D4.4 – Anticipated social implications for RF Holography in dynamic environments following privacy by design approach (WP4)

Grant Agreement Number	101099491
Action Acronym	HOLDEN
Action Title	Ethical Design of Holography with Dense wireless Networks (HOLDEN)
Funding Scheme	HORIZON-EIC-2022-PATHFINDEROPEN-01
Version date of the Annex I against which the assessment will be made	13/12/2022
Start date of the project	1/6/2023
Due date of the deliverable	31/5/2025
Actual date of submission	30/5/2025
Responsible	TWE
Contributors	CNR, ADANT
Dissemination level	Public

European Innovation Council

Funded by the European Union

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 or European Innovation Council and SMEs Executive Agency (EISMEA). Neither the European Union nor the granting authority can be held responsible for them. Grant Agreement No: 101099491.

Authors in alphabetical order

Full Name	Organisation	E-mail
Sage Cammers-Goodwin	TWE	s.i.cammers-goodwin@utwente.nl
Michael Nagenborg	TWE	m.h.nagenborg@utwente.nl
Stefano Savazzi	CNR	stefano.savazzi@cnr.it

Change History

Version	Date	Status	Author (Company)	Description
0.1	01/06/2025		TWE/CNR/ADANT	Version 1 Complete

Executive Summary

This HOLDEN (Ethical Holography of Dense Wireless Networks) report investigates the implications and impacts of Radio Frequency (RF) sensing in dynamic environments to establish appropriate specifications and technical means for implementation considering ethical and privacy constraints.

This report should be viewed in tandem with D4.3: (Identified privacy threats and counter measures). While D4.3 centers on the technology in its current state and technical mitigation measures to combat misuse and harm, this report steps further into the future to consider planned applications for the technology and how it might fit into and affect society.

In Section I we first introduce two features of dynamic sensing that differ from the static use case: machine learning and networked systems. These domains build a foundation from which this report grows from D3.4 (Privacy and Ethical Constraints in Static Environments). Based on mediation analysis and scenario development through public feedback, we investigate the implications for the sensing technology itself and society. Part of this analysis includes unveiling potential ethical challenges with the chosen application for Innovation II- Elderly Care.

Following, in Section II we apply our ethical framework to the dynamic sensing context. This involves applying the Elderly Care (and to a lesser degree other) use case(s) to Mediation Theories, Guidance Ethics, Techno-moral Scenarios, and Value Sensitive Design (VSD).

Finally, in Section III we go into further detail on the features that might shift the ethical and privacy outcome of the technology, providing possible design recommendations.

Table of Contents

Abbreviations

Abbreviation	Description
AALTO	Aalto University
AI	Artificial Intelligence
CNR	Consiglio Nazionale delle Ricerche
ESM	Ethics Status Monitor
HOLDEN	Ethical Holography of Dense Wireless Networks
ML	Machine Learning
RF	Radio Frequency
TET	Technological Environmentality Theory
TGT	Technological Gaze Theory
TMS	Techno-moral Scenarios
TMT	Technological Mediation Theory
VSD	Value Sensitive Design
WP	Work Package

1. Introduction

1.1. Overview

The European Union based HOLDEN (Ethical Holography of Dense Wireless Networks) Project is divided between three technical radio frequency innovations. This report details the second, dynamic imaging using WIFI systems. This innovation is under development by the HOLDEN team at Italy's National Research Council (CNR).

The first innovation was analyzed in D3.4 (Privacy and Ethical Constraints in Static Environments). As expected, some of the challenges are similar for both the first and second innovation and, therefore, similar mitigation strategies can be employed. To minimize overlap between the two reports, we focus on particular characteristics of the second innovation, namely the use of Machine Learning and the goal of creating interconnected networks capable of monitoring larger environments and infrastructures.

The use of Machine Learning (ML) for realizing Innovation II and III, leads to the necessity to engage with the recently established European AI Act and related frameworks. It is essential to build on the challenges and opportunities of ethics-by-design in RF sensing presented in D4.3 to make sure the system can ethically make decisions in acceptable contexts.

This deliverable focuses specifically on dynamic RF systems in the context of elderly care and more broadly on smart environments (including digital twins and industrial robotics). We discuss design opportunities in the spirit of Privacy-by-Design and Value-Sensitive Design (VSD). Further analysis of regulatory questions in the field of medical ethics will be addressed later in WP 9 in the context of palliative care.

We will also build upon and respond to issues identified in:

- D8.9 Ethics Status Monitor (ESM) Version 2

1.2. Description of Technology

The second HOLDEN innovation aims to identify human behavior (e.g., counting people, identifying if someone is standing, walking, walking fast, falling, or moving between rooms) in the built environment. Identification here refers to being able to detect a single individual in a dataset that might contain others. Only once an individual is identified as a separate entity from others in the dataset can activity recognition be applied. Identification here does not refer to uncovering personal data about the individual, although that might be possible through combining datasets or using biometric algorithms (e.g., using the data for gait recognition). In crowded environments identifying individuals is more difficult.

The system's performance depends on the number of receivers and antennas used. More devices will increase the accuracy, and complex environments require more receivers than simple spatial arrangements. Devices can be set up to operate across various frequencies to improve the performance.

Algorithms can be applied to both the access point and the receivers. The latter allows for setting up local instances to be combined to operate on a larger scale. Creating a scalable solution would enable

one to study a whole building or build digital twins. The impact of scalability is a key domain that is reviewed in this report.

Innovation II uses Machine Learning (ML) Algorithms to process the input signals and detect and classify events. The ML model is trained for real-time detection using simulated training data (generated by AI) and data from a test house. The consortium does this to generate diverse bodies in the simulated data without needing to train on people. There is a trade-off between increasing accuracy and reducing bias between varying groups. The measurement difference of two groups having positive results is called statistical parity. To safeguard fairness, we work to lower statistical parity. This is explained in further detail in D4.3.

1.3. Implementation Options

As demonstrated in D3.4 we can consider several design opportunities to achieve desirable ethical and societal outcomes. Within the Value-Sensitive Design approach, the focus of this report will be on "Technical Inquiry," that is: understanding the effects of design alternatives. Although we focus on the implementation of AI and the network capabilities of the technology, some considerations from D3.4 (wave density and type, location and positioning, penetration capacity, and passive tags) are still relevant depending on the application and use case.

1.4. Application – Elderly Care

While there are potentials for use in industrial robotics and facility management, in the context of this deliverable, we focus on the use case identified in D6.1 (Functional Requirements, Privacy Profiles for the Scenarios) for RF sensing in dynamic environments:

"A health monitoring system uses RF sensing to track the health and movement of an elderly person (primary user) within their home. This system operates without wearables, relying on RF technology integrated into existing IoT and Wi-Fi devices throughout the home for continuous, non-intrusive health monitoring, including movement detection, fall alerts, heart rate monitoring, and location tracking" (D6.1, p. 32).

In accordance with D6.1, we will also consider different locations and applications:

"The primary user resides at home or in assisted living. Secondary users may be remote (e.g., family members) or on-site (e.g., caregivers) and can access data via an app, dashboard, or telemedicine platform" (D6.1, p. 33).

Our currently considered use case in healthcare, thus, extends beyond the preliminary goals described in section 1.2. Health care applications have already been explored by other research groups, for example by Dina Katabi and her colleagues to monitor sleep [1] and behavioral symptoms associated with dementia [2]. The latter work is a reminder about the kind of services that can be built on a RF sensing system: RF sensing does not directly detect dementia in the brain, yet, through long-term monitoring and ML assisted data analysis, can detect dementia symptoms from specific bodily behavior. This may have far reaching consequences for the person monitored as well as their friends and relatives.

Focusing on the use case of "Elderly care" brings specific challenges in view of the AI Act. AI used in the context of healthcare may fall under the category of "high-risk systems" since it can be debatable

if “it does not pose a significant risk of harm to the health, safety or fundamental rights of natural persons” (EU AI Act, Article 6; see also [3]). Kolfschooten and Oirschot (2024) describe AI medical devices of risk class IIa and above “used for medical purposes of diagnosis, prevention, monitoring, prediction, prognosis, treatment, or alleviation of disease, injury, or disability” as high-risk under the AI act, and low-risk systems as those that are not *solely* used for medical purposes (thereby not falling under the EU Medical Device Regulation), such as applications in “wellbeing, health promotion, or activity monitoring” ([4], p. 2). One “concrete example” of a low-risk system given by Kolfschooten and Oirschot (2024) is “AI-based sensors used for assisted living for older people” ([4], p. 2).

Thus, it is important to consider what features HOLDEN wishes to design in a healthcare-oriented system. Under the AI Act, more the system promises, the higher risk it becomes, and the higher the risk, the more obligations need to be fulfilled. At this stage, given how much development the technology still needs to undergo, it is helpful to embrace a lower-risk implementation of the technology. Under a low-risk implementation, the AI Act requires AI literacy measures (Article 4) and Transparency obligations (Article 50) by the providers, meaning the developer of the system, in this case ADANT and the HOLDEN Consortium in addition to *the deployers*, those rolling out the system, for example the nursing home ([4], p. 3).

Patient rights are not a centralized under the AI Act, instead focusing the responsibility on the providers and deployers. Article 85 gives the right for anyone who suspects AI Act infringement to log a complaint and Article 86 gives the right to an explanation of decision-making. Article 86 however only applies to high-risk systems if the individual faced a breach of “health, safety or fundamental rights” ([4], p. 3).

Under the current iteration of the AI Act, low risk systems have few obligations. However, it is important to set the bar higher than the current legal regulation by embedding ethics into the design of the system, not over promising and not underdelivering. Moreover, some questions still need to be clarified in the AI Act itself and in our system implementation. These include clarifying who counts as someone belonging to a vulnerable group and what level of identification is acceptable.

It seems that using AI to uncover identity (e.g., using RF sensing for gate recognition) would fall under “unacceptable risks” under the AI Act thereby also causing GDPR concerns by generating personally identifiable information without notice or consent. For example, a patient might want to be recognized and monitored by the system, but their care providers, guests, and grandchildren might not want their data to be collected stored and analyzed. We imagine incidental biometric identification might be a concern for multiple sensor embedded AI systems under the AI Act. D4.3 addresses how we work to mitigate storing personally identifiable information while still achieving activity recognition using selective obfuscation via machine unlearning techniques.

As indicated earlier, we will leave the specific discussion about the potential use of RF sensing with vulnerable populations in palliative care to WP 9. Therefore, in the following, we will assume that the 2nd HOLDEN Innovation will fall under the category of “limited risk.”

2. Ethical Considerations from HOLDEN

In D3.4, we presented our holistic approach to ethics. In the following, we will discuss how the specific challenges of the second HOLDEN innovation and the questions arising from the specific elderly care use case can be integrated in the existing model **(Figure 1)**.

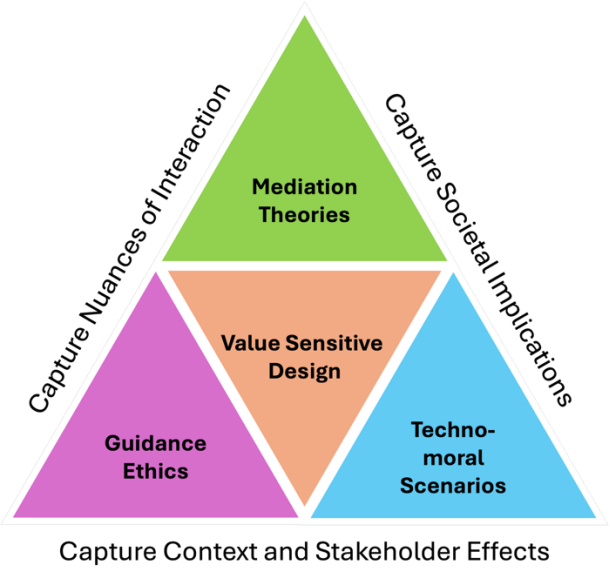

Figure 1 HOLDEN Approach to Addressing Ethical and Social Implications of RF Sensing

2.1. Mediation Theories

Technical Mediation Theory (TMT) [5], Technological Environmentality Theory (TET) [6], and Technological Gaze Theory (TGT) [7] anticipate and enable the responsible societal integration of ethically compliant RF sensing. This technology has the potential to reshape *lived experience* in space depending on the individuals understanding (or misunderstanding) of the technology and its capabilities. In addition, the potential of networked RF sensing systems could influence the *lived experiences of groups* as well as *group identities*. For example, following the Michel Foucault's account of surveillance [8], the envisioned system may separate the inhabitants of a care facility from the care-givers, the ones to be seen from the ones who see. In so far that the inhabitants have access to the same information as the caregivers, we may also assume that these technologies become technologies of the self, which, e.g., constitute the elderly person as a potential patient or as a fragile body at risk.

As has been demonstrated by [9], traditional approaches of Mediation Theory are not well-suited to address technologies such as ML due to their temporal structure, which shapes future activities based on behavior in the past. However, since we will focus on the question of (near) real-time identification of human behavior, the specific challenge caused by constant interactions with an adaptive AI system can be put aside from now.

In this deliverable we make use of these mediation theories to analyze our findings on public perceptions of Techno-Moral Scenarios (see Section 2.3).

2.2. Guidance Ethics

Peter-Paul Verbeek and ECP | Platform for the Information Society developed Guidance Ethics as an innovative approach to philosophy of technology. Three features define the Guidance Ethics approach: 1) operation from the *bottom up* – instead of basing guidelines on pre-existing philosophical literature, stakeholders articulate their needs and values, 2) instead of after the fact assessment, the technology is developed from the start with ethics in mind, and 3) As opposed to centering on worst case scenarios it focuses on *positive ethics* – determining how do we get what we want from the technology [10].

The idea of using HOLDEN applications for healthcare purposes was first officially proposed during our Guidance Ethics Workshop in 2023. At that time, "Ability to live at home longer for those aging or with disabilities" was ranked third out of 17 use cases determined over the course of the day-long workshop and "Long-term health monitoring" came in fifth in the rankings (See Figure 2). Since "Ability to live at home longer for those aging or with disabilities" was chosen as a top context by one of the groups, we have data from stakeholders present at the initial workshop as to how they determined proper use.

#	Use Cases (Contexts)	Group 1	Group 2	Group 3	Group 4	Median Ranking	Average Ranking
1	Built in Safety Measures for Working Environments	3	2	12	6	4.5	5.75
2	Automation (Heating Ventilation)	6	7	5	1	5.5	4.75
3	Ability to live at home longer for those aging or with disabilities	1	5	11	8	6.5	6.25
4	Crowd Monitoring (Festivals, events)	5	8	9	4	6.5	6.5
5	Long-term Health Monitoring	2	4	13	9	6.5	7
6	Indoor Navigation	7	9	2	11	8	7.25
7	Behavior Monitoring and Recognition	13	1	15	3	8	8
8	More advanced Smart Home Functionalities	9	10	8	7	8.5	8.5
9	Intrusion Detection	15	12	3	5	8.5	8.75
10	Facility Management and Monitoring	11	17	7	2	9	9.25
11	Gaming	17	6	4	12	9	9.75
12	Human Recognition for Robotics and AI	12	3	17	10	11	10.5
13	Baby Monitoring	8	16	6	15	11.5	11.25
14	Democratic Engagement through Global Gestures	4	14	10	17	12	11.25
15	Ubiquitous Surveillance for Law Enforcement	10	11	16	14	12.5	12.75
16	Embedded Art Systems	14	15	1	13	13.5	10.75
17	Ubiquitous Transparency (Like Police Body Cams)	16	13	14	16	15	14.75

Figure 2 Context Ranking from 2023 Guidance Ethics Workshop

Stakeholders who ranked the "Ability to live at home longer for those aging or with disabilities" first, thought that the technology could be used "continuously" to "track and monitor people in their daily life without investing in additional infrastructure." They envisioned a "variety of applications, like fall detection or recognizing abnormal behavior outside a normal routine." They noted that consent of all participants involved (patient doctor, etc.…) was essential. They thought that there was potential for "improved quality of life" and cost reductions within the health care system. However, risks were considered, such as dependency on the technology and a need for high reliability. Moreover, it might be possible to hack the home or room, and the technology could lead to "de-socializing."

Participants suggested that the technology could be reshaped to prevent risks and improve gains by having “personalized time window[s]” and “maintaining conventional health care (regular visits of the doctor).” They also suggested “using historical data to suggest future behavior.” Additionally, the technology could be regulated with “periodic control checks by [a] governmental institution” and “limit[ing] signal strength to certain areas.”

In the worst-case scenario, participants thought that “flaws, inaccuracies (or even total failure) in the system [could] prevent somebody from getting the necessary healthcare while solely relying on it.” But thought that overall, the technology would still be worth putting on the market due to “great benefits for large parts of the population.” They believed that the “risks are balanced out.” It is important to keep in mind that the Guidance Ethics workshop included HOLDEN researchers, so there may a more positive bias towards the technology.

2.3. Techno-moral Scenarios

Techno-moral Scenarios (TMS) are a useful tool for contemplating shifts in society, ethics and morals that may arise from a new socially disruptive technology. TMS are not predictions, but instead depictions of internally consistent worlds [11]. These worlds can be used to determine how we should shape the present to create a satisfactory future.

In April 2025, the HOLDEN TWE group hosted an interactive exhibit modelled after Techno-moral scenarios to understand public sentiment and appropriation of HOLDEN Innovations. The exhibit was called “This is Not a Camera: The Future of Monitoring through RF Waves” and was installed in three locations in Enschede, the Netherlands. The exhibit is described in greater detail in D2.3. Three HOLDEN specific innovations were chosen and presented as though they had been developing over the last decade. Patrons of the exhibit were asked to consider the technology in different contexts and vote on how creepy or cool they seemed. In the following, we will present an overview of the patrons’ answers and provide a brief analysis from the perspective of the mediation theories described in Section 2.1. A brief introduction to the framework is included in D3.4. Our analysis follows the outline by Verbeek and Rosenberg (2015), but we minimize use of jargon [12].

For the exhibit, the elderly care innovation was broadened into WIFI Care, “an integrated residential system providing non-intrusive insights and pre-emptive healthcare knowledge. By constantly collecting data on daily activity through WIFI monitoring, [it] assess[es] vitality, vulnerability to diseases like Parkinson’s, sleep health, and mental wellbeing. [The] system will recognize problems before you do!” Participants placed stickers indicating how creepy to cool they thought WIFI care was in three different contexts: rental apartments, elderly care facilities, and their homes (See Figures 3,4 and 5).

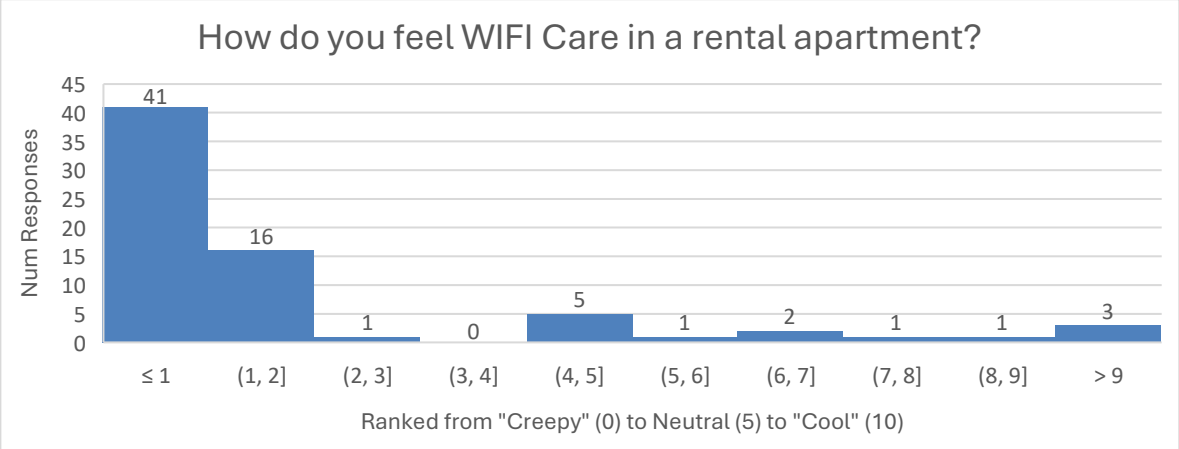

Figure 3 Histogram of responses for rankings of WIFI Care in rental apartments

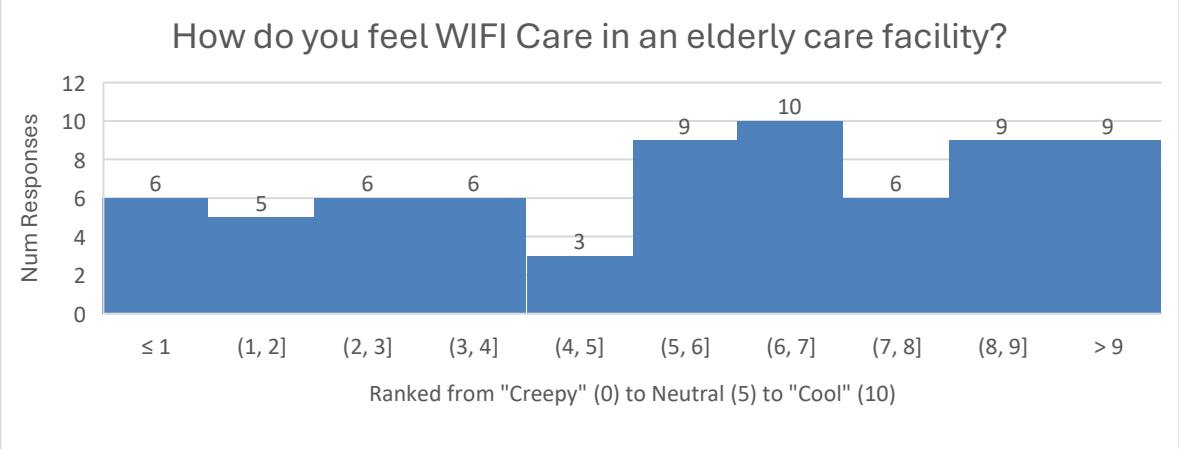

Figure 4 Histogram of responses for rankings of WIFI Care in elderly care facilities

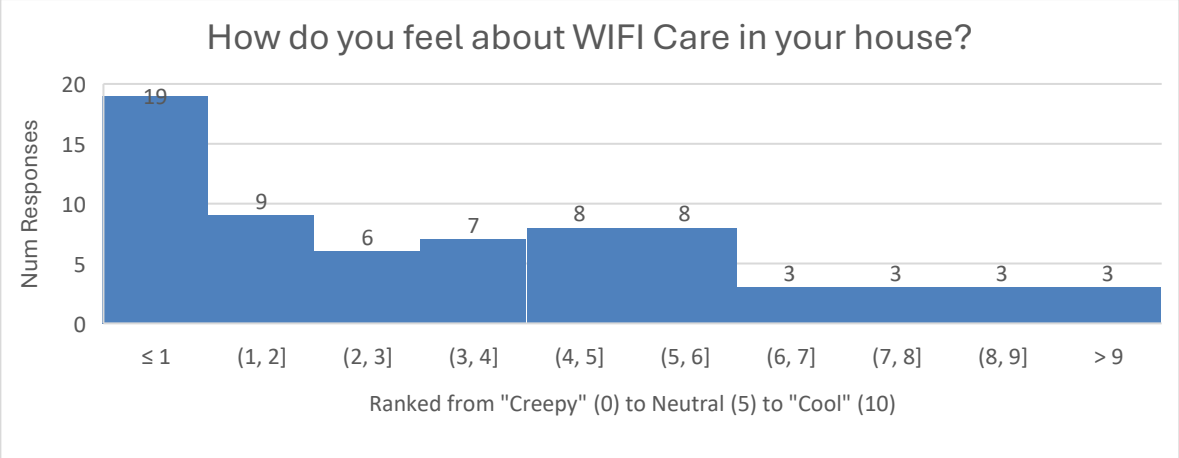

Figure 5 Histogram of responses for rankings of WIFI Care in your home

From the perspective of Mediation Theory, these differences can be explained by acknowledging that technological artifacts (such as RF sensing) only acquire meaning when used in a particular context. For example, it seems easier to understand WIFI Care as means to monitor persons (who might need support), while most participants found it hard to make sense of the same technology when being used – or, maybe, even integrated – in a rental apartment.

Later in the exhibit, participants had the opportunity to share how WIFI Care would impact a resident's life. Several hopes and concerns emerged. One shared, "Invasive, but relevant for better health tracking. Family can respond appropriately & doctors can take swift actions. IF I used it, then I would use it for healthcare without using wearables." From the perspective of Mediation Theory, it is interesting to note that the technology is understood as a way to guide the behavior of other people (family, doctors), while the patient is not mentioned. Thus, the technology is seen as a tool to facilitate actions of others but also seems to come with the risk of reducing the person under observation as a passive object of care. This observation may also explain why the participant would favor the system which does not include wearables.

Several people had positive hopes for the technology, sharing, "Easily warning of diseases, mental or physical," "It will ensure that if there are any anomalies, then the right people will be informed," "It allows [getting] information on […] health and [making] informed decisions," "It can help with the health of the resident," "Enable healthier lifestyle," "Easier more accessible care," and "Potential improvement of service quality + lower price. More profiling & data misuse."

From the perspective of Mediation Theory, we can see how the participants anticipate a hermeneutic relationship, in which "health" can be measured and turned into actionable information. It's noteworthy that at least one participant considers the opportunity that the information is also available to the inhabitant who can act accordingly. There are also two statements, which emphasis the "ease" of use, which could either be seen as a sign of the 'tyranny of convenience' or a desire for the technology to become invisible and to stay in the background.

Multiple people thought that the technology could alleviate elderly care needs, "It can make the residents life safer. There is also the potential that the life on an elderly person gets lonelier due to less personal care. Mentally, it's hard to predict if this is beneficial, it might depend on the health conditions of the individual. They might worry if they're not doing well more due to the tech." Another wrote "It depends on whether the resident has an illness/disability. Which is physical/mental, if so then I think WIFI-CARE can have a very interesting influence. For example, the health of the resident, well-being […] etc."

While the first group of statements rendered the inhabitants invisible, in these statements the technological systems is seen as a replacement for the caregivers, a system that improves the "health of the resident," but also a system that may further reduce the human-human interaction (a common concern also in the academic literature). We may also note that the technology is imagined to either focus on the inhabitant or informs the caregiver but is not considered as something that enables and forms direct interactions between the inhabitant and the caregiver.

It seemed that benefits inherently came with predicted consequences of dependency, loneliness and anxiety. One person shared that WIFI Care "probably [would] make it more efficient. But more control/insight isn't always better. I, myself am in burnout because I have too much control/insight into me. Sometimes just wing it, that's life. That's what separates humans from machines." Two wrote concern over bad healthcare data possibly having a negative impact on mental well-being. Several thought that WIFI Care could lead to less human-to-human connection – "They will miss real human contact! Someone who sympathizes. Real care from person to person. We don't need generators for that. People are […] enough. Distance between people makes us unhappy." Another wrote, "Less physical contact because it becomes less necessary to visit people. What if you only measure that people become unhappier after the system?"

These concerns resonate with our analysis about the lack of human-human interactions in the previous anticipatory statements. They also raise questions about how we relate to ourselves once we have all the information available, and it reveals the implicit hope that knowing about our current state of well-being will promote our well-being, even if it may very well be the case that inhabitants become lonelier and, thus, less happy. We also find a link to questions of politics and resistance, by expressing distrust in the rightful use of the data if they suggest a decrease of well-being.

Other visitors of the exhibition made even more explicit connections between the technology and larger political concerns. One said that they would tear out the system and move houses if it was in their home and they would only have such a system if it was not connected to a larger network and fully under their control. Someone else was concerned about the government possibly accessing the data. Another brought up that the technology is intrusive and relates to fascism and capitalism. – While we may not agree with such strong claims, we need to recognize that to some users, the technology foremost is considered as a (potential) tool of suppression.

To summarize our analysis, we can find that the participants understood the WIFI Care as something that may benefit either the inhabitants or the caregivers. However, the technology is not thought of as fostering a relationship between inhabitants and caregivers. While the technology is often understood as a means to monitor inhabitants' health, there are also concerns about reducing human-human interaction and losing control and access to the data, which brings about concerns about the political implications of such a technology.

2.4. Value-Sensitive Design / Privacy-Sensitive Design

Values Sensitive Design (VSD) was first introduced as a method in the field of Human-Computer Interaction and has been developed by Bataya Friedman and her colleagues since the late 1980s [13].

VSD remains the center of our approach. In this deliverable we focus on technological inquiry to provide a systematic overview of design alternatives and the differences between those alternatives. As in previous deliverables, we consider "Privacy-Sensitive Design" (aka, "Privacy by Design") as a particular form of VSD with a specific focus on the value of Privacy. (Historically, "Privacy by Design" has emerged as a separate field and, in parts, precedes VSD. For example, Goldberg, Wagner, and Brewer (1997) is an early contribution to Privacy-Enabling Technologies [14].)

While VSD in general is a well-established approach by now, the application of VSD to AI (including ML) is relatively new [15], [16], although Friedman and Nissenbaum did already address, e.g., new emerging biases [17]. While Fairness and Bias remain central topics in the discussion of responsible AI, this discussion mostly focuses on the processing of socio-demographic and personal-identifiable data. The acquisition and processing of such data is, however, not a core feature of the HOLDEN innovations.

For certain, in a specific-local context it might be relatively easy to combine personal-identifiable data with the insights enabled through HOLDEN innovation. For example, in an elderly home, inhabitants are likely to have access to private spaces. If an RF system RF system detects *somebody* falling in Mr. Smith's private room, it might be plausible to assume that Mr. Smith has been falling. The same applies to other forms of monitoring functions. However, we would like to argue that the combination of HOLDEN data and other data sources should be considered as a secondary use of such data. We will, therefore, focus on challenges which directly arise from the use of RF sensoring.

Consequently, we will take the findings in the field of the Ethics of Computer Vision as a starting point, which is in particular concerned with bias on the level of object detection and object identification [18], [19].

In the context of this report, "object detection" refers to the capability of a system to extract an individual object (e.g., a person) within an environment. "Object identification" refers to the capability of the system to identify objects as a specific object (e.g., "This object is a table" or "This object is a human person"). Once we distinguish between the levels of detection and identification, we can also address different kinds of bias and source of errors. At the level of detection, we can examine the systems capability to correctly capture individual objects. On this level, we are mostly concerned with visibility/invisibility. On the level of identification, we can focus on questions regarding the risk of misidentification.

2.5. Ethical Challenges

Although this report concentrates on two main features of focus for Innovation II, dynamic sensing (the use of Machine Learning and Networked Sensors), we acknowledge that there are other relevant ethical concerns that must be addressed. Fortunately, throughout the project we have kept track of potential concerns in the Ethics Status Monitor (D8.9).

The Ethics Status Monitor D8.9 organizes potential ethical risks by values that can be considered for each use case. Below we list considerations for the seven values highlighted in the ESM, which overlaps with our findings about HOLDEN Innovation I in D3.4.

1. **Equity and Equality** – It is important that Innovation II works the same across different groups and is not exclusionary, especially to those who may need it most. We are using statistical parity to measure that the system has similarly accurate response rates across groups. Furthermore, it is important that the technology does not further inequality by granting much improved outcomes to some but being completely inaccessible to others, or conversely by oppressing certain groups. The technology needs to continue being trained on different body types and environments. The issue did receive much attention in the Guidance Ethics workshop or in response to the Techno-Moral Scenarios.

2. **Autonomy** – Innovation II has the potential to influence the decisions of a provider and patient when used in a healthcare context, and the potential to influence a robot's actions in industrial robotics, and crowd control and lighting in a facility management context. These responses might impact an individual's autonomy as the system will dictate the truth and appropriate next steps. Just knowing that the technology is in place might also shift someone's behavior, for example enabling recklessness if they know that help will be called in case of emergency, or more careful if they think they will be chastised for not following medical guidance. Ideally the technology might be able to extend user autonomy, by giving them more freedom and control to make informed decisions. It is possible in a multistakeholder context that autonomy might be increased for some and reduced for others. – As we have seen, when confronted with the Techno-Moral scenarios, the envisioned system received mixed evaluations, since it could lead to better informed decisions, but also to information overload and paralysis.

3. **Privacy** – This technology is being designed to be privacy preserving by removing potentially privacy revealing information from the dataset. At the same time, the technology tries to

follow distinct individuals across a networked space and reveal actions such as walking, sitting, and falling. In a sense, the anonymous being is tracked throughout space, revealing where they are and what they are doing in environments where the action itself might have otherwise remained concealed. Is it really a face or name that intrudes on privacy or a revelation of an otherwise discrete series of actions? Moreover, there are potentials for the identity to be revealed by combining datasets. In some instances, such as in elderly care, it is important to make distinct the patient from the caregiver and only collect information on the consenting parties. To maintain privacy, it will be important to delete information that is not relevant to the use case, which might mean merging metrics in historical data to show averages instead of precise logs of individual actions. In our exhibition, (informational) privacy wasn't a major concern, which may have to do with the low expectations towards privacy in a care-related setting.

4. **Transparency** –Transparency is not just mandatory under Article 50 of the AI Act; it is also essential to uphold the prior considered values. If the system does not work well on all subsets of the population, then it is important that the system is transparent about the disparity so that it will be less likely to be covertly used in a way that could harm those groups. It is important for people to know the limitations of the technology and how it works, so that they can fairly make autonomous decisions based on the results. It is important that people know that the system is installed and running so that they can be aware if they are performing actions that they would prefer to remain private. The challenge is how to implement transparency of the system in public space. It may be easier within an app or closed system (such as an elderly care room) to place appropriate signage. Of course, it is entirely possible that someone misses the sign or may be unable to see. This is a larger issue in smart city research [20].

5. **Trustworthiness** – It is important that the system does what it says it is going to do or at least be transparent about how trustworthy it is. Many people assume that data-based systems are more trustworthy or accurate than human systems because data has the connotation of being quantitative and measurable and therefore factual. Although data always has had the potential for bias and inaccuracy, the risk is increased with ML systems because it is challenging to know how it came to its answer. It is important that the system does not abuse people's willingness to trust data-backed decision-making. – Trustworthiness is also a key enabler for the anticipated use in elderly care, where the participants of our exhibition expressed hope that the system would, e.g., guide the caregivers' actions in a timely manner.

6. **Sustainability** – Innovation II aims to work on pre-existing network infrastructure and can likely be operated via a software application on preexisting devices, this reduces the need for additional hardware systems to monitor helpful information. It is not yet clear what the energy toll will be from the ML. As we continue to develop the innovation, we should test the impact this would have in comparison to other interventions.

7. **Responsibility** – There is both an ex-ante and ex-post responsibility that derive from creating and putting a new product out on the market, especially a system that people will ultimately use for monitoring and decision making. If the ML algorithm is inaccurate, it will not be the fault of the AI's fault, but the fault of those who put the system on the market or due to misuse of the system, which may occur due to inadequate transparency.

The ESM also notes potential risks such as hacking, combining datasets, spying and inaccuracy (or malfunction). Limiting these risks will make the technology safer even if it is not misused.

3. Technical Considerations and Design Recommendations

To apply VSD it is necessary to identify morally relevant design alternatives. At the current stage, we consider the following alternatives shared in the introduction as relevant. In the first part, we will briefly review the technical considerations from D3.4. In the second part, we will explain design opportunities regarding ML and networked systems.

3.1. Review of Previous Considerations

RF Wave Density and Type: Innovation II aims to appropriate pre-existing networks to monitor and analyze behavior. Given that the system works better with higher wave density, either the system will work better in certain environments, or there will be pressure to increase the network capacity in monitored environments. If the system works throughout a network, there is also the potential for multiple access points to help each other interpret the data.

Location and Positioning: It might be that the location of the receiver and emitter are less relevant for Innovation II than Innovation I. This is because Innovation II makes quicks decisions using limited point clouds and machine learning as opposed to Innovation I that operates a bit more like a camera, using RF waves as a light source. Still, it might be helpful to test how closely receiver placement must match that of the virtual training environment, especially when moving the system onto the market, where it might not be installed precisely according to instructions.

Penetration Capacity: Despite its network potential, innovation II similarly aims to only work in approved environments. This would mean not collecting data from a neighboring room in the elderly care contexts or ignoring data collected outside of certain zones in an industrial robotic factory context.

Passive Tags: Passive tags could potentially be used to separate anonymous individuals, for example if some need specific access or should be tracked for longer or need to have their data deleted. Of course, passive tags also have a potential for misuse, especially if they have the power to shift responses from a system otherwise striving for neutrality.

Visualization: Innovation II does not aim to "see" in the same way as Innovation I. There are no images to generate but instead shifts in point clouds that are then classified by a machine learning algorithm. How the data is displayed might have a large impact on how well people understand how the data works. For example, there is the option to show what level of certainty the system has in a specific decision. There are also different choices to make in how a person is represented if movement is to be visualized. Are the avatars customizable? Do they reflect the sizes of individuals? Might that minimize the privacy of the system?

3.2. New Considerations

3.2.1. Machine Learning

"Value alignment" remains a central challenge in the development and deployment of Machine Learning, given the black box-nature of the system.[21] In simple words, we can control and know

the input and the output of a system, but it remains difficult to understand how the data is being processed. The focus is especially on biases, that is: the systematic and unfair discrimination against individuals and groups.[17] “Unfair” in this context can be defined as “for no good reasons.” For example, the performance of a fall detection system should not be influenced by certain bodily features of a person such as skin color. These concerns therefore also are connected to the considerations regarding “Equity and Equality” under 2.5.

Bias can occur and can be introduced at various steps of an ML workflow. For example, a biased outcome can be the result of a biased train set or a biased algorithmic (or both). Therefore, different measures can be taken at different stages. In HOLDEN, the focus is on auditing the outcome, but it also seems plausible to evaluate the (training) data or the algorithm itself. The latter could lead to turning to a form of explainable AI, which tries to overcome the black boxing of the algorithm (e.g., [22] on the relate field of computer vision). Explainable AI would also be helpful to address challenges such as Transparency, Trustworthiness and Responsibility (see, section 2.5).

Auditing of ML algorithms for bias detection: As described in D4.3 (Identified privacy threats and counter measures), the HOLDEN projects make use of selected quantitative ethics metrics (based on [23]) to detect and address biases, namely:

1. *Statistical parity* (to measure the difference between the probability of a prediction being positive between two different groups);
2. *Equalized odds* (to ensure equal prediction accuracy across different groups);
3. *Predictive equality* (to measures the accuracy balance by false positive rates), and
4. *Expected Calibration Error* (ECE) (to quantify the mismatch between the accuracy and the confidence of the model).

Additionally, it may be helpful to build in features that might give power back to the user. One approach might be to build in tools for *contestability*. [24] If the system is clearly incorrect in its analysis a user may contest the outcome and fix the data. This might be helpful to correct any shortcomings in the machine learning system as well. Finally, contestability also needs to be understood as a way to open up the blackbox of ML to the political discussion and to address the political questions, which were raised in our exhibition (see, section 2.3).

Auditing of the training data: Different kinds of bias in the training data can also lead to unfair outcome. In the academic literature different kinds of biases are distinguished.[25] Representation bias, for example, refers to the under-representation of a group in a data set. For example, a data set may include less people with disabilities and consequently, the accuracy for identifying people with disabilities is lower than the accuracy for detecting abled-bodied people. Representation bias is often hard to detect, because it requires a comparison with the ‘real’ world. It also often occurs when algorithms are transferred to a different local or social context. Therefore, at least a reliable documentation of the used data sets and, ideally, an evaluation of the quality of the data sets is required.

3.2.2. Networked Receivers

HOLDEN Innovation II has the potential to make use of networked systems to make platforms for wide scale ubiquitous sensing. This means that an individual could theoretically be tracked throughout a space that is owned and operated by one entity, for example an office building, factory,

amusement park, or elderly care facility. The amount of time and frequency that said individual must spend in those locations increases the likelihood that their anonymous data can be attributed back to them. While the data acquisition might be limited in a single location, the overall data may lead to the so-called “mosaic effect”[26] where the combination of sets of non-sensitive, non-private data leads to privacy-sensitive insights.

Additionally, the more time one *must* spend in a sensored environment, the less autonomy they have over whether they are monitored. Additionally, individuals who must engage with monitored environments have less time in true privacy, free from any technological gaze. If these systems become standard approaches for infrastructure management, then there is the potential to only be free from monitoring when literally “off the grid.” Given the popularity of smartphones and widespread use of surveillance cameras, many are already consistently identified and tracked by devices managed by third parties. How the data is used and stored and who has access to it, therefore will have a large impact on how much privacy an individual feels that they have.

Below we list some possible technical and policy considerations that may change outcomes for networked receivers.

1. *Network size moderation* Limit network size and how many networks can be connected together
2. *Historical data limitations* Generalize all historical data that is not necessary for the specific task the system is trying to achieve
3. *Regulation of running times* Limit times when sections of the network can be running
4. *Group Patterns* Recognize group patterns instead of individual people
5. *Decentralisation* In a networked system, we can decide where the data will be processed. In general, it would be preferable to minimize the amount of data exchanged between the nodes of a network. Thus, we need to consider alternatives to centralized infrastructures like edge computing.
6. *Contestability* Build in transparency so people can audit what is monitored by the system
7. *Regulations for Use* Maintain clear standards of what the technology is used for and can be used for. Similarly limit access so that the system may not be combined with secondary databases to reveal otherwise privacy sensitive information.

4. Conclusion

In this deliverable we addressed general and particular ethical challenges of HOLDEN Innovation II. We not only built upon findings from previous work packages and the Ethics Status Monitor but also included the results from our Appropriation Study via an interactive exhibition (Section 2.3). Additionally, we offered a first evaluation of HOLDEN Innovation II from the perspective of the European AI Act.

We demonstrate that most ethical considerations from previous work packages can also be applied to the second HOLDEN Innovation. In addition, we highlighted the challenges of Machine Learning and Networked Receivers, and gave some initial indications of how design decisions can help to promote a desirable outcome. These considerations still need to be verified and specified in future work packages.

5. References

[1] S. Yue, Y. Yang, H. Wang, H. Rahul, and D. Katabi, "BodyCompass: Monitoring Sleep Posture with Wireless Signals," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 4, no. 2, pp. 1–25, Jun. 2020, doi: 10.1145/3397311.

[2] I. V. Vahia *et al.*, "Radio Signal Sensing and Signal Processing to Monitor Behavioral Symptoms in Dementia: A Case Study," *Am. J. Geriatr. Psychiatry*, vol. 28, no. 8, pp. 820–825, Aug. 2020, doi: 10.1016/j.jagp.2020.02.012.

[3] E. Thelisson and H. Verma, "Conformity assessment under the EU AI act general approach," *AI Ethics*, vol. 4, no. 1, pp. 113–121, Feb. 2024, doi: 10.1007/s43681-023-00402-5.

[4] H. Van Kolfschooten and J. Van Oirschot, "The EU Artificial Intelligence Act (2024): Implications for healthcare," *Health Policy*, vol. 149, p. 105152, Nov. 2024, doi: 10.1016/j.healthpol.2024.105152.

[5] P.-P. Verbeek, "Toward a Theory of Technological Mediation: A Program for Postphenomenological Research," in *Technoscience and postphenomenology: the Manhattan papers*, J. K. B. O. Friis and R. P. Crease, Eds., in Postphenomenology and the philosophy of technology. , London: Lexington Books, 2015, pp. 189–204.

[6] C. Aydin, M. González Woge, and P.-P. Verbeek, "Technological Environmentality: Conceptualizing Technology as a Mediating Milieu," *Philos. Technol.*, vol. 32, no. 2, pp. 321–338, Jun. 2019, doi: 10.1007/s13347-018-0309-3.

[7] R. S. Lewis, "Technological Gaze," in *Perception and the inhuman gaze: perspectives from philosophy, phenomenology, and the sciences*, 1st ed., A. Daly, F. Cummins, J. Jardine, and D. Moran, Eds., in Routledge studies in contemporary philosophy. , New York, NY: Routledge, 2020.

[8] M. Foucault, *Discipline and punish: the birth of the prison*, 1st American ed. New York: Pantheon Books, 1977.

[9] J. J. Benjamin, *Machine horizons: post-phenomenological AI studies*. Enschede: University of Twente, 2023.

[10] P.-P. Verbeek and D. Tijink, *Guidance Ethics Approach: An ethical dialogue about technology with perspective on actions*. ECP | Platform voor de InformatieSamenleving, 2020. [Online]. Available: https://ecp.nl/wp-content/uploads/2020/11/Guidance-ethics-approach.pdf

[11] K. Bauer and J. Hermann, "Technomoral Resilience as a Goal of Moral Education," *Ethical Theory Moral Pract.*, vol. 27, no. 1, pp. 57–72, Mar. 2024, doi: 10.1007/s10677-022-10353-1.

[12] R. J. Rosenberger and P.-P. C. C. Verbeek, *Postphenomenological investigations: essays on human-technology relations*. in Postphenomenology and the philosophy of technology. Lanham (Md.): Lexington books, 2015.

[13] B. Friedman and D. G. Hendry, *Value Sensitive Design: Shaping Technology with Moral Imagination*. The MIT Press, 2019. doi: 10.7551/mitpress/7585.001.0001.

[14] I. Goldberg, D. Wagner, and E. Brewer, "Privacy-enhancing technologies for the Internet," in *Proceedings IEEE COMPCON 97. Digest of Papers*, San Jose, CA, USA: IEEE Comput. Soc. Press, 1997, pp. 103–109. doi: 10.1109/CMPCON.1997.584680.

[17] B. Friedman and H. Nissenbaum, "Bias in computer systems," *ACM Trans. Inf. Syst.*, vol. 14, no. 3, pp. 330–347, Jul. 1996, doi: 10.1145/230538.230561.

[20] S. Cammers-Goodwin and N. Van Stralen, "Making Data Visible in Public Space," *McGill GLSA Res. Ser.*, vol. 1, no. 1, pp. 1–32, Nov. 2021, doi: 10.26443/glsars.v1i1.120.

[21] B. Christian, *The alignment problem: machine learning and human values*, First edition. New York, NY: W.W. Norton & Company, 2020.

[22] M. Nauta, *Explainable AI and interpretable computer vision: From oversight to insight*. Enschede: University of Twente, 2023.

[23] G. Palumbo, D. Carneiro, and V. Alves, “Objective metrics for ethical AI: a systematic literature review,” *Int. J. Data Sci. Anal.*, Apr. 2024, doi: 10.1007/s41060-024-00541-w.

Legal references

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonized 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). PE/24/2024/REV/1. OJ L, 2024/1689, 12.7.2024, ELI: http://data.europa.eu/eli/reg/2024/1689/oj