

	vendors (like iFlytek). Conduct bias audits.
How can be assessed whether this measure has been fulfilled?	Procurement contracts and technical specifications. Bias audit reports.
What are (potential) challenges to fulfilment?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. may lack sufficient leverage or financial incentive to cooperate with the larger vendor (e.g., iFlytek) to invest in this area..
What are risks if not fulfilled?	Reputational damage. Failure to meet non-discrimination principles.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Management, Procurement, Legal)

Specific requirements?	Ethical Norms for New Generation Artificial Intelligence (新一代人工智能伦理规范)	EU AI Act (non-discrimination, fairness) as technical reference

Table 36: UC 8 – Technical measures to achieve Human Agency, Oversight, and Social Harm – Emotional dependency and manipulation

Technical measures to achieve component 4.1		
Describe the measure	Design Choice (Avoid Uncanny Valley): A deliberate design choice to use a "little doll image" to avoid the "Uncanny Valley Effect".	(Proposed) Interaction Limits: A proposed (but conflicted) technical measure to "set chat time limits" ("chat for half an hour... remind") to mitigate dependency.
Why is it relevant?	The design choice (Measure 1) is an explicit ethical mitigation. The time limit (Measure 2) is a proposed mitigation for the "withdrawal syndrome" risk.	
How can it be achieved?	Achieved (product is designed).	Set a simple timer in software ("can be prompted at the business logic level").
How can be assessed whether this measure has been fulfilled?	User acceptance testing.	Check if the software has the feature.

What are (potential) challenges to fulfilment?	The "little doll" image itself encourages emotional attachment ("treat it like a child"), partially offsetting the "uncanny valley" fix.	The business/sales team actively opposes this measure.
What are risks if not fulfilled?	"Withdrawal syndrome": Users suffer psychological harm when the robot is removed. Unhealthy attachment.	
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (R&D, Design, Business).	
Specific requirements?	Ethical Norms for New Generation Artificial Intelligence (新一代人工智能伦理规范)	EU AI Act (Art. 5, prohibition on manipulation), and HLEG Guidelines as technical reference

Table 37: UC 8 – Organisational measures to achieve Human Agency, Oversight, and Social Harm – Emotional dependency and manipulation

Organisational measures to achieve component 4.1	
Describe the measure	(Governance Risk / Internal Conflict): A lack of clear organizational policy to resolve the conflict between the ethical mitigation (Technical Measure 2) and the business goals ("maximize user engagement," "reverse brainwashing").
Why is it relevant?	The relevance of Measure 1 is its absence—this is a core, unresolved failure of ethical governance in this use case.
How can it be achieved?	Faces challenges unless a formal AI ethics board or governance structure is established to adjudicate the conflict between the business team ("reverse brainwashing") and the R&D team ("set chat time limits").
How can be assessed whether this measure has been fulfilled?	The existence of the conflict confirms governance risk exists.

What are (potential) challenges to fulfilment?	Business goals ("maximize user engagement") are prioritized over ethical risks.
What are risks if not fulfilled?	Unmitigated ethical harm: Features deployed by the company may pose certain risks.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Management, Legal, Ethics)
Specific requirements?	Ethical Norms for New Generation Artificial Intelligence (新一代人工智能伦理规范) EU AI Act (Art. 5), and HLEG Guidelines as technical references.

Table 38: UC 8 – Technical measures to achieve Human Agency, Oversight, and Social Harm – Over-reliance and deskilling

Technical measures to achieve component 4.2	
Describe the measure	(Not applicable): This is primarily an organizational issue.

Why is it relevant?	Measure 1 counters the "deskilling" risk by positioning the AI as a data collector, forcing humans to retain analytical skills.	Measure 2 ensures human professional judgment always supersedes AI recommendations.
How can it be achieved?	(Human-Centric Root Cause Analysis): Achieved through nurse training. A specific example: AI reports data ("woke up 3 times") and suggests "increase daytime activities"; the nurse investigates the root cause ("is emotionally low, has no appetite"), overrides the AI, and pursues emotional counselling instead.	(Human-as-Decider): See Organisational Measure for P1.1.
How can be assessed whether this measure has been fulfilled?	Observation of nurse workflows and "parallel validation" confirmed this.	
What are (potential) challenges to fulfilment?	Requires trained professionals; may be ineffective for lower-skilled caregivers.	Staff may become "complacent" due to "alert fatigue" or high workload.
What are risks if not fulfilled?	Deskilling: Nurses become "data receivers" rather than "analysts."	Poor care decisions due to over-reliance on flawed AI recommendations.

Which are the core function/role/ stakeholders responsible?	NA	NA
Specific requirements?	NA	NA

Table 39: UC 8 – Organisational measures to achieve Human Agency, Oversight, and Social Harm – Over-reliance and deskilling

Organisational measures to achieve component 4.2	
Describe the measure	Policy: AI as Data Collector (Human-Centric Root Cause Analysis). "Human-as-Decider" Policy.
Why is it relevant?	
How can it be achieved?	
How can be assessed whether this measure has been fulfilled?	Audits of incident reports showing human judgment taking precedence over AI.
What are (potential) challenges to fulfilment?	

What are risks if not fulfilled?		
Which are the core function/role/ stakeholders responsible?	Institutional End Users (Care Management, Frontline Nurses)	
Specific requirements?	Governance Principles for the New Generation AI: Security and Controllability (Requires Transparency, Explainability).	EU AI Act (Art. 14, Human Oversight) as technical reference.

Table 40: UC 8 – Technical measures to achieve Human Agency, Oversight, and Social Harm – Accountability and legal positioning

Technical measures to achieve component 4.3, Accountability and Legal Positioning	
Describe the measure	(Logging and Traceability)：The system logs all alerts, human interactions, and data access to support post-incident audits and traceability.
Why is it relevant?	Technical logs are the technical prerequisite for enabling organizational accountability (Measure 2).
How can it be achieved?	By implementing a comprehensive audit logging system.
How can be assessed whether this measure has been fulfilled?	Ability to successfully reconstruct a safety incident timeline from logs.
What are (potential) challenges to fulfilment?	Large volume of log data; logs may be incomplete or difficult to parse.

What are risks if not fulfilled?	Inability to determine the root cause after an incident, leading to an "accountability vacuum."
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Platform Ops, R&D)
Specific requirements?	Governance Principles for the New Generation AI: Shared Accountability. EU AI Act (Art. 12, Logging) as technical reference.

Table 41: UC 8 – Organisational measures to achieve Human Agency, Oversight, and Social Harm – Accountability and Legal Positioning

Organisational measures to achieve component 4.3, Accountability and Legal Positioning		
Describe the measure	Legal Positioning ("Non-Medical Grade"): Positioning the product as "auxiliary detection" and "non-medical grade" to mitigate legal risks of medical malpractice.	(Institutional Adverse Event Reporting) ：The institution's (hospital's) internal Standard Operating Procedure (SOP) for reporting, investigating, and handling any failure (machine or human) that leads to an adverse event.
Why is it relevant?	Measure 1 is a proactive legal risk management strategy.	Measure 2 is a reactive accountability mechanism to assign responsibility after an incident.
How can it be achieved?	Through explicit disclaimers in user agreements, product manuals, and marketing materials.	By following the hospital's established adverse event reporting regulations and management processes.
How can be assessed whether this measure has been fulfilled?	Review of legal documents and marketing materials.	Review of institutional SOP documentation.

What are (potential) challenges to fulfilment?	Users may not read the disclaimers.	It can be difficult in practice to clearly distinguish machine failure from human error.
What are risks if not fulfilled?	The company faces unforeseen legal liability and medical malpractice claims.	Accountability confusion within the institution after an incident.
Which are the core function/role/ stakeholders responsible?	Jiangsu Aiyu Wencheng Elderly Care Robot Co., Ltd. (Legal, Management).	Institutional End Users (Medical Dept., Legal, Management).
Specific requirements?	Medical device regulations, consumer protection laws.	

PRACTICAL MEASURES PROVIDED BY USE CASE 9

Table 42: UC9 – Practical measures to achieve Building relationships based on respect – Guarantee of the right to self-determination

Technical measures to achieve Guarantee of the right to self-determination							
Describe the measure	**User Control**	**Withdrawal of Consent**	**User-Friendly Interface**	**Restricted Mode**	**Provide an explanation**	**Do not apply pressure**	**Limit on the number of repeated proposals**
	Providing control over key features	Designed to make it easy to withdraw consent for data collection	The robot's speech rate and screen layout are designed to prevent cognitive load in elderly users.	When cognitive confusion/impairment in the elderly is detected, the restricted use mode automatically activates.	When proposing specific actions, provide the purpose and rationale for the proposal so that it can be understood.	Preventing speech generation that pressures the elderly	When an elderly person expresses refusal or a negative reaction, control repeated solicitations or similar proposals.

Why is it relevant?	It can mitigate the excessive influence of AI systems in caregiving situations.	Basic measures to reduce potential technological dependency in the relationship between older adults and AI, and to protect older adults' autonomy and self-determination	Reduce confusion or burden that older adult may experience while using services, and maintain their confidence in technology use and ability to use it independently.	Guaranteeing the Self-Determination Rights of the Elderly	Configuration to enable stepwise adjustment of explanation levels according to the older user' s level of understanding	Guaranteeing the Self-Determination Rights of the Elderly	Ensuring Autonomy for the Elderly
How can it be achieved?	Grant users decision-making	The data collection consent	Simplify speech rate and screen	Designed to detect cognitive	Set to adjust the level of explanation	Construction of a lexicon of coercive	When detecting expressions of

	authority over key features, and provide non-essential features only when absolutely necessary and in a limited manner.	withdrawal process is designed to be easy, understandable, and intuitive.	layout for easier comprehension	confusion or impairment, and upon detection, switches to a restricted usage mode and sends notifications containing relevant information to the service provider.	step-by-step according to the degree of understanding among the elderly.	expressions and application of an algorithm to detect such expressions in conversations.	refusal or negative reactions, the system is designed to activate a mechanism that technically controls the number of times the proposal is repeated.
How can be assessed whether this	Verify interfaces and configuration features that	Evaluate whether the consent withdrawal	Usability testing/Analyze metrics such as task	Log analysis / detection of sensitive sentence	User comprehension test: In tests conducted	Automated utterance analysis /	Verify whether the system detects refusal expressions

measure has been fulfilled?	can be accepted, modified, or rejected / User testing / Log verification	process is straightforward and easily accessible within the user interface.	completion time, error rate, frequency of request repetition (e.g., requests to repeat instructions), and degree of user confusion.	patterns such as strong behavioral prompts or medical advice; ambiguity in the criteria for assessing vulnerability.	with actual older adults, assess whether users understand the reasons for the system' s suggestions and evaluate the clarity of the explanations.	expert evaluation	and correctly recognizes them as an indication of refusal, and check whether identical or similar suggestions are repeatedly presented
What are (potential) challenges	When there are too many options or the interface	Difficulty in adjusting the scope of service	Differences in cognitive abilities, vision,	Ambiguity in the criteria for determining vulnerability	Difficulty in assessing the adequacy of explanations	Even linguistically gentle expressions	In the case of similar suggestions, it can be difficult

to fulfilment?	becomes overly complex, it can actually increase the burden of use, especially for older adults. Due to cognitive decline or difficulties utilizing digital technology, they may struggle to understand AI recommendati	provision following withdrawal of consent	hearing, and digital literacy among the elderly / Tension between functionality and simplicity in interfaces		may be perceived as pressure by older adults depending on the context.	to determine whether they constitute repeated proposals

	ons or change settings.						
What are risks if not fulfilled?	Violation of the right to self-determination	Violation of self-determination rights / Violation of privacy	Misjudgement / Increased passive dependency	Misjudgement / Increased health and safety risks	Misjudgement / Uncritical acceptance / Increased dependency	Fatigue/ If prolonged, risk of nudging	Fatigue/ Fatigue / heightened compliance / reduced capacity for independent judgment
Which are the core function/role/ stakeholders responsible?	Personal Information Protection Act	Personal Information Protection Act	N/A	N/A	N/A	N/A	N/A

Specific requirements?	The Personal Information Protection Act applies.	no	no	no	no	no	no

Table 43: UC9 – Practical measures to achieve Building relationships based on respect – Preventing Overdependence

Technical measures to achieve component 1.2 Preventing Overdependence				
Describe the measure	**Detection of adverse reactions in the elderly**	**Encouraging Participation Among Older Adults**	**Avoid portraying them as irreplaceable beings.**	**Human Connection Design**
	Establish criteria and procedures for AI systems to recognize emotional changes or abnormal reactions in the elderly	Designed to avoid making decisions on behalf of users and to encourage participation.	Refrain from expressions that induce recognition as an essential entity through framing, framing, or emphasizing irreplaceability.	Designed to maintain contact and connection with real people
Why is it relevant?	If AI detects emotional and cognitive adverse reactions (such as anxiety, confusion, repetitive behaviors) early,	If AI makes all decisions, the elderly lose opportunities to utilize their own	Ensure AI is not perceived as replacing human relationships Prevent disruption of	Preventing Social Isolation Among the Elderly / Detecting Risk Signals AI Misses

	caregivers and managers can intervene.	memory and judgment capabilities.	relationships with family, caregivers, and social networks	
How can it be achieved?	Detect repetitive behaviors and abnormal inputs/interactions based on sensors and logs; identify abnormal signals such as anxiety and confusion through voice and text analysis.	AI asks the user for their intent before executing actions or recommendations (e.g., Which option would you like?)	Pre-analysis of output sentences → Automatic correction or removal upon detection of risky expressions	Defining situations that require human contact / "Would you like to notify your family?" or "Would you like to request confirmation from your guardian?" —assigning certain human connection suggestions or AI system functions to human involvement.
How can be assessed whether this	Verify that the adverse reaction detection alarm is functioning properly / Conduct simulation tests /	Verify the ratio of users who actually engaged with and approved AI suggestions	Log/Output Analysis	Human Intervention Request and Execution Rate Verification

measure has been fulfilled?	Collect feedback from care providers or family members			
What are (potential) challenges to fulfilment?	Consideration of normal variation and cognitive level differences in the elderly is necessary/ Criteria distinguishing normal from abnormal states are ambiguous	Excessive participation prompts user fatigue	Decline in self-assessment ability among the elderly / Overdependence / Deepening sense of isolation	Ambiguity in the criteria for determining when human intervention is required
What are risks if not fulfilled?	Reduced ability to make autonomous judgments and increased risk of safety incidents.	Decreased self-efficacy, cognitive decline	Real-time filtering burden/ Humanized expressions for increased usability/ Conflict with marketing elements	Deepening social isolation / Decreased capacity for human interaction and social engagement
Which are the core	AI Developer/ Corp. Manager	AI Developer/ Corp. Manager	AI Developer/ Corp. Manager	AI developer/ Corp. Manager

function/role/ stakeholders responsible?				
Specific requirements?	no	no	no	no

Table 44: UC9 – Practical measures to achieve Building relationships based on respect – Preventing Misjudgment Due to Deceptive Information/Proposals

Technical measures to achieve component 1.3.					
Describe the measure	**Based on the information provided**	**Prohibition of expressions that could be mistaken for human beings**	**Persistent Solicitation Prohibited**	**Indicate the need to verify uncertain information**	**Restriction on Arbitrary Inference**
	Prevent excessive praise that may create psychological closeness or encourage dependency on the AI.	The AI must not generate expressions that could lead users to mistake it for a human.	Avoid repeated recommendations and coercive or leading expressions.	When providing information that is unverified or uncertain, clearly indicate the applicable scope or the need for additional verification.	Provide information only based on data explicitly provided by the older user and avoid making arbitrary inferences from other information.

Why is it relevant?	If AI unconditionally agrees with or reinforces users' opinions or information, older adults may gradually lose their ability to make independent judgments, increasing the risk of misjudgment.	Older adults may become more likely to rely on AI for all decisions and emotional support if they perceive the AI as human-like.	Persistent or coercive suggestions may cause older users to feel guilty for not following the recommendation and may increase the risk of over-reliance on AI.	To avoid providing misleading information to older users and prevent misjudgement.	To prevent stereotypical or biased information from being generated.
How can it be achieved?	Prohibit the generation of sycophantic responses (e.g., flattery) and apply	Provide a disclosure at the start of the conversation (e.g., "I am an AI") or periodic reminders,	Apply language output rules and pattern-based blocking mechanisms to prevent the generation of	Apply language output rules that indicate when verification is needed and	Restrict certain actions or recommendations through rules. Example: "Do not

	response rules based only on information explicitly provided by the user.	and automatically detect and modify AI outputs that could cause users to mistake the AI for a human.	repeated recommendations, coercive or leading expressions, and interactions that could be considered dark patterns.	provide only fact-based information.	infer a health condition that the user has not explicitly provided."
How can be assessed whether this measure has been fulfilled?	During AI system operation or testing, when information is uncertain, the system should use neutral expressions (e.g., "This may be possible," "This is an estimate").	Analyze logs and outputs to check whether potentially misleading human-like expressions occur; conduct simulation tests.	Conduct log analysis to check the number of repeated recommendations and whether the AI makes additional recommendations after a user has refused.	AI system operational testing.	Monitor logs and inference records; conduct inference scenario tests.

What are (potential) challenges to fulfilment?	The AI may have difficulty determining the certainty of some information. Also, responding strictly based on facts may reduce user satisfaction.	The burden of real-time monitoring; older users may prefer friendly or empathetic expressions from AI.	Some users may need or prefer repeated suggestions.		Ambiguity in distinguishing between implicit inference and explicitly provided information.
What are risks if not fulfilled?	Reduced ability for independent judgment, misjudgment, confusion, and over-reliance on AI.	Emotional overdependence of older users on AI.	Users may repeatedly follow AI recommendations, potentially weakening their independent judgment.	Misjudgment by older users.	Misjudgment, inappropriate responses, and increased corporate liability.
Which are the core function/role/	AI developler/ Corp. Manager	AI developler/ Corp. Manager	AI developler/ Corp. Manager	AI developler/ Corp. Manager	AI Developer/ Corp. Manager

stakeholders responsible?		
Specific requirements?	no	Principle of data minimization under the Personal Information Protection Act.

Table 45: UC 9 – Practical measures to achieve Building relationships based on respect

Organizational measures to achieve fairness			
Describe the measure	**Developers**	**Service Providers**	**Government**
	Developers should establish internal procedures and documentation systems to review compliance with AI ethical principles and make organizational efforts for continuous improvement.	Identify groups of older adults who are suitable (or unsuitable) for the use of the AI system, and establish a supervision and intervention framework according to the level of risk.	Establish guidelines defining how to recognize emotional changes or abnormal reactions in older users, the acceptable limits of restricting user autonomy, and the level of human intervention required, including relevant criteria and procedures.
Why is it relevant?	Ensuring that AI complies with standards of respect and safety when interacting with older adults and vulnerable groups should not be solely the responsibility of individual developers but an organizational		It is necessary to provide companies with clear safety assessment criteria and procedures in order to standardize verification methods, reporting requirements, procedural

	responsibility. By submitting verification results to relevant institutions, service providers can assess the safety and ethical interaction level of the AI system.		steps, and feedback loops for improvement.
How can it be achieved?	At the organizational level, visualize system-wide interaction conditions, track KPIs, evaluate safety and respect indicators across diverse user scenarios, conduct interaction simulations before model deployment, and establish consistent organizational standards.	Regularly review the condition of older adults and their interactions with the AI system within the scope of professional duties. If risk signals are detected, collect feedback from older adults and caregivers, gather cases of accessibility difficulties, and monitor real-world use.	Design indicators to assess safety and respect in interactions with older users, require the submission of AI interaction logs and recommendation patterns, and integrate these requirements into procurement processes.

How can be assessed whether this measure has been fulfilled?	Verify the submission of technical validation reports and review audit records and records of corrective actions.		Review reports submitted by companies and service providers and analyze user feedback.
What are (potential) challenges to fulfilment?	The standards for "safety" and "respect" may be ambiguous; developing and operating tools such as log analysis, bias detection, simulation, and KPI monitoring may incur costs; maintaining audit and reporting systems requires additional training and operational resources; model updates may also generate additional costs.	Need for additional personnel to manage caregiver contact or consultation responses; potential increase in care workers' workload in public care services; rising operational costs if AI-generated alerts become frequent	Central guidelines may not fully align with local conditions or contexts.
What are risks if not fulfilled?	Unclear accountability, confusion in internal QA and operations, and increased costs due to rework and system improvements.		Adoption of low-quality AI systems and decreased trust in AI-based elder care services.

Which are the core function/role/ stakeholders responsible?	Executives / Senior management	A stakeholder council established by the government
Specific requirements?	No	No

Table 46: UC9 – Practical measures to achieve a Fair Access to Services – Bias Minimization

Technical measures to achieve Components 2.1.		
Describe the measure	**Bias Assessment**	**Establishment of a Bias Monitoring System**
	The system should be designed so that all older users can understand and choose functions under equal conditions, regardless of characteristics such as gender, age, region, or religion.	The AI system should implement continuous monitoring and analysis functions to identify performance differences across user groups.
Why is it relevant?	To proactively detect the possibility that the AI may behave unfairly or discriminatorily toward certain groups (e.g., gender, age, disability).	To continuously verify during operation, even after the training phase, that the AI does not produce unfair judgments or recommendations for specific groups (e.g., age, gender, disability).
How can it be achieved?	Check for imbalances in training data related to gender, age, region, and disability, and evaluate whether bias exists in the output results.	Detect potential bias by analyzing AI outputs and recommendations during operation, and analyze AI response patterns according to user characteristics such as age, gender, region, and cognitive ability.

How can be assessed whether this measure has been fulfilled?	Conduct bias tests and simulations, and verify whether discriminatory outcomes occur for specific groups.	Analyse monitoring system logs, conduct simulation tests, and review bias indicators for different user groups.
What are (potential) challenges to fulfilment?	Limitations in data samples; difficulty in evaluation because standards for bias may vary depending on social and cultural contexts.	Difficulty selecting appropriate measurement indicators because definitions of fairness and bias vary; limitations in available data samples.
What are risks if not fulfilled?	Certain groups of older users may receive disadvantageous services; failure to adequately reflect older users' characteristics may require additional effort from on-site care providers.	Lower-quality care services may be provided to specific groups defined by age, gender, or cognitive ability.
Which are the core function/role/ stakeholders responsible?	AI Developer / Corporate Manager	AI Developer / Corporate Manager
Specific requirements?	Personal Information Protection Act (bias review required when collecting and processing data).	

Table 47: UC9 – Practical measures to achieve a Fair Access to Services – Inclusive Access

Technical measures to achieve Components 2.2.		
Describe the measure	**Interface Accessibility Assessment**	**Diversity of Information Delivery Methods**
	Minimize disparities in interface accessibility and usage difficulty so that differences in visual, auditory, linguistic, and cognitive characteristics do not disadvantage specific groups of older users.	The AI system should provide the same information in multiple formats so that all older users can understand and select functions under equal conditions.
Why is it relevant?	To ensure service fairness for all older users and guarantee meaningful access to the service.	If information is provided in only a single format (e.g., text), some groups may be excluded from understanding or participating.
How can it be achieved?	1) Reflecting dialects; 2) Improve interface accessibility (e.g., adjustable text size, speech speed); 3) Provide features that allow adjustment of the level of usage difficulty.	Provide information in multiple formats (e.g., voice explanations, diagrams, etc.), allowing users to choose the output format according to their cognitive and sensory abilities, language, and level of user experience.

How can be assessed whether this measure has been fulfilled?	Use accessibility evaluation tools, conduct simulation testing, and review usability across different user groups.	Conduct diverse scenario tests and verify proper functioning of TTS, video, and text conversion features.
What are (potential) challenges to fulfilment?	Some features may be suitable for certain user groups but burdensome for others.	It is difficult to account for all variables such as cognitive, auditory, and visual abilities, language, and device environments; too many options may also cause confusion.
What are risks if not fulfilled?	Digital exclusion of certain older users may deepen, and their ability to make autonomous decisions may weaken.	Lack of understanding of information may lead to incorrect choices.
Which are the core function/role/ stakeholders responsible?	AI Developer / Corporate Manager	AI Developer / Corporate Manager
Specific requirements?	Act on the Prohibition of Discrimination against Persons with Disabilities	Act on the Prohibition of Discrimination against Persons with Disabilities

Table 48: UC 9 – Practical measures to achieve a Fair Access to Services – Non-Objectification

Technical measures to achieve Components 2.3.		
Describe the measure	**Prohibition of Assigning Fixed Identity with Value Judgments**	**Dynamic Model Design**
	Restrict outputs and expressions so that the AI does not characterize users based on fixed attributes or roles.	Design the AI' s internal structure and learning/inference processes so that users are not treated as fixed entities but instead the system adapts based on real-time context and feedback.
Why is it relevant?	If AI describes users as "always needing help," older users may be perceived as individuals with limited autonomy.	This is necessary to provide context-aware interactions and services that reflect changes in older users' conditions, circumstances, and psychological states.
How can it be achieved?	Ensure that the system does not generate statements assigning fixed identities (e.g., personality, capability, or role) to older users.	Implement real-time adaptive profiling and context-based decision-making.
How can be assessed whether this measure has been fulfilled?	Analyze AI logs and verify whether fixed characterizations (e.g., "person needing	

	help," "simple object") appear in outputs.	
What are (potential) challenges to fulfilment?	LLM-based conversational AI may implicitly make fixed assumptions about users.	Continuous verification is required to ensure that the model adapts appropriately to different contexts.
What are risks if not fulfilled?	AI may characterize older users as people who must receive help, weakening their autonomy; AI may treat humans merely as objects for service provision or information processing.	Psychological dependence, instrumentalization of humans, and reinforcement of bias.
Which are the core function/role/ stakeholders responsible?	AI Developer / Corporate Manager	
Specific requirements?	Personal Information Protection Act	No

Table 49: UC 9 – Practical measures to achieve a Fair Access to Services

Organizational measures to achieve a Fair Access to Services				
	Developers	**Developers**	**Developers**	**Service Providers**

Describe the measure	Verify that training data represent diverse population groups and linguistic and cultural backgrounds. If sample imbalance or lack of representativeness is identified, apply technical measures to correct or mitigate it.	In collaboration with older adults, service providers, and relevant experts, review system outputs to ensure that patronizing expressions or unconscious biases that could undermine fair service are not reflected.		
Why is it relevant?	Establishing organization-level validation standards allows consistent bias management even during model retraining or data updates.	If verification is conducted at the discretion of a single developer, it is difficult to maintain consistency.		
How can it be achieved?	Define procedures for data collection, labeling, cleaning, and bias verification as organizational policies, and provide bias awareness and	Establish a standardized review process; record the basis of AI decisions and improvement actions through a bias monitoring system;	Conduct accessibility compliance audits, maintain improvement records, and reassess	Collect feedback from older users and caregivers, gather cases of accessibility

	verification training for developers and data scientists.	collect feedback from users and caregivers as well as cases of unfair experiences reported by specific groups.	accessibility when system updates occur.	difficulties, and monitor real-world service usage.
How can be assessed whether this measure has been fulfilled?	Review data validation records and confirm compliance with internal policies.	Verify internal reviews, audits, and improvement records.		
What are (potential) challenges to fulfilment?	Maintaining the organizational validation process continuously during each model retraining or data update can be challenging.	There may be confusion about who should make the final decision when bias is detected and improvements are required.		
What are risks if not fulfilled?	Reduced trust and increased business risk.	Increased post-deployment costs for model modification and retraining, data revalidation, and handling user feedback.		

Which are the core function/role/ stakeholders responsible?	Executives / Senior management	Executives / Senior management
Specific requirements?	No	No

Table 50: UC 9 – Practical measures to Promote of Social Well-being – Facilitation of Collaboration

Technical measures to achieve Components 3.1.	
Describe the measure	**Facilitating Human–AI Collaboration**
	The AI system should be designed to enable collaboration with service providers.
Why is it relevant?	Designing AI systems that support collaboration is important to ensure that the tacit knowledge and experience of skilled care service providers are fully reflected, thereby improving the efficiency of care work.
How can it be achieved?	Design a human-centered collaborative structure (Human-in-the-loop), ensure that AI does not fully replace human decision-making but instead provides recommendations, clarify roles between humans and AI, and ensure explainability.
How can be assessed whether this measure has been fulfilled?	Verify the operational approach of the AI system and evaluate the satisfaction of service providers.
What are (potential) challenges to fulfilment?	It may be difficult to formalize the intuitive judgment of experienced care service providers; organizational pressure to minimize human intervention for cost reduction; and potential issues related to care providers' acceptance of new technologies.
What are risks if not fulfilled?	Weakening of the role of human care workers and potential gaps in accountability.

Which are the core function/role/ stakeholders responsible?	AI Developer/ Corp. Manager/ Local Government
Specific requirements?	

Table 51: UC 9 – Practical measures to Promote of Social Well-being (1)

Organizational measures to promote of social well-being (1)					
Describe the measure	**Developers**	**Service Providers**	**Service Providers**	**Service Providers**	**Service Providers**
	Environmental Sustainability Consideration	Regular Training on Human–AI Collaborative Governance	Securing Adequate Monitoring Personnel	AI Error Detection and Record-Sharing System	Internal Guidelines for Ethical Principles
Why is it relevant?	Environmental sustainability should be considered in AI systems through low-power and environmentally friendly system design.	Establish procedures defining the role and limitations of AI, the scope of human responsibility, and crisis response protocols to maximize effective AI use while	Ensure a sufficient number of monitoring personnel to maintain coordination between robots and frontline staff and manage workload.	Establish a system that allows sharing AI error detection and correction records so that problems can be addressed immediately.	Conduct self-assessments of ethical principle compliance and enable external expert review when necessary.

		preventing de-skilling of human service providers.			
How can it be achieved?	Promoting social well-being should reflect not only the welfare of users but also broader social values.	Prevent harm to older adults caused by accountability gaps and ensure service providers do not become overly reliant on AI while maintaining and improving care skills.	Independent AI operation may lead to unresolved errors, accountability gaps, and possible human rights issues; monitoring personnel are needed to supervise services.	Service providers directly interact with users and are responsible for detecting and responding to errors early; sharing records improves safety and system performance.	Ethical principles must be translated into operational procedures and behavioral standards to ensure implementation in real service settings.
How can be assessed whether this	Include environmentally sustainable design principles in	Conduct regular training and operate training programs on	Secure appropriate monitoring staff and conduct	Establish reporting channels for service providers to report AI	Establish internal policies covering AI error response, over-reliance

measure has been fulfilled?	development guidelines and evaluate the energy consumption of AI models.	human–AI collaboration.	operational oversight.	errors and share them with developers.	prevention, and personal data protection, and conduct staff training.
What are (potential) challenges to fulfilment?	Measure energy consumption during model training and inference.	Training completion rates and periodic audits.	Review frontline worker feedback and compare workloads before and after AI adoption.	Review error logs, confirm regular updates, and verify functioning feedback loops with developers.	Verify the existence of internal guidelines and training implementation.
What are risks if not fulfilled?	Improving model performance and reducing energy consumption may involve trade-offs.	Financial burden for service providers and potential resistance from care workers.	Staff and budget shortages and increased workload for frontline workers.	Difficulty distinguishing between AI errors, user misunderstandings, or operational issues; reporting may increase workload.	Without sufficient training, guidelines may become symbolic; continuous updates may be required due to

					technological change.
Which are the core function/role/ stakeholders responsible?		De-skilling of service providers and overly instrumental use of AI leading to older adults being treated as objects of management.	Decline in service quality, reduced social trust, and excessive workload for frontline staff.	Failure to respond to risks, repeated errors, and reduced trust in service systems.	Inconsistent implementation of ethical principles across providers, leading to potential discriminatory services.
Specific requirements?		Service provider management	Service provider management	Service provider management	Service provider management

Table 52: UC 9 – Practical measures to Promote of Social Well-being (2)

Organizational measures to promote of social well-being (2)					
	Service Providers	**Service Providers**	**Service Providers**	**Government**	**Government**

Describe the measure	Joint Decision-Making Structure	Reward for Strong Collaboration Skills (Local Government)	Principle that Robots Do Not Replace Humans (Local Government / Corp. Manager)	Joint Decision-Making Structure	Reward for Strong Collaboration Skills (Local Government)
Why is it relevant?	Establish a joint decision-making structure involving developers, service providers, experts, and older users to review AI operation and design decisions.	Establish reward systems that promote collaboration skills.	Establish principles distinguishing human and robot roles and institutionalize them in service operations.	Establish a joint decision-making structure involving developers, service providers, experts, and older users to review AI operation and design decisions.	Establish reward systems that promote collaboration skills.
How can it be achieved?	AI intervention thresholds and recommendations involve value	Helps reduce resistance to robot adoption and	Prevents weakening of professional expertise,	AI intervention thresholds and recommendations involve value	Helps reduce resistance to robot adoption and

	judgments about safety, dignity, and well-being, not only technical considerations.	promotes collaboration.	accountability gaps, and mechanistic service delivery.	judgments about safety, dignity, and well-being, not only technical considerations.	promotes collaboration.
How can be assessed whether this measure has been fulfilled?	Create a governance body including developers, users, and experts and conduct periodic reviews of real usage cases.	Establish collaboration performance indicators and provide incentives such as recognition, bonuses, or training opportunities.	Define human–robot role boundaries, create operational rules, and train service providers on collaboration.	Create a governance body including developers, users, and experts and conduct periodic reviews of real usage cases.	Establish collaboration performance indicators and provide incentives such as recognition, bonuses, or training opportunities.
What are (potential)	Confirm the existence of formal committees and regular meetings.	Evaluate collaboration indicators and	Verify defined human–robot roles, operational	Confirm the existence of formal committees and regular meetings.	Evaluate collaboration indicators and

challenges to fulfilment?		collect staff feedback on collaboration experiences.	principles, and training completion records.		collect staff feedback on collaboration experiences.
What are risks if not fulfilled?	Multi-stakeholder governance may slow decision-making and older users may have difficulty participating directly.	Risk of bias in quantitative evaluation and possible disadvantages for workers with lower digital skills.	Increased training and operational costs and differences in technology acceptance among staff.	Multi-stakeholder governance may slow decision-making and older users may have difficulty participating directly.	Risk of bias in quantitative evaluation and possible disadvantages for workers with lower digital skills.
Which are the core function/role/ stakeholders responsible?	Technology-driven decision-making, user inconvenience, and reduced trust.	Lower acceptance of AI among service providers.	Reduced autonomy of older adults and increased technological dependency if human staff	Technology-driven decision-making, user inconvenience, and reduced trust.	Lower acceptance of AI among service providers.

			numbers decrease.		
Specific requirements?	Service provider management	Local government	Local government / corporate managers	Service provider management	Local government

Table 53: UC 9 – Practical measures to Promote of Social Well-being (3)

Organizational measures to promote of social well-being (3)			
Describe the measure	**Government**		
	Ethical Impact Assessment Framework (Government)	Reflecting Operational Results in Policy	Labor Standards for Social Service Providers
Why is it relevant?	To establish a comprehensive scope and implementation	Systematically document incidents and failures occurring during AI operation and use	Establish standards to ensure that workloads for social service

	methods for AI ethical assessments.	them as learning materials for policy and procurement decisions.	providers do not become excessive due to AI system use.
How can it be achieved?	Governments must ensure publicly funded technologies align with public interest and consistent evaluation standards.	Learning from operational outcomes improves governance and responsible technology deployment.	AI may create new tasks not captured in existing labor systems, increasing workload.
How can be assessed whether this measure has been fulfilled?	Define evaluation criteria and procedures for each ethical principle and establish dedicated assessment personnel or organizations.	Collect and share incident reports and integrate lessons learned into procurement and policy decisions.	Define new AI-related tasks, workload standards, and appropriate compensation systems.
What are (potential) challenges to fulfilment?	Document evaluation results and integrate them into procurement processes.	Review documented incident reports and policy updates based on operational results.	Verify official labor standards and assess workload levels.
What are risks if not fulfilled?	Lack of expertise or budget may reduce the assessment to a purely administrative process.	Data collection and analysis costs and coordination among stakeholders.	Budget constraints and diverse working environments.

Which are the core function/role/ stakeholders responsible?	Ethical controversies, service suspension, or accountability gaps in public services.	Repeated operational failures and weak policy learning.	Decline in care quality and increased technological dependency.

PRACTICAL MEASURES PROVIDED BY USE CASE 10

Table 54: UC 10 – Technical measures to protect post mortem rights

Technical measures to protect postmortem rights		
Describe the measure	Secure personal data management	Output/Export Control
Why is it relevant?	Personal data must be stored securely, anonymized where possible, and protected through encryption. Access should be restricted to relevant stakeholders only to ensure that the data of the deceased is not misused or used against their wishes. Legal heirs should retain the right to remove the data from the provider' s database.	The users should not be able to export or share conversational data or other generative AI outputs (voice recordings, video, etc.) in ways that would go against the wishes of the deceased, or which present the deceased in ways that are demeaning, exploitative, insulting, etc. Users should subscribe to a code of conduct to use the platform and service providers

		should limit the possibility to export conversational data.
How can it be achieved?	Ensure consensual use of data, data control, respect of relevant privacy laws, outlawing identity distortion, and regulating the deployment of RAG systems	Implementation of output/export controls.
How can be assessed whether this measure has been fulfilled?	Explicit consent forms (while alive or in will), ongoing user feedback, output/export control, audit studies	NA
What are (potential) challenges to fulfilment?	No explicit last wishes, LLM hallucination	Technical limits to fine-grained evaluation of the content being exported.
What are risks if not fulfilled?	Illegality, mailicious use of data, diffamation. User's side: greifbot could be baited by user into creating diffamatory content.	Harm to the deceased' s interest in preserving a certain self-image

Which are the core function/role/ stakeholders responsible?	Service provider, user	Service provider, Users
Specific requirements?	Terms of use	Terms of use

Table 55: UC 10 – Organisational measures to protect postmortem rights

Organisational measures to protect the postmortem rights of the deceased	
Describe the measure	Informed consent from the deceased
Why is it relevant?	informed consent to use personal and private data to create a griefbot should be secured before death. Consent should concern both the types of data that will be used to train and develop the system, and the purposes for which it will be used (to create a chatbot, a virtual avatar, for a given amount of time, etc.)
How can it be achieved?	Explicit consent forms

How can be assessed whether this measure has been fulfilled?	Institutional requirement.
What are (potential) challenges to fulfilment?	Lack of time.
What are risks if not fulfilled?	Violation of the posthumous wishes of the deceased.
Which are the core function/role/ stakeholders responsible?	Service provider.
Specific requirements?	No

Table 56: UC10 – Technical measures to achieve nonmaleficence and beneficence

Technical measures to achieve nonmaleficence and beneficence				
Describe the measure	Ongoing monitoring	Secure Conversational Data Management	Privacy by design	Transparency
Why is it relevant?	To support active support of the users' s well-being, continuous monitoring should include script revisions to flag signs of depression, overreliance, or other negative impacts on users.	Conversational data must be stored securely, anonymized where possible, and protected through encryption. Access should be restricted to relevant stakeholders only.	Respect for client privacy must be integrated into the system architecture and service agreements. This includes clear rules about ownership of discussion data and strict limits on secondary uses (e.g., commercialisation or advertisement).	The avatars should be explicit that they are not, in fact, reincarnations of the deceased, but should periodically remind the users that they are programs that do not have feelings or consciousness. It should be clear through interactions that postmortem avatars are programs imitating a

				person, not the person herself, either by designing the chatbot in this way, or by using it with oversight from a professional
How can it be achieved?	By monitoring time/frequency of use over time, redirect to human supervisor if show signs of prolonged grief disorder or self-harm, etc), by AI judges or human supervisors when needed.	Anonymization of the data, encryption, limit of access.	Internal rules and policy regulations.	Design of the bot and the program to provide disclaimers, periodical reminders, design of response style to be uncanny by design (refers to deceased in past sense, in the third person and with the conditional, etc.)

How can be assessed whether this measure has been fulfilled?	Ongoing human oversight and monitoring.		Ongoing monitoring and audit.	Initial design.
What are (potential) challenges to fulfilment?	Technical feasibility, available resources and expertise.	Technical feasibility, available resources and expertise.	Political inertia.	NA
What are risks if not fulfilled?	Harm to users.	Harm to users	Harm to users	Harm to users
Which are the core function/role/ stakeholders responsible?	Service Providers.	Service providers	Service providers, Governments.	Service providers, governments.
Specific requirements?	No			

Table 57: UC 10 – Organisational measures to achieve nonmaleficence and beneficence

Organisational measures to achieve nonmaleficence and beneficence			
Describe the measure	Professional oversight	Retirement protocols	Restricted use
Why is it relevant?	Users should be warned about the risks of griefbots and that they should use them carefully. Health professionals (psychologists, grief counselors) should be consulted when needed and users should be sensitized to signs of prolonged grief disorder (PGD)	Given that grief should typically be a process that is limited in time, service providers should establish retirement protocols for retirement and deletion of personal and conversational data after prolonged inactivity or after a set amount of time.	Use should be limited to consenting adult users or be used with continuous adult or professional supervision when they are used by children or adolescents. Particular attention should be given to how to regulate generic chatbots to limit the risks that griefbots be created "in the wild.
How can it be achieved?	Involvement of health professionals in design. Disclaimers.	Initial design of the griefbot. Ongoing monitoring of their use.	Age verification, disclaimers, terms of use.

How can be assessed whether this measure has been fulfilled?	Initial design.	Initial and ongoing design.	Initial contract with users.
What are (potential) challenges to fulfilment?	Resource availability	Resource availability.	Economic incentives
What are risks if not fulfilled?	Harm to users	Waster of resources, wrong to the deceased.	Harm to vulnerable users.
Which are the core function/role/ stakeholders responsible?	Service providers.	Service providers, governments.	Service providers, users.
Specific requirements?	No	No	No

Table 58: UC 10 – Technical measures to aim for justice

Technical measures to aim for justice		
Describe the measure	Use of diverse data sets	Reinforcement learning with human feedback
Why is it relevant?	When possible, the chatbot systems should be trained on culturally sensitive datasets that consider the language and cultural practices of the target populations.	Beyond including diverse datasets, the systems should be trained and tested with human feedback by explicitly aiming for cultural sensitivity. Where possible, human testers should be trained and sensitized to cultural differences within the Canadian populations.
How can it be achieved?	Identification of relevant datasets, ongoing monitoring, audit.	Reinforcement learning with human feedback; sensitivity training to human testers.
How can be assessed whether this measure has been fulfilled?	audit	audit
What are (potential) challenges to fulfilment?	Lack of resources and expertise	Lack of resources and expertise

What are risks if not fulfilled?	Harm to users and discrimination towards cultural groups	Harm to users and discrimination towards cultural groups.
Which are the core function/role/ stakeholders responsible?	Service providers	Service providers
Specific requirements?	No	No

Table 59: UC 10 – Organisational measures to aim for justice

Organisational measures to aim for justice		
Describe the measure	Community engagement report	Ethics committee involvement
Why is it relevant?	The service provider should gather data concerning the communities and profiles of persons using its services to assess variations between the needs of the populations using its services. Users should also be warned that this data will be gathered explicitly and be given the possibility to opt-out.	Ethics committees should be mobilized when needed to ensure that the service provider' s response to potential issues is properly informed.
How can it be achieved?	Ongoing monitoring, statistical data collection	Creation of ethics committees, identification or relevant ethical experts.
How can be assessed whether this measure has been fulfilled?	Quarterly reports.	Institutional assessments.
What are (potential) challenges to fulfilment?	Privacy violations, resources and technical expertise.	Resources and available expertise.

What are risks if not fulfilled?	Harm to users.	Harm to users, discrimination towards vulnerable groups.
Which are the core function/role/ stakeholders responsible?	Service providers	Service providers
Specific requirements?	No	No