Ref. Ares(2024)9258665 - 27/12/2024

All-in-one middleware for industrial human-robot-interaction

arise-middleware.eu

Coordinator

Consortium partners

Co-funded by the European Union

Document History

Ver.	Date	Description	Author	Partner
0.5	28.10.2024	ToC	Rita Latikka, Johannes, Anttila, Mariela Urra Schiaffino	Demos Helsinki
1.0	6.12.2024	First version for internal review	Johannes, Anttila, Rita Latikka, Tuukka Puonti, Philipp Andreewitch	Demos Helsinki
1.5	15.12.2024	Commented version by internal reviewer	Séverin Lemaignan	PAL Robotics
2.0	20.12.2024	Final version	Rita Latikka, Johannes, Anttila, Tuukka Puonti, Mariela Urra Schiaffino	Demos Helsinki

List of author(s), contributors(s) and reviewer(s)

Name	Partner (Acronym)	Contribution/Comment
Rita Latikka	Demos Helsinki	Main editor
Johannes Anttila	Demos Helsinki	Main editor
Mariela Urra Schiaffino	Demos Helsinki	Main editor
Tuukka Puonti	Demos Helsinki	Assistant editor
Philipp Andreewitch	Demos Helsinki	Assistant editor
Séverin Lemaignan	PAL Robotics	Internal reviewer
Manuel D. Martín	CARTIF	Internal reviewer

Publication Details

Grant Agreement Number 101135784

Acronym ARISE

Full Title	Agile, human-centric, and Real-tIme enabled SourcE technologies advancing industrial HRI in Europe
Topic	HORIZON-CL4-2023-DIGITAL-EMERGING-01-02
Funding Scheme	HORIZON Innovation Actions
Start Date	01/01/2024
Duration	42 months
Project URL	www.arise-middleware.eu
Project Coordinator	CARTIF
Deliverable	D3.1 Initial Open SSH Framework for human-centric and ethical HRI
Work Package	WP3 - Open SSH Framework for human-centric and agile HRI
Delivery Month (DoA)	M12
Version	2.0
Actual Delivery Date	31.12.2024
Type	Report
Dissemination Level	Public
Lead Beneficiary	Demos Helsinki
Keywords	SSH, human-centric, ethical, agile, human-robot interaction

D3.1 Initial Open SSH Framework for human-centric and ethical HRI

Table of Contents

List of Figures

List of Tables

Acronyms & Abbreviations

ACR.	Description	ACR.	Description
AI	Artificial intelligence	LLM	Large language models
AIS	Autonomous and intelligent systems	LEADOR	Leadership in ethical AI development and operational research
ALTAI	Assessment list for trustworthy artificial intelligence	MaM	Maturity assessment model
BGR	Blue, green, red	MR	Mixed reality
BoM	Bill of materials	PCB	Printed circuit board
CoCo	Coexistence-and-cooperation	PC	Personal computer
DPIA	Data protection impact assessment	PESIA	Privacy, ethical & social impact assessment
DPO	Data protection officer	RDT	Responsible disruptive technology
DT	Disruptive technology	REELER	Responsible ethical learning with robotics
EC	European commission	R&I	Research and innovation
EIA	Ethical impact assessment	SATORI	Stakeholders acting together on the ethical impact assessment of research and innovation
ELS	Ethical, social, legal	SGI	Smart graph interface
ETAPAS	Ethical technology adoption in public administration services	SHERPA	Shaping the ethical dimensions of smart information systems - a European perspective
EU-OSHA	European agency for safety and health at work	SIENNA	Stakeholder-informed ethics for new technologies with high socio-economic and human rights impact
FSTP	Financial support for third parties	SDP	Situated participatory design
GDPR	General data protection	SSH	Social sciences and humanities

	regulation		
HLEG-AI	High-level expert group on artificial intelligence	TEF	Test and experimentation facility
HRI	Human-robot interaction	TRL	Technology readiness level
HRC	Human-robot collaboration	VIRT-EU	Values and ethics in responsible technology in the EU
IEEE	Institute of electrical and electronics engineers	WIA	Well-being impact assessment
IEEE SA	IEEE standards association		

Executive summary

This deliverable provides an initial open ARISE SSH framework for human-centric and ethical human-robot interaction (HRI). Demos Helsinki, the social science and humanities (SSH) partner of the ARISE project, leads the development of the framework, which is ongoing throughout the ARISE project. The purpose of this deliverable is to describe the foundations and key elements included in the framework. We begin by explaining our approach and methodology. This includes introducing concepts of Industry 5.0 and human-centricity, as well as outlining our focus on key ethical, legal and social (ELS) considerations in HRI. We report results from mapping, analyzing and assessing similar projects, initiatives, and methodologies in this area, and provide an overall state-of-the-art of existing approaches. We introduce our transdisciplinary and qualitative research approach involving participatory, ethnographic and human-centric design methodologies, and highlight key preliminary findings from our empirical work. We then move to review the key elements of the initial version of the ARISE SSH framework. These elements are the ARISE HRI ethical principles, the ARISE ethics impact assessment, a collection of related industry standards and European regulation, collaboration and co-creation toolkit, typically encountered questions in AI systems, and a collection of relevant use cases with examples of their ethical concerns and questions. We discuss the evolving nature of the framework, the importance of contextual understanding in promoting human-centered and ethical technological innovation and give an overview of the various possible audiences of our work. We describe our goal to provide practical and evidenced-based recommendations for HRI development and policy in the following versions of the framework and conclude by outlining the key next steps in our work and the way forward.

Altogether, the framework is a structured approach to embed key ELS considerations into technological developments. It is intended to ensure industrial HRI projects are safe and consider a wide range of ethics related questions throughout the lifecycle of the technologies. Rather than a one-size-fit-all solution, the framework serves as an open toolkit for various audiences. The framework builds on the previous similar projects, initiatives, research and methodologies in this area. The development work of the framework is iterative in nature. The first iteration of this initial framework will be released at the end of 2026, and the final version of the ARISE SSH framework will be released at the end of the ARISE project in June 2027. This ongoing process enables us to monitor the broad landscape of the SSH aspects relevant to HRI throughout the project, iterate it based on our learning from collaboration with our technical partners and the financial support for third parties (FSPT) projects taking part in the ARISE open call programs. Finally, we remain open for discussions and collaboration also with those outside the ARISE project, to push our vision for more sustainable, human-centric and ethical technological innovations, work environments, as well as the future overall.

1 Introduction

1.1 Purpose and scope

This deliverable provides an initial open ARISE SSH framework for human-centric and ethical HRI. Demos Helsinki, the SSH partner of the ARISE project leads the development of the framework, which is ongoing throughout the ARISE project. The purpose of the document is to describe the foundations and the key elements of the initial version of the framework. Furthermore, we discuss the evolving nature of the framework, the importance of contextual understanding in promoting human-centered and ethical technological innovation, give an overview of the various possible audiences of our work, as well as explain our goal to provide practical and evidenced-based recommendations for HRI development and policy in the following releases of the framework. We conclude by outlining the key next steps in our work and the way forward.

1.2 Structure of this document

This deliverable describes the foundations and key elements of the initial ARISE SSH framework for human-centric and ethical HRI. The objective of this document is to give an overview of the undertaken and planned activities and results of this research and development work. Chapter 2 describes the approach and methodology. Chapter 3 describes the key elements of the ARISE framework. Chapter 4 is the discussion. Chapter 5 is the conclusion.

1.3 Intended audience

This deliverable is public. Therefore, the intended audience for this deliverable is all parties interested in SSH aspects in the context of HRI and technological innovation. These include, for example, the European Commission, consortium partners, researchers, practitioners and decision makers interested in the topic within the scope of the ARISE project. The framework will be used during the ARISE project and iterated further with ARISE partners and participating FSTP projects.

1.4 Contributions from partners

Table 1 summarizes the main contributions from consortium partners for this deliverable:

Table 1. The main contribution from partners.

Partner	Contribution
PAL Robotics	Internal reviewer of the deliverable
CARTIF	Internal reviewer of the deliverable

2 Approach and methodology

2.1 Industry 5.0 and human-centricity in HRI

The ARISE project envisions a near future which aligns with the principles of Industry 5.0, prioritising human-centric, resilient and sustainable work environments[1]. Rather than focusing on automation and technological advancements, which are characteristic of Industry 4.0, Industry 5.0 is a paradigm shift that places human values at the forefront of technological transformation. This approach emphasizes human-centricity, resilience, and sustainability, ensuring that technological progress serves well-being and societal impact.

Human-centricity, a core tenet of Industry 5.0, originates from design disciplines. It reflects an evolution from a techno-centric to a user-centric approach, but with a broader, more holistic perspective. A human-centric approach expands the concept of "the user" to encompass a more nuanced understanding of the human experience, including societal implications and ethical considerations[2]. It stresses intentional consideration of people's needs at every stage of design processes. Beyond adoption of particular methods or tools, it means that a project exhibits a distinct attitude, which translates in tangible engagements of care with the people and communities who are involved with or affected by the design outcomes in all phases, including the design process itself, while also questioning the exclusion of certain participants. Adopting a human-centric approach is not a technical task, and no single methodology assures the development of human-centric technologies. An engagement of care requires intentional work involving continuous interactions and discussions with different stakeholders to gain contextual insights, explore ideas, test and refine solutions, and evaluate outcomes.

The human-centric approach, which is fundamental to this value-driven transition, helps to ensure that collaborations between humans and machines enhance people's well-being at all levels. These levels include: 1) Empowering workers by automating repetitive tasks, allowing them to focus on more intellectually stimulating work (individual, micro level), 2) Advocating for a paradigm shift that prioritizes the human perspective, using technology and organizational structures as enablers (organizational, meso level), and 3) Designing technologies that embed human values from the outset, emphasizing principles such as trustworthiness and human oversight (societal, macro level).

In the context of Industry 5.0, the human-centric approach can be divided roughly in two levels: the technological level and the organizational level. At the technological level, the aim is to design, develop and implement technologies with human needs, values and experiences as the starting point—rather than pursuing technological advancements solely for their state-of-the-art potential—as well as to enhance interactions between humans and machines. Identifying technologies that influence human-centricity is also key to this initiative. For

[1] European Commission: Directorate-General for Research and Innovation, ERA industrial technologies roadmap on human-centric research and innovation for the manufacturing sector. Publications Office of the European Union; 2024. Available at: https://data.europa.eu/doi/10.2777/0266
[2] Ceschin, F., & Gaziulusoy, I. (2019). *Design for Sustainability: A Multi-level Framework from Products to Socio-technical Systems.* (First edition). Routledge.

instance, robotics is an area where the development process can focus on enhancing both physical and psychological safety, aligning with the core ideas of human-centricity. An emerging theme is humans and robots working together in a shared environment, presenting both opportunities and challenges[3]. At an organizational level, the human-centric approach emphasizes worker well-being. Although well-being is a broad area, it specifically focuses on the contribution and expertise of employees. This approach fosters inclusivity and encourages continuous learning and reshaping of skills that the employees already possess.

2.2 Focus on key ethical, legal and social considerations in HRI

In the ARISE project, our goal is to consolidate key ELS—ethical, legal and social—considerations in HRI within a unified framework. The project specifically focuses on HRI, while also considering related areas such as robotics and artificial intelligence (AI). This approach is particularly valuable for two main reasons: 1) ELS dimensions are often treated somewhat separately, and 2) existing frameworks tend to remain broad without focusing on specific contexts. Although this approach does not cover all possible perspectives, it is a vital step toward a more comprehensive framework by integrating these ELS aspects in the context of HRI.

Within the ARISE SSH framework, ethical considerations encompass principles and moral questions that guide actions and decisions. In practice, the focus is on ethical principles and impact assessment, as well as bringing ethical considerations already into early phases of technology developments. Furthermore, the value of initiating and maintaining broader level ethics-related discussions and raising ethical awareness in the context of HRI is acknowledged. Moreover, the goal is to turn higher-level ethical frameworks into practical tools, by utilizing contextualized knowledge of HRI workflows and practices gained through empirical research in the project.

Legal considerations focus on mapping and analyzing existing regulatory frameworks, some of which are legally binding. While the SSH partner does not possess legal expertise directly, viewpoints from the perspective of legislation relevant to HRI are integrated. Practical and evidence-based recommendations for policy will be one of the outcomes of the work, while relevant legal frameworks for HRI are continuously monitored and analysed. A strong focus is placed on the contextual understanding of HRI practices, which may also inform proposals for revisions of existing legal frameworks.

Social considerations are approached broadly including questions related to HRI's potential to transform human relationships, labor market and societal norms. This involves viewpoints such as experiences and perceptions of workers, automation of work and job displacement, training needs, technology's potential to augment and support human workers, as well as social acceptance of HRI and related technologies. The SSH partner's strong advocacy for participatory approaches and inclusivity supports the integration of social considerations into the ARISE project in practice.

---

[3] Demir, K. A., Döven, G., & Sezen, B. (2019). Industry 5.0 and human-robot co-working. *Procedia computer science*, *158*, 688-695. https://doi.org/10.1016/j.procs.2019.09.104

HRI is viewed as a transdisciplinary field focused on researching, designing and implementing systems in which humans and robots interact meaningfully, including aspects of how humans and robots exchange information (e.g., verbally, non-verbally, digitally) and share an environment without physical barriers between them. A similar, yet distinct concept is human-robot collaboration (HRC), which emphasizes HRI aimed at achieving a common goal[4].

2.3 Mapping, analyzing and assessing similar projects, initiatives and methodologies

The development of the ARISE SSH framework began with mapping, analyzing and assessing existing similar projects, initiatives, and methodologies. The search first resulted in 19 projects, which were analyzed and assessed from the perspective of their relevance to the ARISE project. In total, eight projects were chosen to be presented in the report either because they were similar to the ARISE project or offered useful methods or tools, which can be used within the scope of this project. These most relevant projects are listed in the Table 2 below. This is followed by description of the project's methodological outputs relevant to the ARISE project.

Table 2. Similar projects to ARISE from SSH perspective and their key characteristics.

Name of the project	Focus area of the project	Funder of the project	Link to project website	Main methods used in the project	Main deliverables of the project
SIENNA	Ethics (by Design)	Horizon Europe	SIENNA	Literature review	Frameworks, Impact Assessment
SHERPA	Ethics	Horizon Europe	SHERPA	Delphi study	Frameworks, Guidelines, Impact assessment
REELER	Human-Proximity Model	Horizon Europe	REELER	Ethnographic study	Roadmap, Human-Proximity Model, Toolbox
VIRT-EU	Impact Assessment	Horizon Europe	VIRT-EU	Ethnographic Study	PESIA framework, Ethics Primer and Stack

[4] Gervasi, R., Mastrogiacomo, L., & Franceschini, F. (2020). A conceptual framework to evaluate human-robot collaboration. *The International Journal of Advanced Manufacturing Technology*, *108*, 841-865. https://doi.org/10.1007/s00170-020-05363-1

AEQUITAS	Ethics	Horizon Europe	AEQUITAS	Experiments	Frameworks, Tools
Robotics4EU	Ethics	Horizon Europe	Robotics4EU	Societal Readiness Plan	Maturity Assessment Model, Chatbot, Robotics Compass
ETAPAS	Ethics	Horizon Europe	ETAPAS	Literature Review	RDT Framework, Code of Conduct
SATORI	Impact Assessment	Seventh framework programme of the European Community for research and technological development and demonstration activities (FP7)	SATORI	Foresight	Ethical Impact Framework

2.3.1 Existing frameworks, methods and tools

Ethical Principles and Frameworks

A comprehensive ethical framework was developed in the SIENNA project, in its deliverable report 4.7. entitled "An ethical framework for the development and use of AI and robotics"[5]. It proposes a set of methods and procedures for developing, deploying and using AI and robotics systems in a way that adheres to ethical principles. Taking into account earlier proposals of the Ethics Guidelines for Trustworthy AI presented by the High-Level Expert Group on AI (HLEG-AI)[6] and IEEE's Ethically Aligned Design framework[7], SIENNA introduces six high-level ethical principles for incorporating ethics into research and development of AI & robotics, which are based on the ethics guidelines by the SHERPA project, namely human agency, liberty and dignity; privacy and data governance; transparency; diversity, non-discrimination and fairness; individual, societal and environmental wellbeing; accountability.

For the development of robotics, SIENNA assumes the existence of shared practices or phases, described as the "V-model" and that can then be accompanied with guidelines for the

[5] Brey, P., Jansen, P., Maas, J., Lundgren, B., & Resseguier, A. (2021). SIENNA D4.7: An ethical framework for the development and use of AI and robotics technologies (1.1). Zenodo. https://doi.org/10.5281/zenodo.7266848
[6] Ethics guidelines for Trustworthy AI. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai
[7] IEEE Ethically aligned design. Available at: https://standards.ieee.org/wp-content/uploads/import/documents/other/ead_v2.pdf

incorporation of ethical considerations. This V-model approach is accompanied by a checklist, where the different phases of robotics development are measured against the six ethical principles mentioned before. Figure 1 illustrates the V-model. An overview of the V-model checklist is illustrated in Figure 2.

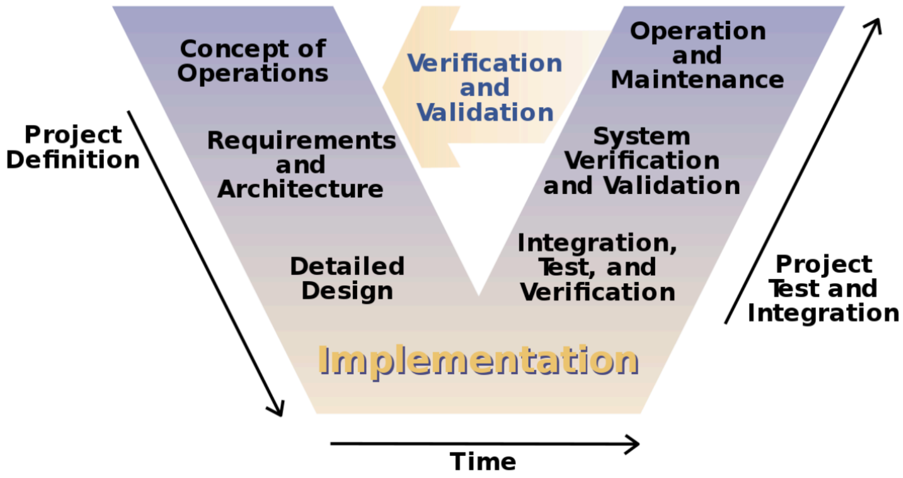

Figure 1. V-model from SIENNA framework.

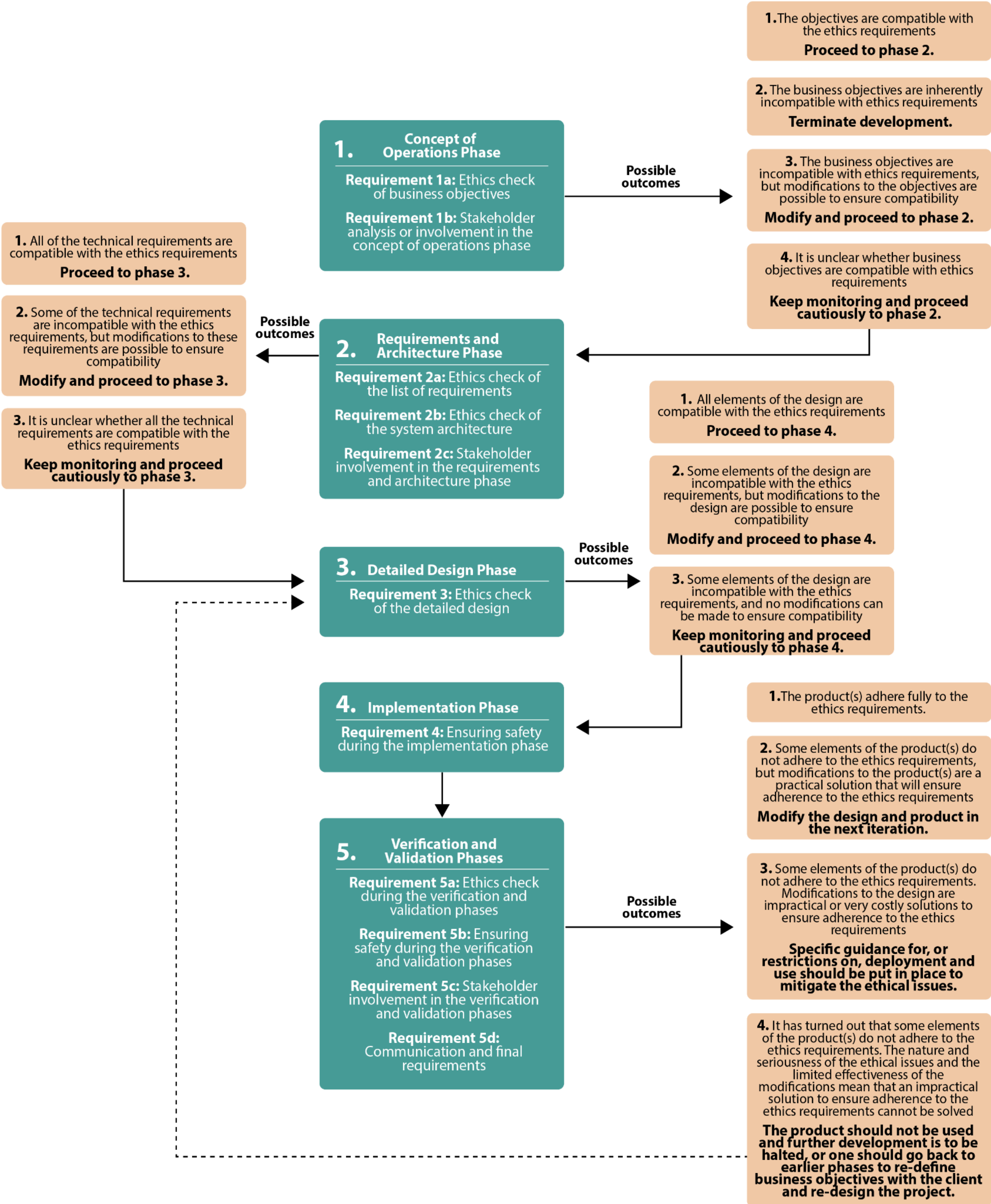

Figure 2. The V-model checklist adapted from the SIENNA framework.

The ETAPAS project, while focused on public administration services, offers a responsible disruptive technology (RDT) framework[8], including a Code of Conduct outlining ethical principles, an overview of potential ethical risks and social impacts of DTs, RDT indicators to measure these risks and impacts, and alignment with the European legal framework.

Ethics by Design

A relevant concept supporting human-centered approach is Ethics by Design, presented in Ethics By Design and Ethics of Use Approaches for AI framework[9], which aims to make people think about and address potential ethical concerns, while they are developing a system. Ethics by Design is a general model, which can be implemented to various development processes, like V-Model described above, or Agile[10].

The ethics by design model builds up on six phases, which are: 1) Specification of objectives, meaning the determination of what the system is for and what it should be capable of doing; 2) Specification of requirements, which means the development of technical and non-technical requirements for building the system; 3) High-level design, which is about development of a high-level architecture; 4) Data collection and preparation, which includes collection, verification, cleaning and integration of data; 5) Detailed design and development, which means the actual construction of a fully working system; and 6) Testing and evaluation, which includes the testing and evaluation of the system in context of the ethical requirements. These are not linear steps that follow another, rather steps that should be implemented into appropriate phases of the chosen development process. Figure 3 illustrates the six phases included in the Ethics by Design framework.

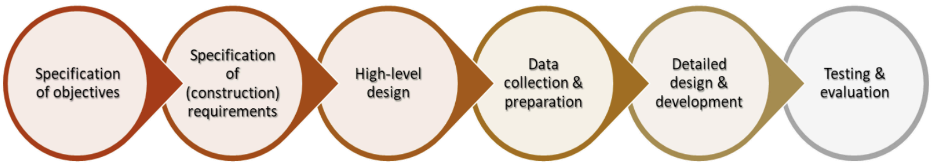

Figure 3. The six phases included in the Ethics by Design framework.

Human-Proximity Model

Within the REELER project, ethical concerns emerging from the disconnect between robotics developers and directly affected stakeholders were coined as "proximity gap". This entails a physical and cultural distance, for example in the form of values and motives, between those parties. One solution provided by REELER was to close these situation-specific gaps by seeking alignment. They suggest the introduction of alignment experts, whose tasks would involve spanning the space between the robotic bubble with their core experts and the

[8] Responsible disruptive technology framework. Available at: https://onedrive.live.com/view?id=8362975DAACA6A1!1320&resid=8362975DAACA6A1!1320&authkey=!APpu8T8bV-nocX4&wdEmbedFS=1&wdo=2&cid=08362975daaca6a1
[9] Ethics By Design and Ethics of Use Approaches for Artificial Intelligence. Available at: https://ec.europa.eu/info/funding-tenders/opportunities/docs/2021-2027/horizon/guidance/ethics-by-design-and-ethics-of-use-approaches-for-artificial-intelligence_he_en.pdf
[10] Agile manifesto. Available at: https://agilemanifesto.org/

affected stakeholders' communities to identify their separate motives, and to communicate them to one another in a move toward alignment and collaborative learning. Figure 4. Illustrates the human-proximity model[11], which is included in the broader awareness-raising toolbox by the REELER project.

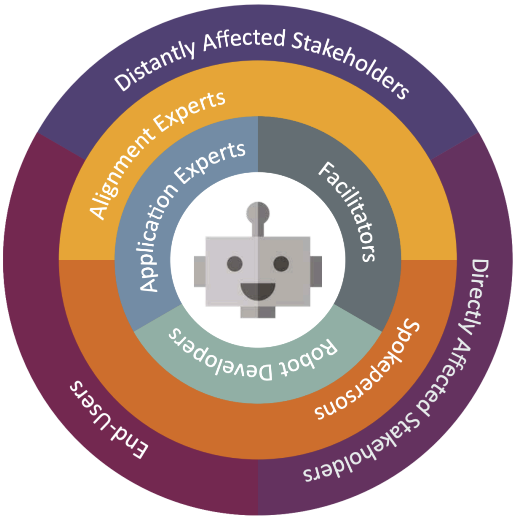

Figure 4. The human-proximity model by the REELER project.

Impact Assessment

The ethical impact assessment (EIA) mentioned in the SIENNA project derives from the SATORI project, which provides a comprehensive EIA framework[12], that is a structured methodology for evaluating the ethical implications of research and innovation (R&I) projects. The framework consists of six stages: 1) Threshold analysis: Determines if ethical issues are present that require a full EIA; 2) EIA plan: Developed if threshold analysis identifies ethical issues (including team, budget, stakeholder engagement); 3) Ethical impact identification: Identifies potential ethical impacts using literature reviews, foresight methods (e.g., horizon scanning, Delphi), and stakeholder input; 4) Ethical impact evaluation: Evaluates the severity, likelihood, and importance of ethical impacts; 5) Remedial actions: Develops recommendations and design interventions to mitigate negative impacts and enhance benefits; 6) Review and audit: Ensures accountability through regular reviews and audits of the EIA process.

VIRT-EU has developed a Privacy, Ethical & Social Impact Assessment (PESIA)[13] questionnaire. PESIA is designed to reflect upon, evaluate and take into account data protection, security and privacy aspects of new technologies, as well as ethical and social concerns. VIRT-EU also offers a multitude of tools, such as the Ethical Stack[14], an interface designed to structure and facilitate ethical considerations in the development of new connected technologies. This is a series of tools to support creators of new connected technology to reflect on their products'

[11] The human-proximity model by the REELER project. Available at: https://reelertoolbox.ab-acus.com/
[12] SATORI EIA framework for research and innovation projects. Available at: https://satoriproject.eu/media/CWA17145-23d2017.pdf
[13] The PESIA framework of the Virt-EU project. Available at: https://www.virteuproject.eu/servicepackage/pesia/
[14] The Ethical Stack of the Virt-EU project. Available at: https://www.virteuproject.eu/servicepackage/ethical-stack/

ethical and social impacts. Furthermore, VIRT-EU has created an Ethical Tools Index[15], that is a list of existing ethical tools.

Maturity Assessment

The Maturity Assessment Model[16] (MaM) being developed by Robotics4EU is a tool designed to evaluate the societal readiness of robotics solutions. In this context, it highlights to what extent the robot meets society's ethical values and its economic, legal and social needs and encompasses common ethical principles, protocols, and best practices, aiming to harmonize ethics assessment across various scientific fields and organizations. Furthermore, Robotics4EU has developed the Responsible Robotics Compass[17], an online tool designed to assess the non-technological aspects of responsible robotics.

Bias Mitigation

Initially, the AEQUITAS project aimed to address various manifestations of bias and unfairness in AI by offering a controlled experimentation environment for AI developers through an online testing platform. Beside it, they offer various tools and methodologies, which can be helpful in this project's scope: fair-by-design guidelines and bias awareness and measurement tools. The project is still ongoing and some deliverables will only be available in the future.

2.3.2 Overall state of the art of existing approaches

The research of existing projects showed that the European landscape of ethical frameworks and tools is broad and varied. Particularly, since the funding through the Horizon 2020 program, the interest in ethical concerns and solutions around AI and robotics has picked up and is reflected in the funding of the project analysed in this section: most of them were funded by the Horizon program.

The range of projects offers different ethical frameworks, which also build upon each other, for instance, the SIENNA framework uses SHERPA's and and IEEE's Ethically Aligned Design approaches and refines them. The ethical principles discussed in the different frameworks resemble each other, with the most common principles being: Agency, liberty, dignity, privacy, data protection, fairness, diversity, well-being and accountability. Along with the frameworks, different practical methods and tools for developing and implementing ethics in AI and robotics were presented: Ethics By Design approach, the human-proximity model, impact and maturity assessment and bias mitigation. Together, the multiple frameworks, methods and tools offer developers of robotics theoretical and practical ways of taking ethical considerations into account.

[15] The Ethical Tools Index of the Virt-EU project. Available at: https://docs.google.com/spreadsheets/d/1Tv1qtlGug99h2-G34CLKhpjIUpvMr56f/edit?gid=2070281166#gid=2070281166

[16] The maturity assessment model: https://robotics4eu.eu/wp-content/uploads/2021/08/D1.3.Maturity_Assessment_Model_final.pdf

[17] The Responsible Robotics Compass of the Robotics4EU project. Available at: https://robotics4eu.eu/robocompass/

The projects and their outputs presented above are not of course an exhaustive list of existing frameworks, tools and recommendations relevant to the scope of the ARISE project. There are various initiatives, ongoing research and development, especially when considering the rapid development of AI and the associated regulatory discussions in Europe and around the world. Some international initiatives that have gained attention in European landscape include a guide to AI ethics by the Alan Turing Institute[18], ACM Code of Ethics and Professional Conduct[19], AI4People's Ethical Framework for a Good AI Society[20], UNESCO recommendation on the ethics of AI[21], OECD AI Principles[22], Montreal Declaration for Responsible Development of AI[23], as well as work of CEN-CENELEC AI Focus Group[24], and IEEE Standards Association (IEEE SA)[25]. Specifically in Europe, current key frameworks include the Ethics Guidelines for Trustworthy AI by HLEG, and Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment[26], alongside the Ethics By Design and Ethics of Use Approaches framework, which we have also included into our frameworks' collection of related European regulation and industry standards. Common to these approaches is that they are centered around establishing ethical standards and guidelines for the development, deployment and use of AI, encouraging innovation within a clear, regulated context. In the European ecosystem, an important initiative is also AI, data and robotics (ADRA)[27], which, among other things, supports educational resources and materials around its topics.

Keeping track of this extensive field, in terms of the development of current models and new initiatives, is part of the development work on our own ARISE SSH framework. In addition to our work building on a wide range of existing approaches, we aim to deepen our understanding of their uses and means for their improvement for the context of HRI. This is done through transdisciplinary and qualitative research we conduct throughout the project.

2.4 Contextual knowledge through ethnography

Our transdisciplinary and qualitative research approach combines participatory[28], ethnographic[29] and human-centric design methodologies to understand the practices,

[18] Leslie, D. (2019). Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute. https://doi.org/10.5281/zenodo.3240529
[19] ACM Code of Ethics and Professional Conduc. Available at: https://www.acm.org/binaries/content/assets/about/acm-code-of-ethics-and-professional-conduct.pdf
[20] The AI4People's Ethical Framework for a Good AI Society. Available at: https://eismd.eu/featured/ai4peoples-ethical-framework-for-a-good-ai-society/
[21] UNESCO recommendation on the ethics of AI. Available at: https://unesdoc.unesco.org/ark:/48223/pf0000373434
[22] OECD AI Principles. Available at: https://www.oecd.org/en/topics/ai-principles.html
[23] Montreal Declaration for Responsible Development of AI. Available at: https://declarationmontreal-iaresponsable.com/la-declaration/
[24] CEN-CENELEC AI Focus Group. Available at: https://www.cencenelec.eu/areas-of-work/cen-cenelec-topics/artificial-intelligence/
[25] IEEE Standards Association. Available at: https://standards.ieee.org/
[26] Assessment List for Trustworthy AI (ALTAI) for self-assessment. Available at: https://digital-strategy.ec.europa.eu/en/library/assessment-list-trustworthy-artificial-intelligence-altai-self-assessmen t
[27] AI, data and robotics (ADRA) ecosystem. Available at: https://adra-e.eu/
[28] Schuler, D., & Namioka, A. (Eds.). (1993). *Participatory design: Principles and practices*. CRC press.
[29] Brewer, J. (2000). *Ethnography*. McGraw-Hill Education (UK).

workflows, and processes of HRI, as well as its professionals' values, aspirations and needs in relation to SSH perspectives. We also seek to understand what the current and possible future practices are where SSH aspects, and ethics considerations particularly, could be embedded[30]. Our research emphasizes the practical day to day work[31] of the ARISE project—a co-researching approach to ethics[32] that brings together the SSH experts, the ARISE partners and their varied technical expertise, and in later stages the FSTP projects. In practice, our research activities have thus far included interviews and workshops online and offline, as well as onsite visits to our test and experimentation (TEF) partners.

In the spring 2024, alongside a literature review, we conducted initial interviews with our project partners to scope the HRI professionals current use and familiarity of existing frameworks and tools in HRI development. In these approximately 30-60 minutes long semi-structured online interviews we discussed our technical partners' views and experiences of ethics and SSH aspects in HRI. The goal of these initial interviews was to exchange information and direct the initial framing of the SSH framework. In autumn 2024, we conducted a series of online workshops with our partners focusing on ethics. In these approximately hour-long meetings, we discussed the Ethics By Design principles, and how they are reflected in our everyday work in the ARISE project and beyond. In autumn 2024, we visited all our four TEF partners to continue investigating the practices, workflows and processes of HRI, as well as its professionals' values, aspirations and needs in relation to SSH perspectives. During these visits, we explored what worked well, identified best practices, and examined ways to embed existing ethics frameworks and processes in HRI development to ensure that the guidelines, tools, specifications and recommendations included in the ARISE SSH framework are relevant, understandable, and fit for purpose. The qualitative data collected includes interview transcriptions, field notes, pictures, as well as the tacit knowledge gained through spending time in the research context. We have content analyzed[33] the data to report our preliminary findings in this deliverable. We first report the key findings from the initial partner reviews and then showcase the preliminary insights from the onsite visits.

2.4.1 The mismatch between the supply of existing approaches and their use

Despite the proliferation of the existing frameworks, tools and methods around ethics and human-centricity in AI and robotics, our initial interviews and work with the technical partners in the ARISE project pointed to a gap in the number of these tools and their use in practice. Although in our initial scoping the partners could very readily identify and elaborate on the sorts of ethical questions and concerns related to the TEF use cases—with much the similar

[30] Elhadj, E., Van Horenbeeck, Z., Lievevrouw, E., & Van Hoyweghen, I. (2024). Brokering responsible research and innovation in in silico medicine. *Journal of Responsible Innovation*, *11*(1), 2414484. https://doi.org/10.1080/23299460.2024.2414484
[31] Pols, J. (2024). Making things specific: towards an anthropology of everyday ethics in healthcare. *Medicine, Health Care and Philosophy*, 1-11. https://doi.org/10.1007/s11019-024-10204-z
[32] Banks, S., Armstrong, A., Carter, K., Graham, H., Hayward, P., Henry, A., ... & Strachan, A. (2016). Everyday ethics in community-based participatory research. In *Knowledge Mobilisation and Social Sciences* (pp. 97-111). Routledge. https://doi.org/10.1080/21582041.2013.769618
[33] Krippendorff, K. (1989). Content analysis. *International encyclopedia of communication*, *1*(1), 403-407.

results that a more process oriented tool approach would have led to—there did not seem to be existing standardized processes or particular ethics related tools and processes that were in regular use within the TEFs. Within the scope of EU-funded projects, where especially in the newer projects there were clear guidelines to set up and follow such processes, partners did describe similar ethics-related reviews and processes as the tools described above prescribe. Most described developing an "ethics sensibility" across the organization and some saw it as a core competence to develop further. The use of such processes and building ethics-related capacities was especially true in EU-funded projects where this was seen as a requirement for funding, whereas anecdotal evidence suggests that similar requests outside the legal requirements were not as much a part of privately funded projects. Altogether, the initial partner interviews indicated a mismatch between the supply of existing frameworks and their utilization. This observation underscores the rationale for our approach to develop the ARISE SSH framework based on existing approaches and resources and refine them for the context of HRI, rather than creating something entirely new solely for the sake of novelty.

2.4.2 Toward a contextualized, human-centric and ethical HRI framework

In autumn 2024, we continued to investigate the relationships between HRI practices, SSH, and ethics perspectives. We visited the TEFs to research the topics together with our partners, who introduced us to their work practices, processes, and environments. Based on the qualitative data collected, we identified five key findings related to contextualized, human-centric and ethical HRI, which we will review next. These insights, along with further collaboration with both the ARISE partners and the FSTP projects, will continue to inform the design of the ARISE SSH framework and its associated components.

Responsibilities for ethics are hazier and more unclear at lower technology levels
Many ethics frameworks and tools take as a starting point that the developers involved know their end users very well and that the technology readiness level (TRL)[34] of the developed product is relatively high. In contrast to this, in the context of ARISE much of what is being developed is not at the highest TRL levels and rather seen as a "platform" or prototype of something to be iterated further. At the same time, the solutions being developed are an assemblage of different technologies, with a distributed network of people working on those and also bringing them together. This reality poses difficult, but not impossible questions for considerations of ethics, responsibility and accountability: Design choices affect the affordances of the technological solution up to a point, but also leave open a wide range of uses of which the lower TRL levels might not have certainty. How could questions and considerations of ethics be embedded also in earlier TRL levels? Are there design choices that can be made to preclude certain unethical uses already at the lower TRL levels?

[34] TRL assessment. Available at:
https://ec.europa.eu/research/participants/data/ref/h2020/other/wp/2018-2020/annexes/h2020-wp1820-annex-g-trl_en.pdf

Not even clear principles are clear-cut: There are everyday balancing acts in the design process

One of the key wishes from those working on HRI projects was clarity and usability also when it comes to frameworks, guidelines and tools around ethics. This is a very understandable wish: The more clear-cut and step by step something is, the easier it is to use and to communicate. Yet, concrete situations in the design processes of HRI show that even "simple" things like clear ethical principles are not really that simple and in fact even commonly accepted principles can be in contradiction with one another. An example of this balancing act came through one of the ARISE partners: On one hand, automatically correcting for a safe and ergonomic position would be beneficial to the health of an operator in the long run, but at the same time this sort of automatic control by a robot of an operator's position would be a clear infringement of human autonomy. In this case valuing autonomy was a clear choice, but what are the ways to better prepare developers and designers to discuss and make decisions on such dilemmas within the context of everyday work?

Modules, models and data encode a way of seeing the world and need careful consideration

The now already burgeoning literature and empirical evidence on the existence and effects of bias in datasets related to algorithmic systems is nothing new and is also well recognized by both technical and SSH experts alike. Attention to bias and a thorough understanding of what are its effects is key not only in large language models (LLMs) integration and the use of natural language for operation, but also for computer vision related tasks from face recognition to emotion recognition. The same holds true for different kinds of standards and modules as well and how they are used. Has enough attention been used on whether the HRI solution recognizes different kinds of accents or voices? Is there a danger that the solution has difficulty in recognizing darker skin tones on hands and this could place an operator in danger?

Clear processes and responsibilities for ethics questions ensure inclusion of different perspectives in different project phases

In general, the ARISE partners had also had very good experiences within the scope of previous European projects on processes related to ethics and ethics assessment. These included cases with clear processes and accountabilities, a designated responsibility within each partner organization and use case for ethics and the inclusion of not only project management or separate ethics people but also technical experts in discussions on ethics and ethics assessment. A balancing act was also present in the context of these good experiences: How to ensure that the process is not reduced to a naive checklist that someone will just routinely click through? But on the other hand, how to ensure that the process is clear and simple enough to not place an oversized burden on people while ensuring that it creates sufficient space for reflection?

Finding a common language and "co-contamination" is key – but takes time

Overall, the considerable amount of time taken for discussions with the SSH experts and the technical experts within the ARISE project has felt very fruitful for those involved. Multidisciplinarity in projects is never easy—there are different backgrounds and expertises at play and their incorporation takes time. Finding this sort of "common language" and a shared ground on which to discuss questions of ethical and societal implications of HRI projects is not easy or fast, but we found that the onsite visits were especially helpful for creating this shared understanding and spaces for in-depth discussions. The TEF feedback of being encouraged to think critically about SSH and ethics-related aspects during the visits and their related workshops and interviews—a "co-contamination"—was also encouraging and feels like it has created fruitful ground for further engagement on the development of the SSH framework.

3 Key elements of the ARISE SSH framework

Based on our preliminary findings from the literature review and our empirical research during the first year of the project, the next section presents the key elements of our own framework—the initial open ARISE SSH framework for human-centric and ethical HRI. These elements are the ARISE HRI ethical principles, ARISE ethics impact assessment, a collection of related industry standards and European regulation, collaboration and co-creation toolkit, typically encountered questions in AI systems, and a collection of relevant use cases with examples of their ethical concerns and questions. Considering the wide range of well-researched and designed ethics and SSH frameworks and tools, as well as the seemingly limited use of or familiarity with them, we focus on compiling and modifying existing approaches. Our aim is, when applicable, to make the existing resources more suitable for their potential users. We also seek to contextualize the existing ethical frameworks and tools into a process that would at the same time ensure discussion of ethics and SSH aspects in HRI.

3.1 ARISE HRI ethical principles

Ethical principles within the context of HRI have been previously developed for example in the Ethics By Design guidelines, the SHERPA and SIENNA projects, largely based on the work done previously by the AI HLEG. The consolidation of these principles into a similar format has also had practical implications, with the principles further serving as a base for the creation of assessment tools for ethical aspects of robotics and AI—for example the Ethics By Design and Ethics of Use Approaches for AI that were commissioned by the European Commission (EC), summarized in Table 3. In this approach, each one of the principles is accompanied by a set of general ethical requirements for AI and robotics systems.

Table 3. General level ethical principles in the Ethics By Design approach.

General level ethical principles in the Ethics By Design approach
1. Respect for Human Agency: Human beings must be respected to make their own decisions and carry out their own actions. Respect for human agency encapsulates three more specific principles, which define fundamental human rights: autonomy, dignity and freedom.
2. Privacy and Data governance: People have the right to privacy and data protection and these should be respected at all times.
3. Fairness: People should be given equal rights and opportunities and should not be advantaged or disadvantaged undeservedly.
4. Individual, Social and Environmental Well-being: AI systems should contribute to, and

not harm, individual, social and environmental wellbeing.

5. Transparency: The purpose, inputs and operations of AI programs should be knowable and understandable to its stakeholders.

6. Accountability and Oversight: Humans should be able to understand, supervise and control the design and operation of AI based systems, and the actors involved in their development or operation should take responsibility for the way that these applications function and for the resulting consequences.

In addition to the general level ethical principles for AI and robotics, other current research and projects have also produced ethical principles specific for HRI and Industry 5.0 approaches to robotics. For example, based on a Delphi survey of over 30 ethics experts in the field, Callari et al.[35] developed a HRI in Industry 5.0 specific ethical framework for HRC in manufacturing (see Figure 5), principles and their associated design goals (see Table 4). Although similar to the AI HLEG and the Ethics By Design approach, the HRI and Industry 5.0 specific framework takes into account in more detail the role of workers as the collaborators of robots through its focus on worker autonomy, skilling, and overall human authority over the robots within the labor process. Although geared towards the manufacturing sector, it illustrates the additional worker related question to consider in collaborative robotics in industrial settings.

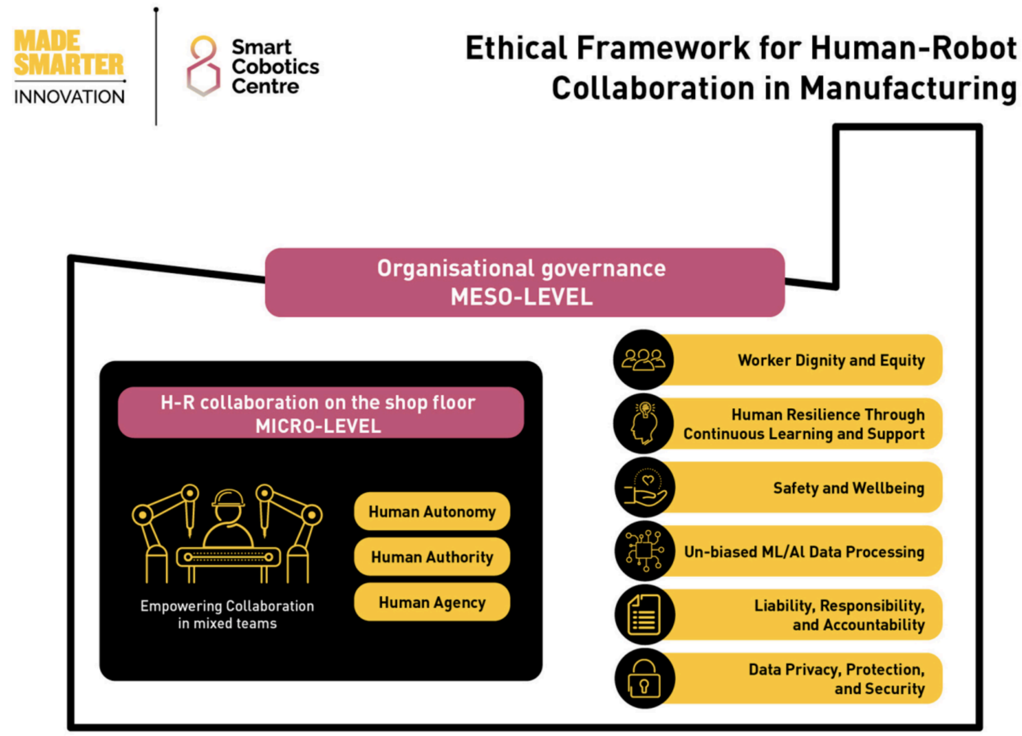

Figure 5. Consolidated ethical framework for HRC in manufacturing (Callari et al., 2024).

[35] Callari, T. C., Segate, R. V., Hubbard, E. M., Daly, A., & Lohse, N. (2024). An ethical framework for human-robot collaboration for the future people-centric manufacturing: A collaborative endeavour with European subject-matter experts in ethics. *Technology in Society*, *78*, 102680. https://doi.org/10.1016/j.techsoc.2024.102680

Table 4. Ethical principles and associated design goals for HRC in manufacturing (Callari et al., 2024)

Ethical principles and associated design goals for HRC in manufacturing
1. Human autonomy: Collaborative tasks allocated to H-R agents should allow (a degree of) autonomy for human workers for potential adaptation to the unpredictability of manufacturing process-related factors, such as temporal task demand variations, or for a change in the state of the operator (e.g. fatigue).Human worker autonomy should not be threatened by any monitoring of collaborative tasks performed with collaborative robots. **2. Human (decision) authority:** Human workers should be empowered to influence changes and improvements in the work environment, and H-R collaboration, fostering meaningful work. Human workers should be given a degree of control over their work processes, allowing them to make decisions, fostering a sense of ownership. **3. Human agency:** Human workers should retain the ability to exert influence over the tasks allocated to collaborative robots, fostering a sense of agency (and trust) in their interactions with collaborative robots. **4. Liability, responsibility, and accountability:** Adequate responsibilities, roles, and procedures should be in place for addressing any potential liability, responsibility, and accountability involved in H-R interaction and collaboration. Guidelines and mechanisms should be established to ensure accountability in cases where H-R interaction and collaboration lead to errors, damages, or accidents. **5. Data privacy, protection, and security requirements in H-R collaboration:** Collect and retain only the necessary data for the intended purpose of improving safety and productivity, minimising the risk of privacy infringement. Informed and voluntary consent must be obtained for data collection and processing related to work activities in H-R interaction and collaboration. **6. Un-biased ML/AI data processing:** Ensure transparency in data collection and processing methods of ML/AI data, and continuously monitor and audit ML/AI systems for algorithm biases, making adjustments as necessary to ensure fairness. Limit ML/AI data processing to the specific purposes disclosed to workers, ensuring that the collection and processing of ML/AI data align with relevant privacy laws and ethical standards. **7. Safety and wellbeing:** Provide safety training to human workers collaborating with collaborative robots, focusing on understanding, and mitigating the risks associated with H-R interaction and collaboration. Incorporate safety measures into the design and development of collaborative robots to ensure the physical safety and wellbeing of human workers, thereby minimising the risk of accidents on the shop floor. Avoid intrusive or constant surveillance practices, as they can contribute to developing a surveillance culture and have adverse effects on the psychological wellbeing of human workers. Encourage a (just) culture of reporting safety incidents, even if they do not result in harm, to improve safety practices continuously. Establish a confidential feedback system that allows human workers to report psychological concerns, ensuring anonymity and protection against retaliation. Ensure that all safety practices and equipment meet or exceed relevant regulatory requirements and industry standards. **8. Human resilience through continuous learning and support:** Support worker personal

development to empower them to navigate their roles safely and effectively, fostering a culture of continuous learning and skill enhancement. Support leadership practices that prioritise the wellbeing and psychological safety in H-R interaction and collaboration. **9. Worker dignity and equity:** Implement policies and procedures to respect, retain and enhance human worker expertise and value. Ensure that human workers receive fair and competitive compensation for their roles in H-R interaction and collaboration on the shop floor, taking into account their skills, contributions, and the changing nature of work.

In this first version of the ARISE SSH framework, we propose to simultaneously test out both the more general level principles set out in the Ethics By Design framework and the more HRI and Industry 5.0 specific ethics principles set out in the framework consolidated by Callari et al. (2024). This is because we aim to have a strong and empirically grounded foundation in Ethics By Design principles, supplemented by more worker related and human-centric approaches. As there is a major overlap between the two approaches, we propose to consolidate their principles and associated requirements or design goals. Table 5 presents the first version of the eight consolidated ARISE HRI ethical principles, namely respect for human agency; privacy and data governance; fairness; individual, social and environmental well-being; transparency; accountability and oversight; worker dignity and equity; human resilience through continuous learning and support. These consolidated principles will act as a first iteration of the ARISE HRI ethical principles to be used with the FSTP projects. The requirements under each principle will be used as a basis for discussion in the ethical impact assessment and mentoring work with the FSTP projects. Our aim is to test what works, identify questions and requirements that seem pertinent, and iterate the ARISE HRI ethical principles and their associated requirements based on the feedback.

Table 5. The first version of the ARISE HRI ethical principles and associated requirements.

The first version of the ARISE HRI ethical principles and associated requirements
1. Respect for Human Agency: Human beings must be respected to make their own decisions and carry out their own actions. Respect for human agency encapsulates three more specific principles, which define fundamental human rights: autonomy, dignity and freedom. **2. Privacy and Data governance:** People have the right to privacy and data protection and these should be respected at all times. **3. Fairness:** People should be given equal rights and opportunities and should not be advantaged or disadvantaged undeservedly. **4. Individual, Social and Environmental Well-being:** Systems should contribute to, and not harm, individual, social and environmental wellbeing. **5. Transparency:** The purpose, inputs and operations of AI and robotics systems should be knowable and understandable to its stakeholders.

6. Accountability and Oversight: Humans should be able to understand, supervise and control the design and operation of AI and robotics systems and the actors involved in their development or operation should take responsibility for the way that these applications function and for the resulting consequences.

7. Worker Dignity and Equity: In line with the core principles of Industry 5.0, dignity and equity are at the core of human-centric, inclusive workplaces. Human worker experience should be respected and the role of human workers in HRC acknowledged and compensated.

8. Human Resilience Through Continuous Learning and Support: Human skills and their development should complement, not compete with advances in robot development. Adequate possibilities for learning new skills and supporting worker psychological safety in HRC should be fostered.

3.2 ARISE ethical impact assessment

Systematic evaluation of the potential impacts of proposed technologies and projects on ethical values or norms is to identify, predict and mitigate the unwanted ethical implications of the initiative. Within the ARISE SSH framework, the initial EIA is grounded on the SIENNA framework and the SATORI EIA process. It will be developed further based on initial findings from the ARISE project's FSTP project phase. The initial ARISE EIA process is first tested with the first batch of the FSTP project in the second year of the project and iterated for the second round based on those experiences. The process will also be updated after the second batch in later stages of the project to capture insights and learnings on EIA, making it readily available for use in future projects. Next, we will describe the various groups related to the ARISE EIA and the six-stage process through which the assessment is conducted.

3.2.1 Different roles within the ethical impact assessment process

The EIA within the ARISE project brings together various groups, namely the FSPT projects, the external ARISE Ethics Committee, ARISE Ethics mentors, and other stakeholders and users that can be engaged in the process. The roles of these groups are emphasized at different stages of the ARISE EIA process.

FSTP projects

The FSTP project who will receive funding from the ARISE project are the main actors in the assessment. The projects will be evaluated by the external ARISE Ethics Committee and helped by the ARISE Ethics mentors. Each FSTP project will be mandated to choose one key accountable person for the ethics mentoring stream as a contact point. In addition, both project management and technical experts are encouraged to attend the ethics related sessions in order to fully capture the different aspects of the projects.

External ARISE Ethics Committee
The external ARISE Ethics Committee (described in greater detail in D3.4 ARISE's Ethics Summary Reports section v1) is an external group of three ethics experts who are in charge of external ethics reviews of the FSTP projects. Based on written material from the FSTP projects, they conduct an initial, interim and final review of ethics-related aspects within the projects and can place recommendations on related actions.

ARISE Ethics mentors
The ARISE Ethics mentors are SSH experts from the ARISE partner Demos Helsinki who will be in charge of facilitating and mentoring of the ethics and SSH related aspects within the FSTP projects, with the help of the ARISE SSH framework for human-centric and ethical HRI. Based on discussions and exercises with the FSTP projects, the ARISE Ethics mentors are in charge of creating an individual ethics plan for the FSTP projects to scope out and meet possible ethics-related aspects coming from the external reviews and further discussions.

Stakeholders
For some projects, it might be pertinent to identify and engage with possible end users and stakeholders related to the project. The ARISE Ethics mentors will help project identify whether or not such stakeholder engagement and participation would be beneficial, who these stakeholders could be and how they should be engaged with.

3.2.2 ARISE ethical impact assessment process

The ARISE EIA process is divided into 6 distinct stages, which are 1) Initial external review, 2) Ethics action plan, 3) Ethics mentoring, 4) Interim external review, 5) Ethics mentoring and roadmap for future use, and 6) Final external review. The EIA process including the six-stages is presented in Figure 6, followed by more detailed descriptions of each stage.

Figure 6. ARISE EIA process for FSTP projects.

1. **Initial external review**

The process begins by the external ARISE Ethics Committee conducting an initial external review of the chosen projects based on the applications and written material from the FSTP projects. The Ethics Committee evaluates each project selected for funding for compliance with the Horizon Europe standards on ethics and according to the Ethics Appraisal rules set up by the European Commission in the standard Ethics assessments conducted in all Horizon Europe calls and programmes. The members of the ARISE Ethics Committee will have autonomy to point out possible concerns and actions for addressing these concerns as well as necessary ethical requirements to be met by the projects, if any.

2. **Ethics action plan**

After the initial external review and its results, the ARISE Ethics mentors will plan individual sessions with each one of the chosen FSTP projects in order to:

- Assign a responsible person as a contact point within the project for the ethics related work and mentoring.
- Conduct a further ethical assessment threshold analysis based on the ARISE SSH framework, including the SATORI EIA process. The process will provide an assessment of possible further ethical concerns to take into account and is based on an easily readable numerical format and will highlight possible concerns in a "traffic light" (green, yellow, red) system based on the answers.
- Go over the results of the threshold assessment and the external ethics review in a discussion with the team and talk over possible points of focus, timelines and actions.
- Co-create an individual ethics mentoring plan to tackle the concerns raised and meet requirements set forth by the external committee and in discussions with the ethics mentors. The ethics action plan will contain key milestones, timelines and actions for the duration of the FSTP phase.

3. **Ethics mentoring**

The ethics mentoring for the FSTP projects will be organized by the ARISE Ethics mentors, that is, SSH experts in the ARISE project according to the individual mentoring plans. The mentoring includes the identification of possible tools and actions (collected together in the ARISE SSH framework), helping understand the principles and how they relate to the project, identifying related legislation and standards as well as identifying the need for possible co-development tools such as stakeholder mapping or design tools. The amount of individual mentoring sessions will depend on the needed requirements and concerns within the project. In addition to individual sessions, the ARISE ethics mentors can also organize common sessions for a few or all FSTP projects if they deem it beneficial for peer-learning. The ethics mentoring will help the FSTP projects prepare for the interim review.

4. **Interim external review**

The interim review is conducted by the external ARISE Ethics Committee. In the review based on written materials provided by the FSTP projects, the external reviewers will focus on whether or not there has been sufficient progress.

5. **Ethics mentoring and roadmap for future use**

After the interim review, the mentoring plan will be updated if necessary and mentoring will continue in an individual format on a needs basis to meet the milestones and requirements set forth in the individual mentoring plans. Further, the ARISE Ethics mentors will facilitate the projects in conducting an ethics roadmap for future use of the technologies and use cases. This roadmap, based on the insight that often lower TRL level technologies will be further developed in the future, is a sort of set of possible questions and concerns that have been identified already at the pilot phase but that will be able to be tackled or addressed only at further stages of real life integration and testing, or at higher TRL levels.

6. **Final external review**

The final review by the external ARISE Ethics Committee will assess how and whether or not the set requirements and milestones were achieved by the projects. The reports will be shared with the FSTP projects, and the final part of funding is tied to the successful review.

3.3 Collection of related European regulation and industry standards

As part of our initial ARISE SSH framework, we provide a collection of European regulation and industry standards related especially to SSH perspectives of the ARISE project. While the current listing does not include all possible perspectives, it aims to be a comprehensive overview of the current approaches relevant to our framework and its use in the project.

Table 6 presents a categorization of multiple different industry standards and EU-level legislation related to HRI. The chosen standards and legislation are EU-centric but also include international standards in cases where they are more applicable for this framework, for example, ISO-standards. The reasoning used for these categories is roughly 1) to make a division between binding and non-binding regulations and/or standards, 2) to make a division between AI and robotics regulations and/or standards, and 3) to specify general safety and human-centric standards. **Binding EU-level legislation** is legally binding within the EU and sanctionable if not followed. **Ethical guidelines and non-binding recommendations** are a set of recommendations, which are highly encouraged to follow but there are no legal ramifications if they are not followed. **AI-specific standards and frameworks** include non-binding standards and recommendations, though most of the standards listed are ISO-standards and countries often refer to these standards in their own legal regulations. **Robotics specific standards** refer to non-binding standards that are focused on robotics. **General safety and human-centric standards** combine non-binding documents that focus on work safety and social aspects of working. Overlap between categories is inevitable, but the different standards and legislation have been themed to their respective categories by the main focus of the document's theme. In addition to the existing legislation, we monitor the field for upcoming and relevant legislation regarding HRI and will be updating the forthcoming iterations of the framework to include these.

Table 6. Industry Standards and Legislation related to HRI.

Binding EU-level legislation
GDPR General Data Protection Regulation
ePrivacy Directive
Machinery Directive
EU AI Act
Proposal tabled by European Commission: AI Liability Directive

Ethical Guidelines and Non-Binding Recommendations
Ethics Guidelines for Trustworthy AI by HLEG
ALTAI Assessment List for Trustworthy Artificial Intelligence (ALTAI) for self-assessment
Ethics By Design and Ethics of Use Approaches
AI-powered robotic strategy by the European Commission (EC)
General Purpose AII Code of Practice (first draft published in November 2024)
IEEE 7000-2021 IEEE Standard Model Process for Addressing Ethical Concerns during System Design
IEEE 7014-2024 IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems
ISO 26000:2010 Guidance on social responsibility
AI-Specific Standards and Frameworks
ISO/IEC 22989:2022 Information technology — Artificial intelligence — Artificial intelligence concepts and terminology
ISO/IEC 23053:2022 Framework for Artificial Intelligence (AI) Systems Using Machine Learning (ML)
ISO/IEC 24029-1:2021 Artificial Intelligence (AI) — Assessment of the robustness of neural networks Part 1: Overview
ISO/IEC TR 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
ISO/IEC 24027:2021 Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making
ISO/IEC TR 24368:2021 Information technology — Artificial intelligence — Overview of ethical and societal concerns
ISO/IEC AWI TR 5469 Artificial intelligence — Functional safety and AI systems
ISO/IEC AWI TS 6254 Information technology — Artificial intelligence — Objectives and approaches for explainability and interpretability of ML models and AI systems
ISO/IEC 24028:2020 Information technology — Artificial intelligence — Overview of trustworthiness in artificial intelligence
IEEE 7001-2021 IEEE Standard for Transparency of Autonomous Systems
ISO/IEC JTC 1/SC 42 (AI Standards) Committee working for standardization in the area of Artificial Intelligence

Robotics Specific Standards
ISO 10218 - 1 & 2 Robots and robotic devices — Safety requirements for industrial robots Part 1: Robots & Part 2: Robot systems and integration
ISO 13482:2014 Robots and robotic devices — Safety requirements for personal care robots
ISO/TS 15066:2016 Robots and robotic devices — Collaborative robots
ISO 9283:1998 Manipulating industrial robots — Performance criteria and related test methods
ISO 9787:2013 Robots and robotic devices — Coordinate systems and motion nomenclatures
ISO/TR 20218-1:2018 Robotics — Safety design for industrial robot systems Part 1: End-effectors
ISO 13399-1:2006 Cutting tool data representation and exchange Part 1: Overview, fundamental principles and general information model
ISO 9409-1:2004 Manipulating industrial robots — Mechanical interfaces Part 1: Plates
ISO/TR 20218-1:2018 Robotics — Safety design for industrial robot systems Part 1: End-effectors
ISO 23795-1:2022 Intelligent transport systems — Extracting trip data using nomadic and mobile devices for estimating C02 emissions Part 1: Fuel consumption determination for fleet management
ISO 15746-1:2015 Automation systems and integration — Integration of advanced process control and optimization capabilities for manufacturing systems Part 1: Framework and functional model
ISO 11354-1:2011 Advanced automation technologies and their applications — Requirements for establishing manufacturing enterprise process interoperability Part 1: Framework for enterprise interoperability
ISO/IEC 20546:2019 Information technology — Big data — Overview and vocabulary
ISO/IEC TR 20547-5:2018 Information technology — Big data reference architecture Part 5: Standards roadmap
ISO 18646 1:2016 and ISO 18646-2:2024 Robotics — Performance criteria and related test methods for service robots Part 1: Locomotion for wheeled robots & Part: Navigation
ISO 18646-3:2021 Robotics — Performance criteria and related test methods for service robots Part 3: Manipulation
ISO 18646-4:2021 Robotics — Performance criteria and related test methods for service robots Part 4: Lower-back support robots

ISO 13849-1:2023 and ISO 13849-2:2012 Safety of machinery — Safety-related parts of control systems Part 1: General principles for design & Part 2: Validation
General Safety and Human-Centric Standards
ISO 12100:2010 Safety of machinery — General principles for design — Risk assessment and risk reduction
ISO 31000:2018 Risk management — Guidelines
EU-OSHA Guidelines European Agency for Safety and Health at Work Guidelines
ISO 27501:2019 The human-centred organization — Guidance for managers
ISO 9241-210:2019 Ergonomics of human-system interaction Part 210: Human-centred design for interactive systems
IEC 61508-1:2010 Functional safety of electrical/electronic/programmable electronic safety-related systems - Part 1: General requirements (see Functional Safety and IEC 61508)
ISO 13849-1:2023 and ISO 13849-2:2012 Safety of machinery — Safety-related parts of control systems Part 1: General principles for design & Part 2: Validation
ISO 9241-210:2019 Ergonomics of human-system interaction Part 210: Human-centred design for interactive systems
IEEE 7010-2020 IEEE Recommended Practice for Assessing the Impact of Autonomous and Intelligent Systems on Human Well-Being

3.4 Collaboration and co-creation toolkit

The ARISE SSH framework ensures that ARISE developments align with a human-centric approach and the principles of Industry 5.0, both of which place special emphasis on human interactions, collaboration and co-creation between the various actors during different phases of the design process. Several methods and tools have been developed for the specific use of HRI or that can be adapted to this context. Table 7 provides an overview of design approaches and co-creation and collaborative tools and methods that are applicable to the context of ARISE. While the list is not exhaustive, it provides ideas for possible approaches that can be tailored according to the specific needs of the project.

Table 7. Design approaches, co-creation and collaborative tools and methods applicable to HRI.

Context	Name of tool or approach	What is it used for?

General design approach	**Design Thinking**[36]	A human-centered approach to innovation that includes empathizing, defining, ideating, prototyping, and testing.
General design approach	**Participatory Design**[37]	Involves users directly in the design process to ensure outcomes meet their needs and preferences.
General design approach	**User-centered (UX) design**[38]	Focuses on designing products based on the explicit needs and requirements of users.
General design methodology	**Double diamond**[39]	Methods aiming to create a visual representation of any existing or new innovation and design processes. It helps recognize key interventions where human-centric approaches are needed.
Human-centered design methodology	**The Human-Centered Design Toolkit**[40]	Toolkit containing several tools and methods to bring up human-centricity at every stage of any design process
Ethics; General design process	**Ethics for Designers toolkit**[41]	Toolkit to explore, discuss and uncover the ethical aspect at different stages of design processes
Human centered design; AI	**Design Toolkit for Human-Centered AI (HCAI)**[42]	Collection of tools and methods to bring up human centricity at different stages of design AI applications.
Co-creation; Design process of social robots	**LEADOR *(Led-by-Experts Automation and Design Of Robots)***[43],[44]	An end-to-end participatory design methodology involving domain experts in the co-design, automation, and evaluation of social robots.
Multidisciplinary collaboration; Design process of social robots	**Social Robot Co-Design Canvases**[45]	Framework and canvas tool to facilitate collaboration between experts of different fields in the design process of social robots, while also

[36] Design Thinking. Available at: https://www.interaction-design.org/literature/article/5-stages-in-the-design-thinking-process
[37] Participatory Design. Available at: https://www.interaction-design.org/literature/topics/participatory-design
[38] UX design. Available at: https://www.interaction-design.org/literature/topics/user-centered-design
[39] Double diamond. Available at: https://www.designcouncil.org.uk/our-resources/the-double-diamond/
[40] The Human-Centered Design Toolkit. Available at: https://www.ideo.com/journal/design-kit-the-human-centered-design-toolkit
[41] Ethics for Designers toolkit. Available at: https://www.ethicsfordesigners.com/tools
[42] Design Toolkit for HCAI. Available at: https://projects.tuni.fi/uploads/2023/02/d7f42d29-kite-hcai-design-toolkit.pdf
[43] Leodor. Available at: https://arxiv.org/abs/2105.01910
[44] Leodor. Available at: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2021.704119/full
[45] Social Robot Co-Design Canvases. Available at: https://dl.acm.org/doi/10.1145/3472225

		incorporating a user perspective
Design process of human–robot collaborative workplaces	**Smart Graph Interface (SGI)**[46]	Structured framework and interactive tool to support strategic decisions in designing human–robot collaborative workplaces.
Collaborative design; Care robots supporting disabled people	**LEGO® Serious Play® and Design Thinking Workshops**[47]	These methodologies engage participants in envisioning useful robots.
Co-creation; Care robots	**Co-Creation Sessions**[48]	This approach involves organizing multiple co-creation sessions with diverse individuals to refine the design of assistive robots.
Participatory design; Care robots supporting older adults	**Situated Participatory Design (SPD)**[49]	Participatory methodology for designing human-robot interactions directly within the environments where they will be deployed.
Participatory design; explainable robots	**Participatory Design for Explainable Robots**[50]	Participatory methodology that emphasizes the creation of robots that can articulate their decision-making processes to users.
Design process in HRI	**HRI Toolkits**[51]	Comprehensive toolkits that provide a suite of tools and methods to design, implement, and evaluate HRC systems.
Design process of Industrial collaborative robots	**SHAREWORK Toolkit**[52]	SHAREWORK offers a modular software/hardware toolbox for industrial collaborative robots, enabling human-like perception, human-aware dynamic planning, and safe maneuvering without safety fences.
Design process of Industrial collaborative robots	**CoCo (Coexistence-and-Cooperation) System**[53]	The CoCo system facilitates seamless human-robot collaboration by dynamically switching between

[46] SGI. Available at: https://link.springer.com/article/10.1007/s12008-023-01607-y
[47] LEGO® Serious Play® and Design Thinking Workshops. Available at: https://www.frontiersin.org/journals/robotics-and-ai/articles/10.3389/frobt.2022.731006/full
[48] Co-Creation Sessions. Available at: https://link.springer.com/article/10.1007/s12008-019-00641-z
[49] SDP. Available at: https://arxiv.org/abs/2302.00588
[50] Participatory Design for Explainable Robots. Available at: https://www.inf.uni-hamburg.de/research/projects/trail/publications/pdfs/2024-03-gebelli-hri24.pdf
[51] HRI toolkits: https://www.researchgate.net/publication/364974022_Tools_and_Methods_for_Human_Robot_Collaboration_Case_Studies_at_i-LABS
[52] SHAREWORK toolkit: https://cordis.europa.eu/project/id/820807/reporting
[53] CoCo system. Available at: https://arxiv.org/abs/2206.01775

		coexistence and cooperation modes based on human intentions.

3.5 Typically encountered ethical questions related to AI systems

Typically encountered ethical questions in HRI development are manyfold. We have organized them according to ethical principles presented in the ethics-by-design framework, which builds on the Ethics Guidelines for Trustworthy AI and SIENNA and SHERPA projects. The Ethics Guidelines for Trustworthy AI is the source for the Assessment List for Trustworthy AI (ALTAI), which provides examples of general level ethical questions related to AI systems presented below. Additionally, we complete the whole by incorporating two ethical principles we propose—1) Worker dignity and equity, 2) Human resilience through continuous learning and support—thus the list becomes complete, covering all 8 consolidated ARISE HRI ethical principles. For the last two principles, we have created example questions ourselves, based on the original work by Callari et al. (2024).

1. Respect for human agency
Human beings must be respected to make their own decisions and carry out their own actions. Respect for human agency encapsulates three more specific principles—autonomy, dignity and freedom—which define fundamental human rights.

- Is the AI system designed to interact, guide or take decisions by human end-users that affect humans or society?
- Could the AI system generate confusion for some or all end-users or subjects on whether they are interacting with a human or AI system?
- Could the AI system affect human autonomy by generating over-reliance by end-users?
- Could the AI system affect human autonomy by interfering with the end-user's decision-making process in any other unintended and undesirable way?
- Does the AI system simulate social interaction with or between end-users or subjects?
- Does the AI system risk creating human attachment, stimulating addictive behaviour, or manipulating user behaviour?

2. Privacy and Data governance
People have the right to privacy and data protection, and these should be respected at all times.

- Did you consider the impact of the AI system on the right to privacy, the right to physical, mental and/or moral integrity and the right to data protection?
- Depending on the use case, did you establish mechanisms that allow flagging issues related to privacy concerning the AI system?
- Is your AI system being trained, or was it developed, by using or processing personal data (including special categories of personal data)?
- Did you put in place any of the following measures some of which are mandatory under the General Data Protection Regulation (GDPR), or a non-European equivalent?
    - Data Protection Impact Assessment (DPIA)

 - Designate a Data Protection Officer (DPO) and include them at an early state in the development, procurement or use phase of the AI system;
 - Oversight mechanisms for data processing (including limiting access to qualified personnel, mechanisms for logging data access and making modifications);
 - Measures to achieve privacy-by-design and default (e.g. encryption, pseudonymisation, aggregation, anonymisation);
 - Data minimisation, in particular personal data (including special categories of data);
- Did you implement the right to withdraw consent, the right to object and the right to be forgotten into the development of the AI system?
- Did you consider the privacy and data protection implications of data collected, generated or processed over the course of the AI system's life cycle?
- Did you consider the privacy and data protection implications of the AI system's non-personal training-data or other processed non-personal data?
- Did you align the AI system with relevant standards (e.g. ISO25, IEEE26) or widely adopted protocols for (daily) data management and governance?

3. Fairness

People should be given equal rights and opportunities and should not be advantaged or disadvantaged undeservedly.

- Did you establish a strategy or a set of procedures to avoid creating or reinforcing unfair bias in the AI system, both regarding the use of input data as well as for the algorithm design?
- Did you consider diversity and representativeness of end-users and/or subjects in the data?
- Did you put in place educational and awareness initiatives to help AI designers and AI developers be more aware of the possible bias they can inject in designing and developing the AI system?
- Did you ensure a mechanism that allows for the flagging of issues related to bias, discrimination or poor performance of the AI system?
- Is your definition of fairness commonly used and implemented in any phase of the process of setting up the AI system?
- Did you ensure that the AI system corresponds to the variety of preferences and abilities in society?
- Did you assess whether the AI system's user interface is usable by those with special needs or disabilities or those at risk of exclusion?
- Did you ensure that Universal Design principles are taken into account during every step of the planning and development process, if applicable?
- Did you take the impact of the AI system on the potential end-users and/or subjects into account?
- Did you consider a mechanism to include the participation of the widest range of possible stakeholders in the AI system’s design and development?

4. Individual, Social and Environmental Well-being

AI systems should contribute to, and not harm, individual, social and environmental wellbeing.

- Does the AI system impact human work and work arrangements?
- Did you pave the way for the introduction of the AI system in your organisation by informing and consulting with impacted workers and their representatives (trade unions, (European) work councils) in advance?
- Did you adopt measures to ensure that the impacts of the AI system on human work are well understood?
- Could the AI system create the risk of de-skilling the workforce?
- Does the system promote or require new (digital) skills?
- Could the AI system have a negative impact on society at large or democracy?
- Are there potential negative impacts of the AI system on the environment?
- Where possible, did you establish mechanisms to evaluate the environmental impact of the AI system's development, deployment and/or use (for example, the amount of energy used and carbon emissions)?

5. Transparency

The purpose, inputs and operations of AI programs should be knowable and understandable to its stakeholders.

- Did you put in place measures that address the traceability of the AI system during its entire lifecycle?
- Did you explain the decision(s) of the AI system to the users?
- Do you continuously survey the users if they understand the decision(s) of the AI system?
- In cases of interactive AI systems (e.g., chatbots, robo-lawyers), do you communicate to users that they are interacting with an AI system instead of a human?
- Did you establish mechanisms to inform users about the purpose, criteria and limitations of the decision(s) generated by the AI system?

6. Accountability and Oversight

Humans should be able to understand, supervise and control the design and operation of AI based systems, and the actors involved in their development or operation should take responsibility for the way that these applications function and for the resulting consequences.

- Did you establish mechanisms that facilitate the AI system's auditability (e.g. traceability of the development process, the sourcing of training data and the logging of the AI system's processes, outcomes, positive and negative impact)?
- Did you ensure that the AI system can be audited by independent third parties?
- Did you foresee any kind of external guidance or third-party auditing processes to oversee ethical concerns and accountability measures?
- Did you organise risk training and, if so, does this also inform about the potential legal framework applicable to the AI system?
- Did you consider establishing an AI ethics review board or a similar mechanism to discuss the overall accountability and ethics practices, including potential unclear grey areas?
- Did you establish a process to discuss and continuously monitor and assess the AI system's adherence to this Assessment List for Trustworthy AI (ALTAI)?

- Did you establish a process for third parties (e.g. suppliers, end-users, subjects, distributors/vendors or workers) to report potential vulnerabilities, risks or biases in the AI system?
- For applications that can adversely affect individuals, have redress by design mechanisms been put in place?
- Have the humans (human-in-the-loop, human-on-the-loop, human-in-command) been given specific training on how to exercise oversight?
- Did you establish any detection and response mechanisms for undesirable adverse effects of the AI system for the end-user or subject?
- Did you ensure a 'stop button' or procedure to safely abort an operation when needed?
- Did you take any specific oversight and control measures to reflect the self-learning or autonomous nature of the AI system?

7. Worker Dignity and Equity

In line with the core principles of Industry 5.0, dignity and equity are at the core of human-centric, inclusive workplaces. Human worker experience should be respected and the role of human workers in HRC acknowledged and compensated.

- How is it ensured that human workers are treated with dignity and respect in a workplace shared with robotic/AI systems?
- Are there measures in place to address any form of discrimination or harassment in the workplace, particularly in scenarios involving human-robotic/AI system interactions?
- How does the organization's compensation structure reflect the value of human workers contribution, especially in comparison to the efficiencies gained through automation?
- In what ways the unique contributions of human workers in technically integrated workplaces is recognized?
- How are diverse workers' needs in settings that include both human and robotic/AI workers accommodated?
- How is the significance of human roles communicated within the organization and in external communication?
- What mechanisms are in place for workers to provide feedback or raise concerns about their working conditions, particularly regarding their interactions with robotic/AI systems?
- How is training for new technologies, especially those involving collaboration with robotic/AI systems, made accessible and equitable for all employees?

8. Human Resilience Through Continuous Learning and Support

Human skills and their development should complement not compete with advances in robot development. Adequate possibilities for learning new skills and supporting worker psychological safety in HRC should be fostered.

- How are learning opportunities designed to complement and enhance human skills alongside advancements in robotics/AI systems?
- What strategies are in place to ensure that the skills development offered keeps pace with the evolution of technology in the workplace?
- What types of learning resources are available to workers to help them adapt to and collaborate effectively with robotic/AI systems?

- What measures are taken to support the psychological safety of workers interacting with robotic/AI systems?
- How are potential anxieties or concerns associated with increased automation and human-machine interactions addressed?
- How is feedback from employees about the effectiveness of the learning and support programs provided gathered and used?
- How are learning and development opportunities integrated into worker's regular work schedules without causing disruption or undue stress?
- What career pathways are available for workers in a hybrid environment of humans and robotic/AI systems?
- How are career pathways communicated to employees, and how do they reflect the potential for long-term career growth despite technological changes?
- What long-term support systems are in place for employees to continue developing their skills over time?
- How is the culture of continuous learning and resilience among its workforce promoted?
- What incentives and rewards are provided to encourage ongoing personal or professional development?

3.6 Collection of relevant use cases and examples of ethical concerns and questions

The ARISE project is structured around four TEFs which all provide two relevant HRI-related challenges, each accompanied by corresponding illustrative use cases. The eight proposed challenges provided by the ARISE TEF (CARTIF, INTELLIMECH, PAL and POLIMI) partners are: 1) dismantling and assembly of high-value products, 2) complex product picking in industrial warehouses, 3) flexible collaborative robots, 4) smart programming, 5) enhancing robot functionality through multimodal HRI interactions, 6) developing robotic systems able to autonomously perform fetch and carry tasks in healthcare environments, 7) leveraging HRI to improve the efficiency of workers in high precision flexible tasks, and 8) leveraging HRI for improving ergonomics in high precision tasks. These 8 use cases serve as examples and inspiration for FSTPs, which will design and execute their own experiments within the open call programs and the scope of the ARISE project. The use cases, although illustrative in nature, provide a good starting point to anticipate potential ethical concerns associated with the proposed HRI-related challenges. In this section, all eight use cases are introduced, followed by examples of ethical concerns and associated questions that should be accounted for—from the perspective of the ARISE HRI ethical principles.

Illustrative use case for challenge 1 dismantling and assembly of high-value products: Battery Disassembly

The primary objective of this application case is to optimize and expedite the disassembling process of batteries by utilizing a collaborative approach between humans and robots. This strategy integrates mixed reality and artificial vision technologies to accurately detect and localize screws within the battery, enhancing efficiency and accelerating the disassembly procedure. The actors of the use case are: Human workers, PC (System), Robot/Camera, and MR HoloLens glasses.

The general flow of the work task is as follows: the dimensional camera captures BGR and depth image of the object and forwards them to the PC for analysis. The PC then proceeds with coordinate calculations and transformations to both robot reference space and world coordinates. These coordinates are then sent forward to the robot, which guides the human. The PC's role is to oversee the interaction between the human and the robot. Meanwhile, the collaborative robot selects screws for the disassembly process. The actors of this use case, and their roles, are summarized in Table 8.

Table 8. The actors and their roles in the battery disassembly use case.

Human workers	PC (System)	Robot/Camera	MR HoloLens glasses
Supervising the disassembly process.	Calculating and transforming coordinates.	Providing accurate data.	Provide real time visual information.
Establishing the framework and making critical decisions, such as selecting screws to retrieve.	Coordinating robot actions.	Perform physical actions.	Track user interactions for an immersive user experience.
	Relaying data to AR glasses for enriched visual experiences.		

Examples of ethical concerns and their associated questions

Respect for Human Agency

- How is workers' informed consent ensured before the HRI solution is deployed?

Privacy and Data governance:

- How is it ensured that the camera embedded in the system collects data solely for the purpose of the HRI solution?

Fairness:

- How does the use case address diversity among workers to ensure the Voice Command Recognition capability is inclusive and recognizes non-standard accents?

Individual, Social and Environmental Well-being:

- What potential health and safety risks to workers are assessed and what measures are implemented to address them?

Transparency:

- Has the project completed a dual-use compliance statement?

Accountability and Oversight:

- How are accountability and liability for the HRI system's performance established, especially in the case of system failures?

Worker Dignity and Equity:

- Are there measures in place to ensure that all team members benefit from the time savings and reduced physical strain provided by the collaborative robot?

Human Resilience Through Continuous Learning and Support:

- What ongoing training and support are provided to help workers adapt to using MR glasses and working alongside a robot?

Illustrative use case for challenge 2 complex product picking in industrial warehouses: Fruit picking

The main purpose of this application case is to make certain tasks that would be done by a human worker, like picking fruits, can be done by a collaborative robot, making HRI possible and perform the work more efficiently. The actors of the use case are: Human workers, MR HoloLens glasses, system, Collaborative robot and camera. The users engage with the mixed reality glasses through intuitive gestures, voice commands and gaze. The MR glasses then send forward this input to the user by providing a live camera view, object information and system feedback. The glasses also forward relevant data to the PC via a Wi-Fi connection (WebSockets protocol). The PC then processes the data and sends the processed data back to the user. The PC's role is to act as a central hub and facilitate the communication between the user, the glasses and the robot, sending commands to the robot based on the input information received from the user. The robot provides updates on its actions and its environment to the PC. The camera's role is to capture visual data and send it to the PC for adjustments of camera settings. The actors of this use case, and their roles, are shown in Table 9.

Table 9. The actors and their roles in the fruit picking use case.

Human workers	MR HoloLens glasses	System	Collaborative robot and camera
Supervise operations	Provide interface between human workers and the system	Receive instructions	Robot sorts based on instructions from human workers
Provide instructions to the collaborative robot via MR HoloLens glasses	Provides location of objects as a digital sphere so that the user is aware of the pick-up position of the robot	Provide real-time visual feedback	Camera captures visual data for object recognition and situational awareness
			Aids decision-making in the system

Examples of ethical concerns and their associated questions
Respect for Human Agency
- How does the use case consider the risk of worker displacement due to automation and what measures are implemented to mitigate this risk?

Privacy and Data governance:
- If the HRI solution uses data from a third party, how is API integration compliance ensured, and is a compliance form provided?

Fairness:
- How does the project ensure that visually impaired, far-sighter or near-sighted workers can effectively use MR HoloLens glasses?

Individual, Social and Environmental Well-being:
- What protocols are in place for safe physical interaction, including detailed fail-safe mechanisms?

Transparency:
- Are the participants informed about the purposes, risks, and benefits of their involvement in the research activities?

Accountability and Oversight:
- What processes are in place to ensure compliance with ethical, legal and organizational standards throughout the research activities?

Worker Dignity and Equity:
- How is training for the HRI solution made accessible and equitable for all employees?

Human Resilience Through Continuous Learning and Support:
- How are workers encouraged to provide feedback about the technology and their training needs?

Illustrative use case for challenge 3 flexible collaborative robots: Assembly and packing robotic assistant
In this application case, the robot should recognize and manage the heavy product, safeguarding the operator from injuries and safety problems, while the operator should monitor the robot's activity and intervene, for example, if inaccuracies occur or unexpected forces are requested to perform the task.

A product is selected to be packed from the company management system's database by an operator. After this, the robot receives the product code information and obtains the necessary information to recognise the product in the space (CAD and images). The robot also receives the information on the related tray. Subsequently the operator prepares the packing area by placing the first tray on a pallet. Through physical interface and voice input commands, the operator advises the robot so that it can recognise the object and its pose. Then the robot picks up the object and places it in one of the free slots. Because the pre-formed tray constrains the product's positioning and requires a specific force to achieve the correct coupling, the positioning sequence in the slots is pre-programmed. In case of incorrect coupling in the product's final slot, the operator can intervene by physically acting on the robot for a successful positioning of the product in its slot. In this phase, the robot bears the product's weight and cooperates with the human to achieve precise positioning. Once the tray is filled, the operator will perform a quality check and set up another tray until the desired number of trays for the current pallet is reached. At any time, the operator can stop, start or give high-level commands to the robot. During the whole process, a camera monitors the operator's posture, and they are alerted if their ergonomic indexes fall below a threshold. The actors of this use case, and their roles, are summarized in Table 10.

Table 10. The actors and their roles in the assembly and packing robotic assistant use case.

Collaborative robot and operator	Human workers
Execute force operations	Supervise the packing process
Execute fine manual operations	Handle exceptions
Share some tasks with human operators	Manage the empty pre-formed tray layer or empty boxes
	Manage orders from company management system
	Coordinate shipments

Examples of ethical concerns and their associated questions

Respect for Human Agency

- What strategy is adopted to ensure that workers possess the necessary knowledge and skills to collaborate efficiently with the robot?

Privacy and Data governance:

- Will research data be anonymized or pseudonymised? If so, what techniques will be implemented?

Fairness:

- How are the voice command and camera features designed to consider diverse backgrounds, facial features, languages, and accents of participants?

Individual, Social and Environmental Well-being:

- What sustainability practices are adopted for dismantling, assembly and the environmental impact of the use case?

Transparency:

- Has the project completed a misuse compliance statement?

Accountability and Oversight:

- How will the implementation of health and safety procedures, data collection measures and informed consent processes be monitored and evaluated?

Worker Dignity and Equity:

- How does the system ensure equitable safety measures for all operators, particularly in interventions involving physical interactions with the robot?

Human Resilience Through Continuous Learning and Support:

- What mechanisms are in place for operators to provide feedback on the robotic system and the overall workflow?

Illustrative use case for challenge 4 smart programming: Robotic Kitting Assistant

One of the standard and crucial phases in producing assembled parts is kitting. A specific product is selected from the company's management system, obtaining its Bill of Materials (BoM) with the various components and their location within the warehouse. Following a lean paradigm, all the items constituting the kit should be retrieved from a warehouse, bringing only what is needed to the assembly station. This application case considers both automated warehouses, involving a robotic cell, and non-automated warehouses, consisting of shelving, which humans should manage.

The flow of the process starts when the operator selects a product to be assembled from the company management system, and the robot receives a list of objects, their quantity, and their location within the company. The needed items could be stored in an automated or non-automated warehouse. Then the operator reads the object list from the robot's touchscreen. The robot can follow the operator to the automated warehouse in follow-me mode or independently. Once in the warehouse, the operator takes items from the shelves and places them inside the bin, which will constitute the assembly kit. The operator then updates the list of items to be picked via the screen or with natural language. While this operation is happening, the robot will communicate with the warehouse, requesting the bins with the remaining components. Once the selected bin is retrieved, the robot should pick the right number of items and move them to the assembly kit bin, simultaneously updating the list of items. Once all the objects of the list have been collected, the operator performs a quality check and sends the robot to the assembly station that requested the kit. The actors of this use case, and their roles, are summarized in Table 11.

Table 11. The actors and their roles in the robotic kitting assistant use case.

Autonomous mobile robot and collaborative robot	**Human workers**
Offer a simple way to codify different objects	Teach the robot on what it has to do
Take input naturally from operators	Supervise the kitting process
Autonomously create the kit	Handle exceptions
Notify lack of necessary items	Provide additional instructions if it is necessary
Navigate autonomously inside the warehouse	Task and process scheduling

Examples of ethical concerns and their associated questions

Respect for Human Agency

- What are the detailed procedures for obtaining informed consent, and can templates of the Informed Consent Forms and Information Sheets in language intelligible to participants be provided?

Privacy and Data governance:

- How does the use case demonstrate compliance with relevant data protection regulations, such as GDPR, and what security measures are in place to protect collected data from unauthorized access?

Fairness:

- How is inclusivity ensured in addressing the challenges posed by diverse practitioner characteristics?

Individual, Social and Environmental Well-being:

- How does the use case ensure compliance with relevant local, national or European health and safety guidelines and regulations?

Transparency:

- How are the details of informed consent, data processing and other relevant procedures communicated in an understandable and accessible manner to all participants?

Accountability and Oversight:

- How are accountability and liability for the HRI system's performance established, especially in the case of system failures?

Worker Dignity and Equity:

- How does the HRI solution address potential biases, especially in recognition technologies used in interfaces like touchscreens and natural language processing?

Human Resilience Through Continuous Learning and Support:

- How does the organization promote a culture where continuous improvement and innovation are encouraged, and learning from mistakes is seen as part of growth?

Illustrative use case for challenge 5 enhancing robot functionality through multimodal HRI interactions: Interactive learning session

This use case demonstrates the robot's ability to adapt technology to the user's needs and environment, emphasising a human-centred approach in HRI. It showcases how robots can support personal growth and enrichment activities, making technology an empowering tool rather than a constraint.

The focus of this example use case emphasises the patient-robot flow, which has been deemed more interesting from an interaction perspective. The robot first assesses the user's (patient) skill level and specific interests in using speech recognition and the robot proceeds to personalise the session based on the user's responses and observable ques. After this, the robot provides step-by-step instructions for the session, utilising multimodal communication like gestures and verbal instructions. The robot then monitors the user's progress during the session and adjusts the pace or repeats instructions as needed. In addition, the robot uses facial expression recognition and activity recognition to observe signs of frustration and confusion. If these signs are observed, the robot will offer encouragement, suggest breaks, or modify the teaching method to keep the session enjoyable. During the session, the robot will also provide real-time feedback and tips for improvement, encouraging exploration and creativity while simultaneously adapting the session to incorporate the user's ideas and preferences. After the session, the robot reviews the user's achievements and areas for improvement. Based on the user's engagement and feedback, the robot suggests future sessions or related hobbies. The personalised profile of the user is stored for future user interactions with the robot. The actors of this use case, and their roles, are shown in Table 12.

Table 12. The actors and their roles in the interactive learning session use case.

Care staff	Patients	Interactive robot	Care environment
Scheduling physical activities via a mobile-based interface	Different needs, limitations and preferences	Range of sensing and interaction capabilities	Physical building layout
			Organisational constraints

Examples of ethical concerns and their associated questions

Respect for Human Agency

- How will the informed consent of participants (staff, employee, volunteers, patients, etc.) be obtained to ensure ethical requirements are met?

Privacy and Data governance:

- What technical and organizational measures are implemented for further processing of previously collected personal data to safeguard the rights of data subjects?

Fairness:

- How are the voice command and camera features designed to consider diverse backgrounds, cognitive abilities and technical capabilities?

Individual, Social and Environmental Well-being:

- What potential health and safety risks to workers are assessed and what measures are implemented to address them?

Transparency:

- Has the project completed a misuse compliance statement?

Accountability and Oversight:

- What processes are in place to ensure compliance with ethical, legal and organizational standards throughout the research activities?

Worker Dignity and Equity:

- What mechanisms are in place to continuously monitor and correct biases in algorithms, especially those related to facial expression and activity recognition?

Human Resilience Through Continuous Learning and Support:

- What features does the robot have to create a psychologically safe learning environment where users feel free to take risks and express creativity?

Illustrative use case for challenge 6 developing robotic systems able to autonomously perform fetch and carry tasks in healthcare environments: Asking for help while carrying medical samples

This case focuses on using assistive robots for the safe and efficient transportation of medical samples between patient wards and laboratories within healthcare environments. In case of the robot being blocked (for instance, because of a hospital bed blocking the passage in a corridor), the robot looks for someone and asks for help.

The main flow consists of two interactions that need to be considered, the first one being fetch & carry flow. In this scenario a staff member calls the robot and requests it to bring a medical sample to a known location, for example a laboratory. In this case a nurse places the sample on the robot and then lets the robot go. The second main flow is a help request flow, which follows the first flow. In this case the robot has faced an interruption in its navigation. This could be due to, for example, an object blocking the passage or a door being closed. This results in the robot looking for help to overcome these obstacles in navigation. The actors of this use case, and their roles, are illustrated in Table 13.

Table 13. The actors and their roles in the asking for help while carrying medical samples use case.

Hospital staff members	Mobile manipulator (robot)
Send requests for the robot for samples	Carry samples
Load the samples on to the robot	Navigate obstacles
Assist robot when interruptions in navigation happens	Ask for help when navigation is interrupted

Examples of ethical concerns and their associated questions

Respect for Human Agency

- What are the procedures and criteria for identifying and recruiting research participants (e.g., number of participants, inclusion/exclusion criteria, direct/indirect incentives, risks, and benefits for participants?)

Privacy and Data governance:

- How does the use case ensure that the data processed is relevant and limited to the purposes of the project, in compliance with the data minimization principle?

Fairness:

- How are the voice command and camera features designed to consider diverse backgrounds, facial features, languages, and accents of participants?

Individual, Social and Environmental Well-being:

- What protocols are in place for safe physical interaction, including detailed fail-safe mechanisms?

Transparency:

- Are the participants informed about the purposes, risks, and benefits of their involvement in the research activities?

Accountability and Oversight:

- How will the implementation of health and safety procedures, data collection measures and informed consent processes be monitored and evaluated?

Worker Dignity and Equity:

- Is the robot's interface designed to be easily usable by all staff, regardless of their technical skill levels or physical capabilities?

Human Resilience Through Continuous Learning and Support:

- What channels are available for staff to report issues or suggest improvements regarding the robot's operations?

Illustrative use case for challenge 7 leveraging HRI to improve the efficiency of workers in high precision flexible tasks: PCB Desoldering

Reworking of printed circuit boards (PCBs) which did not pass the quality control check is considered a high-precision task. Defective PCBs should be recovered, i.e., defective components must be identified and replaced by operational ones. The process of replacing components is intricate and time-consuming, in the sense that an operator must (i) recognize the defective components on the board by the information received from the quality control station, (ii) pick the functional component among the pool of components (it may be time-consuming, due to similarity of components, for example, resistors with identical shape and different resistance values), (iii) select the suitable tools among the available tools, depending on the type of components and the level of operator expertise, (iv) solder and desolder components taking into account to avoid applying excessive heat to components and board, to orient components correctly, to inspect new soldered joints, and etc., which causes high levels of physical and mental stress on them.

The main flow starts by the automated guided vehicle carrying the defective PCBs from the control station room to the reworking stations room. During this, the UR5e robot is preparing the rework station by the data received from the control unit. Preparation of the rework station includes identification of the proper tools and functional components from the available pool of tools and components via computer vision. The robot also picks and places all the needed objects on the rework station. The identification of correct tools happens through BGR and depth images that are sent from the robot to the PC, where the PC runs object detection algorithms to identify the correct tools and components. In addition, their positions are calculated. In addition, the automated guided vehicle is provided with a map of the building, so that it can navigate across the shop floor. The actors of this use case, and their roles, are summarized in Table 14.

Table 14. The actors and their roles in the PCB desoldering use case.

Human workers	UR5e Robot	AGV
Soldering/desoldering	Pick the functional component among the pool of components	Carry defective PCBs from the control station room to the reworking station
	Select the suitable tools among the available tools	
	Set up working space	

Examples of ethical concerns and their associated questions

Respect for Human Agency

- How are participants, such as employees, being given opportunities to train their skill in HRI?

Privacy and Data governance:

- What are the detailed procedures for obtaining informed consent regarding data processing?

Fairness:

- How is inclusivity ensured in addressing the challenges posed by diverse practitioner characteristics?

Individual, Social and Environmental Well-being:

- What sustainability practices are adopted for dismantling, assembly and the environmental impact of the use case?

Transparency:

- Has the project completed a dual-use compliance statement?

Accountability and Oversight:

- How are accountability and liability for the HRI system's performance established, especially in the case of system failures?

Worker Dignity and Equity:

- Are there safeguards to prevent certain operators from consistently receiving more challenging or stressful rework tasks based on the system's performance assessment?

Human Resilience Through Continuous Learning and Support:

- Are there mechanisms for operators to report issues, provide suggestions and contribute to continuous improvement efforts without fear of reprisal or negative impacts on their employment?

Illustrative use case for challenge 8 leveraging HRI for improving ergonomics in high precision tasks: Workpiece holder

The robot holds the workpiece in the appropriate position and orientation while an operator is working on it. Through the machine vision, the physical features of the operator including the height, and neck length are provided to the robot for calculating the right position and orientation of the workpiece considering the ergonomics and comfort of the operator. In addition, for the further adjustment of orientation, the operator sends a speech command to the robot.

The main flow of the process initiates by capturing BGR and depth images of the operator and sending these images to the PC. The PC then extracts the relevant physical features of the operator, and the appropriate ergonomics and comfort of the operator are defined. In the next step, the desired joint angles of the robot are obtained and sent to the robot for holding the workpiece. The operator's speech is captured and used for voice commands to the robot to make required adjustments. The actors of this use case, and their roles, are shown in Table 15.

Table 15. The actors and their roles in the workpiece holder use case.

Operator	UR5e robot
Send move commands to the robot verbally	Adjusts the workpiece orientation

	Estimate the pose of the workpiece according to physical parameters of operators through images.

Examples of ethical concerns and their associated questions
Respect for Human Agency
- How does the project consider the risk of worker displacement due to automation and what measures are implemented to mitigate this risk?

Privacy and Data governance:
- Has the DPO been appointed and are the contact details of the DPO made available to all data subjects involved in the research?

Fairness:
- How are the voice commands designed to consider diverse backgrounds, and non-standard accents?

Individual, Social and Environmental Well-being:
- How does the use case ensure compliance with relevant local, national or European health and safety guidelines and regulations?

Transparency:
- How are the details of informed consent, data processing and other relevant procedures communicated in an understandable and accessible manner to all participants?

Accountability and Oversight:
- How will the implementation of health and safety procedures, data collection measures and informed consent processes be monitored and evaluated?

Worker Dignity and Equity:
- How can the system be audited for bias in the recognition and adjustment algorithms to ensure fair treatment of all operators?

Human Resilience Through Continuous Learning and Support:
- What training and support mechanisms are provided to operators to effectively interact with the robot, especially for those less familiar with advanced technology.

4 Discussion

4.1 Evolving nature of the ARISE SSH framework

In this report, we have described the foundations and key elements of the initial version of the ARISE SSH framework for human-centric and ethical HRI. We have given an overview of the undertaken and planned activities and results of this development work, which will continue throughout the project. The three key milestones in the development work are the release of the initial version of the ARISE SSH framework at the end of 2024, the release of first iteration of the framework at the end of 2025, as well as release of the final version of the framework at the end of the project in June 2027. This ongoing process enables us to monitor the broad landscape of the SSH aspects relevant to HRI, iterating our own work by gathering new empirical knowledge in the context of the FSTP projects and including the newest information to the evolving framework.

4.2 Emphasis on contextualized, human-centric innovation

While this initial framework provides a wide overview of the existing guidelines, tools and methods relevant to HRI from the SSH perspective, our aim is not just to provide a collection of those. The central aim of the SSH research in the ARISE project is to turn the higher-level guidelines into practical tools and meaningful recommendations and specifications. This is done throughout our qualitative research, in which we apply participatory, ethnographic and human-centric design methodologies. We study the workflows and processes of HRI and its professionals' values, aspirations and needs in relation to SSH perspectives to understand what the current and possible future practices are where SSH aspects, and ethics considerations particularly, could be embedded. This paves the project's way toward contextualized, human-centric and ethical innovation in the context of HRI. One of the key contributions of the ARISE project from the SSH perspective will indeed be the insights from the contextual understanding gained through bottom-up qualitative research, with a strong emphasis on participatory research and design. This will be a valuable, novel contribution to the discussions around SSH aspects in HRI and beyond.

4.3 Open toolkit for various audiences

The initial version of the ARISE SSH framework is purposefully public. Therefore, the intended audience of this work at this stage is all parties interested in SSH aspects in the context of HRI and technological innovation. These include parties such as the European Commission, researchers, practitioners and decision makers interested in the topic within scope of the ARISE project. We hope that our work also reaches the European HRI community. This includes our own project partners as well as the FSTP projects that will be supported by the ARISE project through cascade funding.

Our work builds on previous similar projects, initiatives and research and development in this area. As for researchers, we hope this framework provides practical viewpoints and insights into key ethical, legal and social considerations related to HRI. For practitioners, we hope this toolkit provides a comprehensive overview of the existing guidelines and tools and provides ideas for their further development. For decision-makers, in the following releases of the framework, we seek to provide practical and evidenced-based recommendations for HRI development and policy, in order to pave away for human-centric and sustainable technical developments, work environments and the future overall. For our partners, the framework provides a practical toolbox to the SSH foundations and key elements to be considered within the course of the project. Importantly, the framework will function as an open toolbox of SSH related aspects to be considered in the ARISE open call programs, first running in 2025-2026 and the second in 2026-2027. Hence, a key target group of this work is the FSTP projects associated with the ARISE project, who will be supported in the use of this framework.

4.4 Recommendations for HRI development and policy in future iterations of the ARISE SSH frameworks

Throughout the ARISE project, our goal is also to gather together lessons learned and possible policy recommendations for HRI development and policy in the European context. Based on our research with the ARISE partners and the FSTP projects, the insights gathered during the mentoring phase of the FSTP projects and more broadly the discussions within the ARISE consortium, we hope to provide decision-makers with up to date and empirically grounded recommendations on both the level of policy around HRI as well as EU-project related insights into how to incorporate SSH aspects into multidisciplinary projects like ARISE. These lessons, insights and recommendations will be included in the further iterations of the framework and published separately as a briefing paper/s in the following years of the project.

5 Conclusion

This report describes the initial open version of the ARISE SSH framework for human-centric and ethical HRI. The framework is a structured approach to embed key ethical, legal, and societal considerations into technological developments. The framework builds on the previous similar projects, initiatives, research and methodologies. The key elements of the initial open ARISE framework include: the ARISE HRI ethical principles, the ARISE ethics impact assessment, related industry standards and European regulation, collaboration and co-creation toolkit, typically encountered questions in AI systems, as well as a collection of relevant use cases with examples of their ethical concerns and questions. Rather than a one-size-fit-all solution, the framework serves as an open toolkit for various audiences. Altogether, the framework is to ensure industrial HRI projects are safe and consider a wide range of ethics related questions throughout the lifecycle of the technologies.

The framework is developed throughout the ARISE project in an iterative manner, always enabling the inclusion of fresh insights available at the time of its newest releases and incorporating the experiences from its use within the ARISE project. Important next steps in our work include continuing development of the key elements of the framework and preparing ourselves for the FSTP mentoring phase of the project. Our close collaboration with project partners continues, in addition to which, we are always open for discussions and partnerships with those outside of the ARISE project but sharing our vision for promoting more sustainable, human-centric and ethical technological innovations, work environments, as well as the future overall.