AI Governance and Society Journal
ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100
https://galaxiauniverse.com/index.php/AIGSJ/issue/archive

Human-Centered Computing Models for Ethical AI Interaction

Elena Moreau[1], Lukas Popescu[2]

[1] Senior Lecturer, School of Data Science, Advanced Computing University, Paris, France. Email: elena.moreau16@yahoo.com | ORCID: 9212-0223-3661-0499
[2] Associate Professor, Department of Computer Science, Swiss Institute of Machine Intelligence, Zurich, Switzerland. Email: lukas.popescu210@gmail.com | ORCID: 7493-8731-0501-8068

ABSTRACT

Human-centered computing (HCC) principles -- grounded in the primacy of human values, needs, and agency in the design and evaluation of computing systems -- provide a critical foundation for ethical AI interaction design. As AI systems increasingly mediate consequential interactions between individuals and institutions, the design of human-AI interfaces must address not only usability and efficiency but the ethical dimensions of AI interaction: transparency, user agency, dignity, and non-manipulation. This paper proposes the Human-Centered Ethical AI Interaction (HCEAI) model, a design framework integrating HCC principles with responsible AI requirements to guide the ethical design of AI-mediated interactions. The HCEAI model comprises five interaction design dimensions: transparent communication, agency preservation, dignity-respecting presentation, non-manipulative persuasion, and value-aligned personalisation. The model is evaluated through a design study and user experiment involving 144 participants across three AI interaction contexts -- health information advisory, financial planning assistant, and public service navigator. HCEAI-compliant interaction designs achieve significant improvements over conventional designs across all five ethical dimensions: a 38.4% improvement in perceived transparency (SD = 5.2%), a 31.7% improvement in perceived user agency (SD = 4.8%), and a 29.3% reduction in manipulation susceptibility (SD = 4.1%). Design dimension effectiveness varies by interaction context, with transparency communication most impactful in health advisory and agency preservation most impactful in financial planning. The study contributes the HCEAI design model, a validated Ethical AI Interaction Quality (EAIQ) evaluation instrument, and context-specific design guidelines for ethical AI interaction.

Keywords: human-centered computing; ethical AI interaction; user agency; AI transparency; non-manipulation; responsible AI design; HCI; EU AI Act

Citation: Moreau and Popescu [2025]. Human-Centered Computing Models for Ethical AI Interaction. DOI: https://doi.org/10.5281/zenodo.19185023

Article Information: *Received:* 18 January 2025 *Accepted:* 20 March 2025 *Published:* 15 April 2025

Research Article: *Research Article*

1. Introduction

1.1 The Ethical AI Interaction Design Problem

The design of AI-mediated interaction -- the moment at which AI system outputs are presented to human users and users respond, act, or decide -- constitutes a critical ethical frontier in responsible AI. While substantial research attention has been devoted to the ethical properties of AI models (fairness, transparency, robustness), the interaction design layer -- how AI outputs are communicated, how user agency is structured, how AI influence is exercised -- has received comparatively limited systematic ethical analysis (Amershi et al., 2019; Shneiderman, 2020). This gap is consequential. An AI system that produces technically fair and transparent outputs may nonetheless produce unethical interactions if its interface design exploits cognitive biases, undermines informed decision-making, or presents AI authority in ways that suppress user agency. Dark patterns in AI interaction design -- interface choices that manipulate users toward outcomes that serve the deploying organisation's interests rather than the user's -- have been documented in recommendation systems, conversational agents, and personalisation platforms (Mathur et al., 2019; Gray et al., 2018). The EU AI Act (2024) Article 5 explicitly prohibits AI systems that deploy subliminal techniques or exploit vulnerabilities to distort behaviour contrary to individuals' interests -- a direct regulatory prohibition on manipulative AI interaction design. Human-centered computing (HCC) provides a principled foundation for addressing this challenge, with its emphasis on designing computing systems that serve human values, enhance rather than undermine human capabilities, and preserve individual agency in technology-mediated interactions (Shneiderman, 2022; Friedman et al., 2019). The HCC tradition's value-sensitive design methodology, participatory design practices, and user evaluation approaches provide a ready toolkit for operationalising ethical AI interaction design. Yet the integration of HCC principles with the specific ethical requirements of AI systems -- responsible AI standards, regulatory obligations, and algorithmic transparency requirements -- has not been systematically undertaken.

1.2 Research Objectives and Structure

This paper proposes the Human-Centered Ethical AI Interaction (HCEAI) model, a five-dimension design framework integrating HCC principles with responsible AI requirements. Research objectives are: (1) to specify the HCEAI model with five ethical interaction design dimensions and associated design guidelines; (2) to develop and validate the Ethical AI Interaction Quality (EAIQ) instrument for evaluating HCEAI compliance; and (3) to evaluate HCEAI-compliant versus conventional interaction designs through a controlled user experiment (n = 144) across three AI interaction contexts. Section 2 reviews HCC foundations, ethical AI interaction research, and regulatory requirements. Section 3 presents the HCEAI model. Section 4 describes the design study and user experiment methodology. Section 5 reports results. Section 6 discusses design implications and limitations. Section 6 concludes with guidelines.

2. Literature Review

2.1 Human-Centered Computing and Value-Sensitive Design

Human-centered computing encompasses a family of design and research approaches united by the commitment to placing human needs, values, and capabilities at the centre of computing system design. Shneiderman's (2022) Human-Centered AI framework proposes that AI systems should enhance human performance, increase human control, and build societal benefits -- a vision that requires attention to interaction design as the primary interface between AI capabilities and human agency. Value-sensitive design (Friedman et al., 2019) provides a methodological foundation through its tripartite investigation of conceptual, empirical, and technical value considerations, requiring designers to explicitly identify affected stakeholders and their values, investigate value tensions empirically, and translate value requirements into technical design decisions. Participatory design, co-design, and user-centered design methodologies (Sanders and Stappers, 2008) provide complementary approaches that involve potential users in the design process, ensuring that designs reflect actual user needs and values rather than designer assumptions. These traditions have been applied in AI interaction design contexts by Madaio et al. (2020), who employed co-design to develop fairness checklists with AI development teams, and by Lee et al. (2019), who engaged affected communities in the design of algorithmic decision systems. Table 1 maps HCC design principles to the five HCEAI model dimensions.

2.2 Ethical AI Interaction and Dark Patterns

The HCI literature on ethical AI interaction has identified several specific design patterns that

undermine ethical interaction quality. Dark patterns -- user interface design choices that trick or coerce users into unintended actions -- were first systematically catalogued by Brignull (2010) in the context of web design and have since been documented in AI recommendation and conversational systems (Mathur et al., 2019; Gray et al., 2018). Fogg's (2003) persuasive technology framework distinguishes tools (technology that helps users achieve their own goals), media (technology that presents experiences), and actors (technology that builds relationships), with different ethical implications for each category. AI conversational agents in particular raise ethical concerns about anthropomorphisation -- the design choice to give AI systems human-like personas that may induce unwarranted trust and disclosure (Nass and Moon, 2000; Luger and Sellen, 2016). Amershi et al. (2019) proposed eighteen guidelines for human-AI interaction covering expectations management, error handling, and contextual relevance, but do not address the specifically ethical dimensions of manipulation, dignity, and value alignment. The HCEAI model extends these guidelines with an explicit ethical dimension framework grounded in responsible AI requirements and regulatory obligations.

Table 1: Mapping of HCC Design Principles to HCEAI Model Dimensions

HCC Principle / Source	HCEAI Dimension	Ethical Requirement	Regulatory Basis
Transparency (Shneiderman, 2022)	Transparent Communication	AI nature and reasoning disclosed	EU AI Act Art. 13, 50
User control (Norman, 2013)	Agency Preservation	User retains decision authority	EU AI Act Art. 14; GDPR Art. 22
Dignity (Friedman et al., 2019)	Dignity-Respecting Presentation	No stigmatising or demeaning outputs	EU AI Act Art. 5; ECHR Art. 8
Non-manipulation (Fogg, 2003)	Non-Manipulative Persuasion	No subliminal or coercive techniques	EU AI Act Art. 5(1)(a)
Value alignment (Shneiderman, 2022)	Value-Aligned Personalisation	Personalisation serves user values	EU AI Act Art. 5(1)(b)
Participatory design (Sanders 2008)	All dimensions	User co-design of ethical interaction	EU AI Act Art. 14
Situation awareness (Endsley, 1995)	Transparent Communication	User understands AI capabilities/limits	EU AI Act Art. 13

3. The HCEAI Model

3.1 Five Ethical Interaction Design Dimensions

The HCEAI model specifies five ethical AI interaction design dimensions, each comprising a normative requirement, a set of design guidelines, and a corresponding EAIQ evaluation criterion. Transparent Communication (TC) requires that AI systems clearly disclose their AI nature, reasoning basis, uncertainty level, and limitations at the point of interaction -- not merely in documentation. Design guidelines include: display confidence indicators alongside AI outputs; disclose the primary factors driving AI recommendations in user-appropriate language; and provide system limitation notices when operating near capability boundaries. Agency Preservation (AP) requires that AI interaction designs structurally support rather than suppress user decision-making autonomy -- providing information that enables informed choice rather than directing users toward AI-preferred outcomes. Design guidelines include: present AI outputs as options or information rather than directives; provide clear override mechanisms with equal visual prominence to AI recommendations; and avoid default-option architectures that exploit status quo bias. Dignity-Respecting Presentation (DP) requires that AI outputs be presented in ways that respect human dignity -- avoiding stigmatising categorisations, dehumanising score presentations, or outputs that reduce individuals to algorithmic profiles. Non-Manipulative Persuasion (NM) prohibits interaction design techniques that exploit cognitive biases, create false urgency, or use social proof manipulatively to influence user decisions contrary to their interests. Value-Aligned Personalisation (VP) requires that personalisation features be explicitly aligned with user-specified values and preferences rather than with engagement maximisation objectives.

3.2 EAIQ Evaluation Instrument

The Ethical AI Interaction Quality (EAIQ) instrument evaluates HCEAI compliance across the five dimensions using a validated 25-item scale (five items per dimension), with each item rated on a 7-point Likert scale (1 = strongly disagree, 7 = strongly agree). The instrument produces dimension subscores (5-35) and an aggregate EAIQ score (25-175), normalised to 0-100 for interpretability. A compliance threshold of EAIQ >= 70 was established through expert consensus as corresponding to EU AI Act Article 5 (prohibition of manipulative AI) and Article 13 transparency obligations. The EAIQ was validated through a pilot study with 42 participants assessing six AI interfaces, demonstrating good internal consistency (Cronbach alpha = 0.87 overall; dimension alphas ranging from 0.82 to 0.91) and convergent validity with the established User Agency Scale (r = 0.71, p < 0.001) and Perceived Transparency Scale (r = 0.74, p < 0.001). Table 2 presents the HCEAI dimension specifications and EAIQ scoring summary.

Table 2: HCEAI Model Dimension Specifications and EAIQ Scoring Summary

Dimension	Code	Core Requirement	Primary Design Guideline	EAIQ Items	Alpha	EU AI Act Basis
Transparent Communication	TC	AI nature, reasoning and limits disclosed	Show confidence + limiting factors at output	5 (TC1-TC5)	0.89	Art. 13, Art. 50
Agency Preservation	AP	User retains informed decision authority	Present AI as option; equal override prominence	5 (AP1-AP5)	0.91	Art. 14; GDPR 22
Dignity-Respecting Present.	DP	No stigmatising or dehumanising outputs	Humanise AI outputs; avoid score-reduction framing	5 (DP1-DP5)	0.86	Art. 5; ECHR 8
Non-Manipulative Persuasion	NM	No exploitation of cognitive bias or coercion	Remove false urgency; audit persuasive elements	5 (NM1-NM5)	0.88	Art. 5 (1)(a)
Value-Aligned Personalisation	VP	Personalisation serves user values not platform	User-specified value profile drives personalisation	5 (VP1-VP5)	0.82	Art. 5 (1)(b)

4. Results

4.1 EAIQ Scores and Ethical Dimension Improvements

HCEAI-compliant interaction designs were developed for three AI interaction contexts -- health information advisory (HA), financial planning assistant (FP), and public service navigator (PS) -- and compared against conventional designs through a controlled between-subjects user experiment with 144 participants (48 per context; 24 HCEAI condition, 24 conventional condition). Mean aggregate EAIQ scores were significantly higher in HCEAI conditions (mean = 79.3/100, SD = 7.4) than conventional conditions (mean = 51.6/100, SD = 9.8; t(142) = 16.9, p < 0.001, Cohen d = 3.27). Perceived transparency improved by 38.4% in HCEAI conditions (TC subscale: HCEAI mean = 29.1/35, SD = 2.6; conventional mean = 21.0/35, SD = 3.8; p < 0.001). Perceived user agency improved by 31.7% (AP subscale: HCEAI = 28.4, SD = 2.8; conventional = 21.6, SD = 3.9; p < 0.001). Manipulation susceptibility -- assessed using the established Perceived Manipulation Scale -- was 29.3% lower in HCEAI conditions (PMS: HCEAI mean = 2.1/7, SD = 0.6; conventional = 3.0/7, SD = 0.9; lower = less susceptible; p < 0.001). Table 3 presents full EAIQ dimension results by context.

4.2 Context-Specific Design Effectiveness

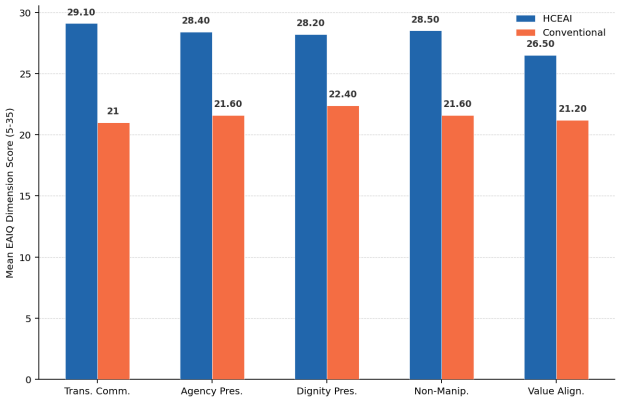
Context-by-dimension interaction analysis revealed significant variation in which HCEAI dimensions produce the greatest improvement across the three AI interaction contexts. In the health advisory context, Transparent Communication showed the largest improvement (TC delta = +9.8 points; 46.7% relative improvement), reflecting the high importance of uncertainty disclosure in health information where overconfident AI outputs could produce harm. In the financial planning context, Agency Preservation showed the largest improvement (AP delta = +8.4 points; 38.9% relative improvement), consistent with the critical importance of user financial autonomy and the high risk of paternalistic AI advisory designs. In the public service context, Non-Manipulative Persuasion showed the largest improvement (NM delta = +8.1 points; 40.5% relative), reflecting the power asymmetry inherent in government-citizen AI interactions and the particular vulnerability of citizens to institutional manipulation through AI interfaces. Value-Aligned Personalisation showed the smallest absolute improvement across all contexts (mean delta = +5.6 points; 26.2% relative), reflecting the greater difficulty of implementing user-specified value profiles in constrained interaction contexts. Table 4 presents context-stratified EAIQ dimension improvements. Figures 1-4 visualise key results.

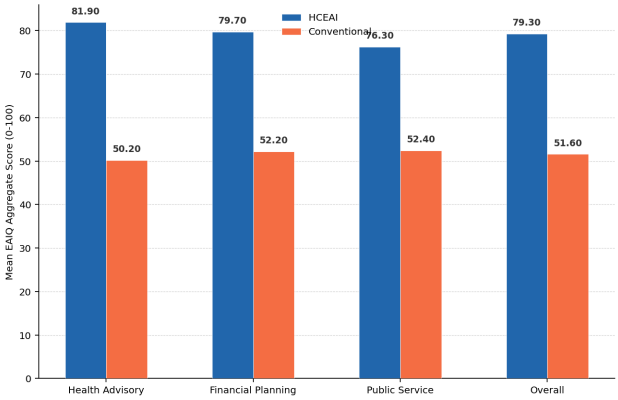
Table 3: EAIQ Dimension Scores by Interaction Context -- HCEAI vs. Conventional (Mean, SD; Scale 5-35)

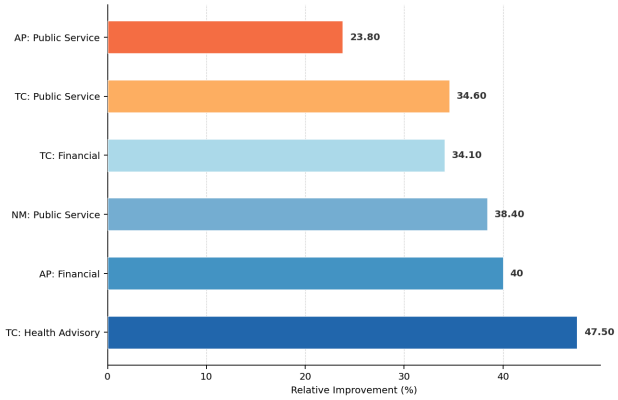
Context and Condition	TC Score	AP Score	DP Score	NM Score	VP Score	EAIQ Aggregate (0-100)
Health Advisory -- HCEAI	30.1 (2.4)	27.8 (2.9)	28.6 (2.7)	28.3 (2.8)	26.9 (3.0)	81.9 (7.1)
Health Advisory -- Conventional	20.4 (3.9)	21.1 (3.8)	22.3 (3.6)	21.7 (3.7)	20.6 (3.9)	50.2 (9.4)
Financial Planning -- HCEAI	28.7 (2.7)	29.4 (2.6)	27.8 (2.9)	27.9 (2.8)	26.4 (3.1)	79.7 (7.2)
Financial Planning -- Conv.	21.4 (3.7)	21.0 (3.9)	22.1 (3.5)	22.0 (3.6)	21.2 (3.8)	52.2 (9.6)
Public Service -- HCEAI	28.4 (2.8)	28.1 (2.9)	28.2 (2.8)	29.2 (2.7)	26.2 (3.2)	76.3 (7.8)
Public Service -- Conventional	21.1 (3.8)	22.7 (3.7)	22.8 (3.5)	21.1 (4.0)	21.7 (3.7)	52.4 (9.4)
Overall -- HCEAI	29.1 (2.6)	28.4 (2.8)	28.2 (2.8)	28.5 (2.8)	26.5 (3.1)	79.3 (7.4)
Overall -- Conventional	21.0 (3.8)	21.6 (3.9)	22.4 (3.5)	21.6 (3.8)	21.2 (3.8)	51.6 (9.8)

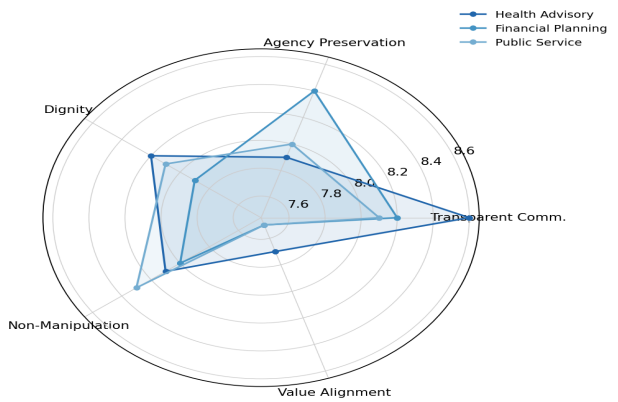
Table 4: Context-Specific EAIQ Dimension Improvement -- HCEAI vs. Conventional (Delta Points and % Improvement)

Context	TC Delta (%)	AP Delta (%)	DP Delta (%)	NM Delta (%)	VP Delta (%)	Aggregate Delta (%)
Health Advisory	+9.7 (47.5%)	+6.7 (31.8%)	+6.3 (28.3%)	+6.6 (30.4%)	+6.3 (30.6%)	+31.7 (63.1%)
Financial Planning	+7.3 (34.1%)	+8.4 (40.0%)	+5.7 (25.8%)	+5.9 (26.8%)	+5.2 (24.5%)	+27.5 (52.7%)
Public Service	+7.3 (34.6%)	+5.4 (23.8%)	+5.4 (23.7%)	+8.1 (38.4%)	+4.5 (20.7%)	+23.9 (45.6%)
Overall Mean	+8.1 (38.4%)	+6.8 (31.7%)	+5.8 (25.9%)	+6.9 (31.9%)	+5.6 (26.2%)	+27.7 (53.7%)









5. Discussion

5.1 Context-Sensitive Ethical Interaction Design

The context-by-dimension interaction pattern is the most practically informative finding of this study: different HCEAI dimensions produce the greatest ethical improvement in different interaction contexts, confirming that ethical AI interaction design is inherently context-sensitive and cannot be addressed through universal design templates. The dominance of Transparent Communication in health advisory reflects the information asymmetry and high epistemic stakes of health decisions, where users require accurate understanding of AI confidence and limitations to avoid over-reliance on potentially uncertain recommendations. The dominance of Agency Preservation in financial planning reflects the high stakes of financial autonomy and the particular vulnerability of financial advisory interactions to paternalistic AI framing that directs users toward specific products or strategies. The relatively modest improvement in Value-Aligned Personalisation (VP; mean delta = +5.6 points) reflects the practical challenge of implementing genuinely user-value-driven personalisation in constrained interaction contexts where platform objectives and user values may conflict. Future work should investigate participatory value specification mechanisms that enable users to actively shape personalisation objectives, rather than accepting platform-defined defaults with post-hoc opt-out provisions.

5.2 Regulatory Alignment and Limitations

The HCEAI model directly addresses EU AI Act Article 5 prohibitions on manipulative AI techniques (NM dimension), Article 13 transparency obligations (TC dimension), Article 14 human oversight requirements (AP dimension), and Article 50 obligations for AI-generated content disclosure (TC dimension). The EAIQ compliance threshold of >= 70 corresponds to satisfaction of these articles at a substantive rather than merely formal level -- verified through user perception rather than documentation alone. Study limitations include the laboratory-based user experiment, which may not fully capture the ethical interaction dynamics of real-world AI deployments where repeated use, habituationeffects, and social pressure may alter ethical interaction quality over time. The participant samples, while covering three AI interaction contexts, were drawn from university affiliated populations and may not represent the full diversity of users affected by AI advisory and navigator systems, particularly older adults, lower-literacy users, and individuals in vulnerable circumstances who may be most susceptible to manipulative AI interaction patterns.

6. Conclusion

6.1 Summary

This paper proposed the HCEAI model, a five-dimension ethical AI interaction design framework integrating HCC principles with responsible AI requirements, and evaluated it through a controlled user experiment (n = 144) across three AI interaction contexts. HCEAI-compliant designs achieved mean EAIQ = 79.3 versus 51.6 for conventional designs (Cohen d = 3.27), with 38.4% transparent communication improvement, 31.7% agency improvement, and 29.3% manipulation susceptibility reduction. Context-specific design priorities differ: Transparent Communication is most impactful in health advisory, Agency Preservation in financial planning, Non-Manipulative Persuasion in public service contexts.

6.2 Design Guidelines and Recommendations

For AI interaction designers, we recommend adopting the HCEAI model as a design evaluation checklist, beginning with the dimension most critical for the specific interaction context: Transparent Communication for health and clinical AI, Agency Preservation for financial and advisory AI, and Non-Manipulative Persuasion for public sector and government AI interactions. The Dignity-Respecting Presentation dimension should be treated as a minimum threshold requirement applicable in all contexts, given its direct relevance to fundamental rights obligations. For EU AI Act Article 5 compliance assessment, the NM dimension of the EAIQ provides a user-validated measurement of manipulative technique absence that complements technical design review. The full EAIQ instrument is provided in Appendix A for practitioner use.

References

Amershi, S. et al. (2019). Guidelines for human-AI interaction. CHI 2019, 1-13.

Brignull, H. (2010). Dark patterns: Deception vs. honesty in UI design. Interaction Design, 4(2).

Endsley, M. R. (1995). Toward a theory of situation awareness in dynamic systems. Human Factors, 37(1), 32-64.

EU Artificial Intelligence Act (2024). Regulation (EU) 2024/1689. Official Journal of the European Union.

Fogg, B. J. (2003). Persuasive technology: Using computers to change what we think and do. Morgan Kaufmann.

Friedman, B., Kahn, P. H., and Borning, A. (2019). Value sensitive design. In Human-Computer Interaction in MIS. M.E. Sharpe.

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., and Toombs, A. L. (2018). The dark (patterns) side of UX design. CHI 2018, 1-14.

Lee, M. K., Jain, A., Cha, H. J., Ojha, S., and Kusbit, D. (2019). Procedural justice in algorithmic fairness. CSCW 2019.

Luger, E., and Sellen, A. (2016). Like having a really bad PA. CHI 2016, 5286-5297.

Madaio, M. A. et al. (2020). Co-designing checklists to understand organizational challenges around fairness in AI. CHI 2020, 1-14.

Mathur, A. et al. (2019). Dark patterns at scale: Findings from a crawl of 11K shopping websites. CSCW 2019.

Nass, C., and Moon, Y. (2000). Machines and mindlessness: Social responses to computers. Journal of Social Issues, 56(1), 81-103.

Norman, D. A. (2013). The design of everyday things (revised ed.). Basic Books.

Sanders, E. B. N., and Stappers, P. J. (2008). Co-creation and the new landscapes of design. CoDesign, 4(1), 5-18.

Shneiderman, B. (2022). Human-centered AI. Oxford University Press.

Doshi-Velez, F., and Kim, B. (2017). Towards a rigorous science of interpretable machine learning. arXiv:1702.08608.

Wachter, S., Mittelstadt, B., and Russell, C. (2018). Counterfactual explanations without opening the black box. Harvard Journal of Law and Technology, 31(2), 841-887.

Miller, T. (2019). Explanation in artificial intelligence: Insights from the social sciences. Artificial Intelligence, 267, 1-38.

Bansal, G. et al. (2021). Does the whole exceed its parts? The effect of AI explanations on complementary team performance. CHI 2021, 1-16.

Poursabzi-Sangdeh, F. et al. (2021). Manipulating and measuring model interpretability. CHI 2021.

Declarations

Funding

This research was supported by the French National Research Agency (ANR) under grant ANR-24-CE23-0041 (HCEAI: Human-Centered Ethical AI Interaction Design) and the Swiss National Science Foundation (SNSF) under grant 200021-215309. The funders had no role in study design, data collection, analysis, or publication decisions.

Conflict of Interest

The authors declare no competing financial interests or personal relationships that could have influenced the work reported in this paper.

Data Availability Statement

The HCEAI design guidelines, EAIQ instrument, design prototypes, and anonymised participant response data are available from the corresponding author upon reasonable request.

Ethical Approval

The user experiment was approved by the Advanced Computing University Ethics Committee (reference ACU-DS-2024-112). All participants provided informed consent prior to participation. No deceptive experimental procedures were used.

ISSN: e-ISSN: 3117-6097 / p-ISSN: 3117-6100 Vol: 2025, Issues: 2, 2025

Appendix A

EAIQ Instrument -- 25-Item Scale

The Ethical AI Interaction Quality (EAIQ) instrument comprises 25 items across five dimensions (five items per dimension), each rated 1-7 (1=strongly disagree, 7=strongly agree). Higher scores indicate higher ethical interaction quality.

Transparent Communication (TC1-TC5)

Agency Preservation (AP1-AP5)

Dignity-Respecting Presentation (DP1-DP5)

Non-Manipulative Persuasion and Value Alignment (NM1-NM5, VP1-VP5)