2025 International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA) | 979-8-3315-3562-9/25/$31.00 ©2025 IEEE | DOI: 10.1109/ACDSA65407.2025.11166650

Proc. of International Conference on Artificial Intelligence, Computer, Data Sciences and Applications (ACDSA 2025)
7-9 August 2025, Antalya-Türkiye

ETHAI implementation. A Co-Constructive way for Embedding Ethics in Healthcare AI Development

Lorena Volpini
CyberEthics Lab. Rome, Italy
l.volpini@cyberethicslab.com

Valerio Prosseda
CyberEthics Lab.
Rome, Italy
v.prosseda@cyberethicslab.com

Francesca Morpurgo
CyberEthics Lab.
Rome, Italy
f.morpurgo@cyberethicslab.com

Aristeidis Tsitiridis
DEA group, UBITECH, Athens, Greece; FRAV group, Rey Juan Carlos University, Spain; Cranfield University, Defence Academy, Shrivenham, UK
atsitiridis@ubitech.eu

Špela Glišović Krivec
Zavod vsesorte, izobraževanje o naravi in zdravju
Tolmin, Slovenia
spegli@gmail.com

David Krivec
Zavod vsesorte, izobraževanje o naravi in zdravju
Tolmin, Slovenia
david.krivec@i-vste.org

Jakob Sajovic
Department of Neurology University of Ljubljana, Faculty of Medicine
Ljubljana, Slovenia
jakob.sajovic@mf.uni-lj.si

Teja Štrempfel
Department of Neurology University of Ljubljana, Faculty of Medicine
Ljubljana, Slovenia
teja.strempfel@gmail.com

***Abstract*— This paper presents the ETHAI (ETHics of AI) methodology, a novel multidisciplinary approach to integrating ethical values within artificial intelligence systems designed for healthcare applications. While theoretical frameworks for AI ethics have proliferated, significant gaps persist between abstract principles and practical implementation. ETHAI addresses this disconnect through a co-constructive process that brings together ethicists, technical developers, clinical experts, and potential users to collectively define and operationalize ethical requirements. The methodology is characterized by four key principles: contextual awareness, practical application, cross-disciplinary collaboration, and iterative refinement. In specific, this paper reports on the preliminary results of implementation of ETHAI within COMFORTage, a European Commission-funded project developing AI systems for monitoring and supporting patients with dementia and age-related frailty. The findings highlight how the participatory risk assessment provided by the method helped identify domain-specific considerations requiring adaptation of generic ethics principles, while revealing significant cross-disciplinary communication challenges in translating ethical concerns into technical specifications. The semi-structured approach proved particularly effective in balancing normative frameworks with stakeholder engagement, demonstrating that meaningful ethical integration remains possible even when introduced at the design phase instead of project conception. This case study contributes valuable empirical evidence to the field of applied AI ethics and design, where practical implementation experiences remain underreported despite their importance for advancing responsible technology development.**

Keywords*— *AI ethics implementation, cross-disciplinary co-design, healthcare AI, ethics requirements translation

I. INTRODUCTION

As artificial intelligence research continues to advance, its applications increasingly penetrate social life, pervading several spheres of the service economy, public administration, and our daily lives, raising ethical concerns of growing urgency and complexity. While artificial intelligence as a field of research dates back to the 1950s—with theoretical debates on AI ethics nearly as old [1]—translating abstract ethical principles into concrete development practices remains a fundamental challenge. The last decade has witnessed an increase in research on AI ethics, and a consequent proliferation of theoretical frameworks. Policy statements, regulations, and domain-agnostic guidelines have somewhat rationalized the landscape, yet discussions focused on concrete applications remain rare [2],

Funded by the European Union Horizon Europe Programme, under agreement 101137301 - COMFORTAGE HORIZON-HLTH-2023-STAYHLTH-01

979-8-3315-3562-9/25/$31.00 ©2025 IEEE

Authorized licensed use limited to: Kennesaw State University. Downloaded on October 13,2025 at 18:38:34 UTC from IEEE Xplore. Restrictions apply.

while reports of practical experience in developing ethical AI systems are even harder to find. This implementation gap stems from multiple factors. [3] highlights how ethical considerations (such as privacy) must account for social contexts in ways that are difficult to encode programmatically. Within industry settings, commercial pressures and rapid development cycles often hamper ethical implementation despite genuine intentions [4]. Moreover, cross-disciplinary communication presents significant obstacles, with different disciplines developing distinct languages and frameworks—what [5] described as the "two cultures" problem—making it difficult to bridge ethics, computer engineering, and domain expertise. The disconnect between ethics and implementation reflects a deeper fracture between diverse communities with distinct languages, methodologies, and objectives. While ethics experts tend to formulate general principles and normative recommendations, developers usually require specific, measurable, and implementable guidance compatible with software architectures and system constraints. In the following sections the preliminary results of a Research and Innovation (R&I) initiative are presented as a way of addressing the above-mentioned gap. Section II presents the R&I project in which this experiment is being carried out and briefly introduces ETHAI methodology. Section III presents the results obtained, which are then discussed in Section IV. Conclusions are drawn in Section V.

II. The ETHAI methodology within COMFORTage Project

A. The COMFORTage project

The COMFORTage project addresses dementia and frailty challenges in aging populations through a comprehensive care model that spans the entire care continuum. This initiative implements a staged approach to intervention and prevention that combines clinical expertise with advanced computational methods. At its foundation, the research component employs statistical analysis and data modelling to identify contributing factors in dementia progression. These findings inform the development of preventive strategies and patient stratification methods, targeting identified risk pathways before significant cognitive decline manifests. This proactive approach represents a shift from traditional reactive care models. For individuals requiring clinical intervention, COMFORTage develops personalized care frameworks adaptable to specific patient circumstances. These frameworks incorporate AI-powered computational monitoring systems that analyze patient data to support evidence-based follow-up protocols, to simulate interventions and study their effects through digital twins in order to promote personalized care approaches. A distinguishing feature of the project is its implementation of data-driven assistive technologies that process multiple data streams to provide appropriate support and monitoring capabilities, balancing independent living with safety considerations.

The project's innovation lies in its integration of medical advances with artificial intelligence technologies. This includes novel approaches to risk factor analysis, personalized prediction models, AI-based medical devices, and aggregated clinical evidence sources. These clinical innovations work alongside AI developments such as explainable AI, patient digital twins, and virtual assistive technologies for clinical decision support and patient care. Solutions developed within COMFORTage, particularly those related to clinical decision-making, undergo comprehensive health technology assessment that includes ethics evaluation. This paper focuses specifically on this ethics assessment component. It shows how the ETHAI methodology developed and tailored to this project addresses the gap between domain-agnostic ethics requirements and practical implementation through a collaborative process involving ethicists, developers, and clinicians. It finally presents some partial results and lessons learnt.

B. The ETHAI methodolgical framework

The ETHAI framework represents an innovative approach to embedding ethical considerations within AI-based systems. Moving beyond traditional models that address ethics retrospectively, ETHAI adopts a proactive stance where ethical values guide the entire development journey from the earliest conceptual stages through to implementation [6]. The ETHAI methodology developed and implemented within COMFORTage seeks to be more than a technology assessment tool, it wants to enable a multidisciplinary co-design environment to facilitate “ethical” AI development. Unlike previous approaches that have attempted to “translate” ethics for developers, this method is based on the premise that AI ethics must be co-constructed through continuous interaction between ethicists, developers and dominion experts, as well as potential users (particularly clinical researchers, practitioners, care givers, patients and other users), recognizing the unique expertise and potential contribution of each group. For example, in contexts involving older adults with mental health conditions, clinicians, caregivers, and patients' families possess privileged perspectives on the care journey and its challenges. Their insights must be incorporated to identify potential impacts effectively. The ETHAI methodology places strong emphasis on process design, as this is where most uncertainties in ethical AI implementation reside. While the methodology's theoretical foundation has been detailed elsewhere [6,7], the focus in this paper is more on its implementation aspects.

At its core, ETHAI is characterized by four fundamental principles. The first is *contextual awareness*. It recognizes that ethical frameworks must be adapted to specific environments—particularly crucial in fields like healthcare, where vulnerable populations are involved. The second pillar is *practical application*, focusing on converting theoretical ethical concepts into specific, measurable requirements for technical teams. The third element is *cross-disciplinary collaboration*, bringing together ethicists, technical experts, field specialists, and users to ensure comprehensive perspective integration. The fourth pillar, *iterativity*, embraces continuous improvement and responsiveness to new challenges through repeated cycles of development and assessment [7].

The process unfolds through four interconnected phases that form a continuous cycle. The Ethical Requirements Identification phase involves an interdisciplinary team analyzing potential ethical implications of the AI system. This analysis draws on established frameworks such as EU Guidelines, Regulation (e.g. the EU AI Act) bioethics principles, and care ethics, while also incorporating stakeholder values and concerns [7] as well specific dominion-related perspectives,

Authorized licensed use limited to: Kennesaw State University. Downloaded on October 13,2025 at 18:38:34 UTC from IEEE Xplore. Restrictions apply.

through risk assessments. In the second phase, Requirement Translation, high level ethics and regulatory requirements are transformed into actionable system requirements. For instance, the principle of transparency might be operationalized through requirements for explainable algorithms, user interfaces that clearly indicate AI involvement and highlight crucial decisions' output with an explanation and then translated in specifications. The third phase, Implementation and Refinement, sees technical developers implementing these requirements while maintaining ongoing dialogue with the ethics team. This collaborative approach facilitates mutual learning and adaptation between ethical and technical perspectives, ensuring that implementation challenges inform requirement refinement. The final phase, Assessment and Evaluation, determines whether the system meets the established ethics requirements through technical testing, user feedback and expert review where needed. Identified shortcomings trigger a return to the first phase, creating a continuous improvement loop.

This cyclical structure acknowledges that ethical considerations in AI are dynamic, evolving alongside technological advances and shifting societal values. By integrating ethics into every development stage, ETHAI helps create AI systems that not only perform their intended functions effectively but also dynamically align with human values and ethical principles. This methodology is particularly valuable in sensitive contexts like healthcare, where ethical missteps can have significant consequences for vulnerable populations.

C. *Positioning ETHAI in the Landscape of AI Assessment Methodologies*

Most current AI ethics assessment, evaluation tools are designed for specific contexts - either internal company audits, external regulatory or standard compliance auditing, or academic research settings. These approaches differ primarily along two dimensions: the scope of inquiry and the degree of independence of the examiner [8]. This latter dimension reflects how much control the entity being evaluated has over the nature of the assessment and over the framing of its outcomes. Within this landscape, ETHAI as implemented within COMFORTage project occupies a position of moderate independence due to the unique governance structure of European-funded R&I Actions. The organization leading this work is a member of the COMFORTage consortium yet it maintains a distinct role representing public interest without the vested incentive to declare AI systems as "ethical" at all costs. This balanced positioning enables critical assessment while maintaining collaborative relationships with technical partners.

The scope dimension regards the breadth of an AI system inquiry. Holding that the difference among scopes exists on a spectrum from narrowest to broadest rather than distinct types, it is possible to differentiate "specific" assessments, when the analysis revolves around a specific harm or impact using a defined benchmark, "focused", when a specific harm is evaluated against a procedural requirement, "structured", when a set of harms is considered within a defined taxonomy and "exploratory", which corresponds to a broad exploration of possible harms and impacts of a system [8]. Again, the distinctive environment of COMFORTage as a EU funded R&I project allows an implementation of ETHAI through a hybrid approach that combines structured assessment with exploratory inquiry into wider impacts. In practice, this means starting with the EU Guidelines for Trustworthy AI [9,10] as a defined taxonomy based on established policies and practices [8], while simultaneously conducting exploratory investigations that identify potential impacts in an open-ended manner. Such a complementary inquiry is facilitated by opportunities to engage with external stakeholders (provided by the Community Care Forum, established and mobilized by the consortium) and representatives of potentially affected communities and social groups. Within this framework, a team of AI ethicists from CyberEthics Lab. interacts with diverse participants possessing domain expertise and the relevant perspective to conduct robust preliminary explorations of potential impacts.

The methodology enables a comprehensive assessment of system impacts across technical, human, and societal dimensions. It examines immediate effects, cumulative harms that may develop over time, and systemic impacts arising from complex interactions or second-order effects. This phase represents an initial risk identification (see Fig. 1) as a preliminary investigation of potential unwanted impacts that will be systematically refined in subsequent phases. The dual focus—combining structured evaluation based on established frameworks with expansive exploration of potential impacts—positions this approach as a distinctive methodology that bridges the gap between abstract ethical principles and practical implementation within the evolving landscape of AI assessment approaches. Furthermore it illustrates ETHAI's ability to adapt across different scopes as well as diverse independence levels, making it much more versatile and practically useful also beyong European R&I actions context.

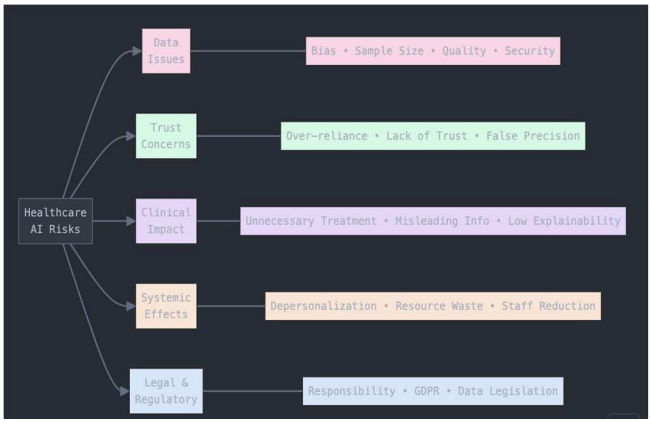

Fig. 1. First classification of risks identified through participatory approach

III. Results: ETHAI Implementation in the COMFORTage Project

A. *Interdisciplinary Implementation Process*

The ETHAI methodology was implemented within the COMFORTage project, revealing significant insights into embedding ethics in AI healthcare development. The consortium included 39 organizations with diverse expertise: technical partners, university researchers, clinical practitioners, computer scientists, social scientists, and humanists. This

Authorized licensed use limited to: Kennesaw State University. Downloaded on October 13,2025 at 18:38:34 UTC from IEEE Xplore. Restrictions apply.

diversity created an ideal environment for testing the methodology's effectiveness. Due to the project's 4-year timeline, we adopted a parallel approach rather than sequential development. This allowed ethics development and technical work to proceed simultaneously, demonstrating ETHAI's adaptability to real-world conditions.

B. Implementation of ETHAI in COMFORTage

Given the distributed expertise across project partners, embedding ethicists directly within technical teams was not feasible. Instead, we created multidisciplinary working groups. Ethicists and social scientists first collaborated with technical managers to establish a high-level organizational framework. Developer groups were then organized according to their involvement in AI system components, with multiple contact points to the ethics team

From an operational perspective, an **AI inventory** was conducted to identify the types of applications that would be developed within the high-level architecture. The inventory examined several key aspects, such as: What AI models and components each organization would develop, the types of applications (Machine Learning, Generative AI, etc.), the learning approaches used (supervised, unsupervised, deep learning, NLP). each model's purpose and role in the overall architecture, data inputs for training and execution and expected model outputs. This inventory was designed as an iterative practice, updated throughout the project as changes occurred. Not all information was clear at project inception, allowing for adaptation over time.

Once the consortium agreed on system architecture and produced diagrams outlining the AI components and their interactions, development teams began implementation. The ethics team conducted a comprehensive study on AI risks in healthcare for mental health and older adults. Based on this research, they identified 12 high-level requirements inspired by EU Guidelines for Trustworthy AI principles. These requirements were compiled into a "translation table" that acts as a support to convert ethical principles into technical requirements and specifications. Each requirement was developed into approximately 10 specific sub-requirements, each to be implemented through specifications identified and suggested by the developers.

Figure 2 shows the high-level ethics requirements, while Figure 3 demonstrates, as an example, the unpacking of R1 Autonomy and Oversight requirement into sub-requirements (requirement suggestions) and specifications (developers' team feedback). Once all the specifications will be agreed, the same table will integrate information collected across different developers' teams with focus on different components and allow for assigning different levels of priority.

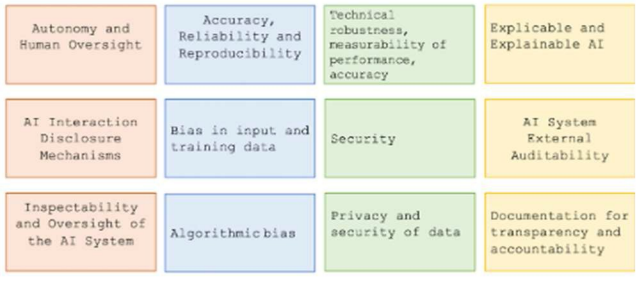

Fig. 2. High-Level Requirements

ID	REQUIREMENT SUGGESTIONS	DEV. TEAM FEEDBACK	TYPE	COMPONENT
R1.1	Develop a classification module to automatically flag decisions as 'crucial' based on predefined criteria stored in a configurable ruleset. Human approval shall be required before executing any decisions classified as 'crucial.'	Introduce a decision classification mechanism within ICML that flags AI-driven outputs as informative, suggestive, or crucial.	Non functional	ICML & XAI
R1.2	Human approval workflows for flagged decisions.	Introduce a flagging mechanism in the AI model output pipeline, When flagged, route the decision to a human reviewer panel or clinician dashboard. Require explicit approval (e.g., checkbox + digital signature or credentials).	Non functional	ICML & CDSS
R1.3	Options to provide feedback on decisions and override system outputs.	Provide users, particularly clinicians and healthcare professionals with the ability to provide feedback on AI-generated decisions and, when necessary, override the system's outputs.	Functional	CDSS & ICML
R1.4	Clear visual cues for flagged decisions requiring human attention.	Provide visual indicators both inline and in list views (e.g., overview tables). Connect UI elements with underlying ASM tags (e.g., flagged=True, urgency=high).	Non functional	CDSS & ICML
R1.5	Implement **user control buttons** to accept, reject, or modify AI-generated decisions, ensuring this functionality is prominently displayed in the interface. Simple and intuitive for users to access.	CDSS handles the presentation, the core logic for handling AI decision feedback and override actions is tied directly to ICML.	Functional	ICML & CDSS
R1.6	A feedback module shall be implemented that allow users to : 1) flag specific decisions as critical or requiring review	This requirement supports a human-in-the-loop feedback loop, enabling users (e.g., clinicians or reviewers) to proactively signal concerns about AI outputs. ICML must be able to accept these feedback signals, record them in the audit trail, trigger downstream workflows or notifications when flagged.	Functional	ICML & IKB
R1.7	2) Provide qualitative feedback on decision accuracy or relevance.	The ICMI should support continuous learning, model validation, and human-AI collaboration. ICML must accept qualitative user input tied to specific predictions, log this feedback for review or possible future retraining and use decision-makers' domain expertise contributes to AI improvement.	Functional	ICML
R1.8	3) Design a **stop-and-seek-advice mechanism** where flagged decisions trigger system alerts and pause execution pending further review.	Mechanism must be implemented within the ICML logic that classifies, flags, and controls execution flow for AI outputs	Functional	ICML & CDSS & IKB
R1.9	4) Log user feedback in a structured database to inform continuous system improvements.	ICML is directly responsible for structuring and storing feedback tied to specific decision outputs, enabling this feedback to be accessed for model auditing, retraining, or user experience adjustments. This supports a feedback loop architecture, critical for a learning health system.	Non functional	ICML & IKB & CDSS
R1.10	Implement an **analytics module** to monitor user feedback patterns and identify recurring issues or areas for refinement.	User feedback may be collected at the UI level, ICML is responsible for analysing feedback data for quality improvement, detecting patterns related to decision errors, clinician concerns, or low-trust interactions, enabling continuous learning and system evolution based on real-world usage.	Functional	ICML & IKB & CDSS

Fig. 3. A picture of the Requirements Translation Table

C. Current implementation status

At the time this paper was written, technical partners were evaluating the high-level requirements and sub-requirements, assessing the relevance for their component and deriving specifications, through dialogue with the ethicists. After consolidating a list of specifications, with indicators and acceptance criteria, both technical partners and ethicists will prioritize agreed requirements through collaborative discussion. This represents the first iteration of ETHAI methodology focused on designing the overall COMFORTage AI system, which includes: Integrated Care Model Library (ICML) of AI algorithms, Explainable Models and Algorithms (XAI), Real-time Feedback and Behavioral Analytics, Clinical Decision Support System Suite (CDSS) with Visual Analytics Tools.

As a preliminary result of developers-ethicists collaboration, consider requirement *R1. Autonomy and Oversight*. This requirement applies to the entire AI-based COMFORTage system. The complete R1 formulation is: *Implement mechanisms to preserve human decision-making authority in crucial matters. The AI system shall facilitate "human in the loop" and "human on the loop" frameworks, enabling operators to provide feedback on AI-generated outputs and allowing the system to learn and enhance its performance*. Development teams working on different components must translate this requirement into specifications. Fig. 3 shows the ICML developers team's work on the translation table. Feedback from the other developer teams will be integrated to gather specifications for all the main components.

Authorized licensed use limited to: Kennesaw State University. Downloaded on October 13,2025 at 18:38:34 UTC from IEEE Xplore. Restrictions apply.

In the second iteration, requirements will be tailored to 13 specific clinical studies as pilot use cases. This phase will incorporate: specific needs and concerns of users (especially from clinical partners) and input collected through participatory risk assessment (cf. fig. 1). At this stage, clinical and technical partners will work in a more integrated manner than in previous phases. External stakeholders' input will be gathered through interviews and workshops.

Parallel to requirements co-design, an in-depth study examines ethics risks, as perceived by clinicians, within each pilot use case. This builds upon initial risk identification shown in Figure 1.

D. Case Study: "Mind & Mouth" Clinical Study

The Faculty of Medicine at the University of Ljubljana is conducting a key implementation through the "Mind & Mouth" study. This research investigates the complex relationships between oral health, cognitive function, and overall wellbeing in older adults. The study employs multiple assessment approaches including oral health evaluations, neuroimaging techniques to assess brain activity, and cognitive assessments. These are combined with personalized interventions such as social learning programs and serious games delivered via tablets. Data from these assessments is fed into AI-based analytics tools that stratify patients according to their risk of dementia and frailty. Clinical researchers evaluate these AI-generated risk assessments against their own established methods of diagnosis. Among the goals there is to enhance diagnostic with AI-based analytics. The role of human medical expertise is key to advance these methods that in the future may support in determining appropriate interventions to manage cognitive decline risk, mild cognitive impairment, and early-stage dementia.

Within this pilot study, five specific AI ethics risks have been identified. The first is related to algorithmic bias leading to potential discrimination in AI decision-making. Second, possible privacy vulnerabilities raise questions about protecting sensitive patient data. Third, difficulties in understanding AI decision processes is considered through the lens of explainability challenges (partially addressed through XAI implementation). Fourth, there is a risk of clinicians' overreliance on AI, potentially overshadowing clinical judgment. Last, the challenge of balancing AI recommendations with clinical judgment relates directly to maintaining human professionals' autonomy while integrating AI support systems. These risks are addressed by the EU Guidelines for trustworthy AI and technical requirements already outlined (see section III B and Fig. 2). Clinical researchers remain actively involved in ongoing assessment to ensure mitigation, through requirements refinement.

E. Bridging theory and practice

Many concerns identified by clinical researchers relate to the COMFORTage AI system use rather than design development. However, specific discussion sessions will focus on design-based mitigations for each use case's anticipated impacts. These targeted sessions will bridge the gap between theoretical risk assessment and practical implementation, ensuring that ethical concerns from the initial phase are adequately addressed, where possible, in the technical design. Requirements obtained through this process will be refined and improved for specific applications. Indicators and acceptance criteria will be added to fine-tune the assessment questionnaires, evaluate case-based implementation, and produce recommendations for future phases.

IV. Discussion

The described ETHAI implementation stages used the EU Guidelines for Trustworthy AI as a foundation, providing a domain-agnostic ethical framework. This process established baseline ethical requirements [6,7] that would undergo subsequent improvement and refinement.

A. Integration of Domain-Specific Ethical Considerations

Generic ethical guidelines have inherent limitations when applied to specific contexts. To address this, a parallel participatory risk assessment process was initiated, focusing on healthcare-specific and neuroethical considerations, relevant to older patients with dementia. This process revealed that while the EU guidelines provided valuable general principles, they required significant adaptation to address several critical areas: healthcare process integration challenges, neuroethics considerations specific to dementia, vulnerabilities of elderly patients and their families, and care relationship dynamics. Through structured discussions with clinical researchers and practitioners, we identified several critical risks. Most were considered potentially mitigable through AI governance and requirements for users and deployers. However, some risks—such as system over-reliance—require technical design interventions at the development stage to be effectively addressed.

B. Interdisciplinary Communication Challenges

The translation phase, still underway, has already provided valuable insights into interdisciplinary communication. Early meetings between ethics and technical teams revealed significant terminology and conceptual gaps. Technical partners noted they were accustomed to receiving requirements from users rather than co-designing them with ethics specialists. This necessitated adjustments to established development workflows and increased attention to communication. However, this process resulted in a positive secondary outcome: increased awareness of AI ethics implications among developers.

Requirements underwent multiple reformulations, being reframed using terminology familiar to technical teams. This process also educated ethicists about technical reasoning approaches. Moving forward, we anticipate that ambiguous concepts will need to be operationalized into measurable criteria, requirements will undergo negotiation regarding formulation and specification, and technical feasibility assessment will be conducted.

The refinement process is converging towards 12 consolidated primary requirements, unpacked into more specific implementation guidance (referred as sub-requirements and consequent specifications). This increased granularity provides actionable direction for development teams working on the COMFORTage system.

C. Manifestations of Cross-Disciplinary Barriers

Face-to-face collaborative sessions between ethicists, technical researchers and clinical researchers revealed

Authorized licensed use limited to: Kennesaw State University. Downloaded on October 13,2025 at 18:38:34 UTC from IEEE Xplore. Restrictions apply.

communication challenges influenced by domain-specific factors. For example, ethicists might raise concerns about patient autonomy in cognitive monitoring systems without specifying operational details. For instance they may not explain how concerns should translate into consent protocols, data access controls, or override mechanisms that technical professionals could implement. At the same time, varying ethical priorities have been observed across different clinical studies' contexts, suggesting the need for context-specific technical approaches and risk mitigation strategies for each use case. These findings will shape the second iteration of our methodology.

D. Methodological Insights

The semi-structured approach here adopted proved advantageous, if compared to a rigid methodology. Beginning with normative agreement on the EU Guidelines and then adapting them through participatory engagement, achieved greater stakeholders' buy-in than a more prescriptive approach would have allowed.

The timing of ethics integration emerged as a crucial factor. While true "ethics by conception" would ideally begin at project formulation, our results demonstrate that meaningful ethical integration can begin at the design phase and still significantly influence system development.

Some of the above-mentioned disciplinary barriers could be addressed with improved tools and processes. For example, ethics-to-implementation mapping tools could visually connect abstract ethical principles and concrete technical features, improving communication. For example, mapping "human autonomy" to specific features like user control functions and interfaces and explainability modules. These would outline specific ethical priorities for each clinical context, mapping them to risks and technical requirements. Such profiles could guide customized implementations and assessments, preventing potential communication challenges.

V. CONCLUSION

The ETHAI methodology developed and implemented within the COMFORTage project demonstrates a promising approach to bridging the persistent gap between abstract ethical principles and concrete technical implementation in AI systems for healthcare. This experience reveals that effective ethics sensitive AI development requires not merely translation but co-construction of requirements through sustained interdisciplinary dialogue. The iterative, participatory process adopted has shown several advantages over conventional approaches. By engaging diverse stakeholders from ethics, technical development, and healthcare domains in collaborative requirements development, this method addresses cross-disciplinary communication barriers while increasing ownership and understanding across teams. This co-design approach fosters greater appreciation of ethics considerations among technical partners and deeper understanding of implementation constraints among ethicists.

A number of goals may benefit from this approach. These range from mapping relevant characteristics of a system to identify potential gaps and issues to determining targets for further study, scrutiny, or monitoring. The findings from such processes can help facilitate deliberation and generate consensus around how impacts should be defined and prioritized, and the appropriate methods to detect and remediate them.

The semi-structured, flexible nature of this methodology proved particularly valuable in the complex healthcare context, where domain-specific ethical considerations—such as those related to dementia care and older adults' vulnerability—require careful attention. While implementation remains ongoing, preliminary results suggest that meaningful ethics integration can occur even when introduced at the design phase rather than project conception.

Future work should focus on refining methods for operationalizing abstract ethical concepts into measurable technical criteria, expanding participatory risk assessment across diverse stakeholder groups, and developing more sophisticated tools to support interdisciplinary communication. Additionally, longitudinal studies tracking the impact of early ethical integration on final system outcomes would provide valuable evidence for the efficacy of such approaches.

By documenting both the successes and challenges encountered in implementing ETHAI within a real-world healthcare AI development project, this paper contributes to the growing body of practical experience in ethical AI development—an area where concrete case studies remain scarce despite abundant theoretical frameworks.

REFERENCES

[1] L. Floridi and J. Cowls, 'A Unified Framework of Five Principles for AI in Society', *Harvard Data Science Review*, vol. 1, no. 1, Jul. 2019, doi: 10.1162/99608f92.8cd550d1.

[2] J. Burstein, G. T. LaFlair, K. Yancey, A. A. von Davier, and R. Dotan, 'Responsible AI for Test Equity and Quality: The Duolingo English Test as a Case Study', 2024, *arXiv preprint arXiv:2409.07476.* [Online]. Available: https://arxiv.org/pdf/2409.07476

[3] H. Nissenbaum, 'Privacy as Contextual Integrity', *Wash. L. Rev.*, vol. 79, p. 119, 2004.

[4] M. Whittaker *et al.*, 'AI now report', AI Now Institute at New York University, New York, 2018. [Online]. Available: https://www.elindependiente.com/wp-content/uploads/2018/12/Informe-AI-Now.pdf

[5] C. P. Snow, *The Two Cultures*. Cambridge University Press, 2012.

[6] COMFORTage, 'D7.1- Ethics requirements, guidelines, and best practices for trustworthy AI I', project report, 2024.

[7] L. Volpini et al. "The ETHAI Methodology A context-sensitive, actionable approach to ethics of AI in Healthcare," in *Artificial Intelligence Applications and Innovations. AIAI 2025 IFIP WG 12.5 International Workshops*, in IFIP Advances in Information and Communication Technology, vol. 754. , 2025. (forthcoming)

[8] M. Bogen, 'Assessing AI: Surveying the Spectrum of Approaches to Understanding and Auditing AI Systems', Center for democracy and technology, 2025. Accessed: Mar. 13, 2025. [Online]. Available: https://cdt.org/insights/assessing-ai-surveying-the-spectrum-of-approaches-to-understanding-and-auditing-ai-systems/

[9] N. A. Smuha, 'The EU Approach to Ethics Guidelines for Trustworthy Artificial Intelligence', 2019, *Rochester, NY*: 3443537. Accessed: May 06, 2024. [Online]. Available: https://papers.ssrn.com/abstract=3443537

[10] Directorate-General for Communications Networks, Content and Technology (European Commission) and Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji, *Ethics guidelines for trustworthy AI*. Publications Office of the European Union, 2019. Accessed: Mar. 16, 2024. [Online]. Available: https://data.europa.eu/doi/10.2759/346720

[11] A. Jobin, M. Ienca, and E. Vayena, 'The global landscape of AI ethics guidelines', *Nat Mach Intell*, vol. 1, no. 9, pp. 389–399, Sep. 2019, doi: 10.1038/s42256-019-0088-2

Authorized licensed use limited to: Kennesaw State University. Downloaded on October 13,2025 at 18:38:34 UTC from IEEE Xplore. Restrictions apply.