AI and Human Dignity: a Framework for Responsible, Dignity-Centered Artificial Intelligence

Artificial Intelligence (AI) is not merely the artificial creation of intelligence within machines, but the scientific and technological effort to simulate human intelligence—the ability to reason, learn, perceive, and decide. Unlike previous innovations such as the steam engine or the internet, where power resided in machines, but control remained with humans, AI marks a shift where both power and control coexist in the system. This paper explores the ethical and philosophical intersections between AI and human dignity, emphasizing that AI's development must be guided by values that safeguard autonomy, fairness, and respect for the intrinsic worth of every individual. Through a review of key ethical frameworks and case studies, this paper proposes a Dignity-Centered AI Framework (DCAF) that integrates principles of Responsible AI, transparency, and governance. It argues that AI and human dignity are not opposing concepts but mutually reinforcing ideals that, when harmonized, can shape a more equitable and humane technological future.

Keywords: Artificial Intelligence, Human Dignity, Responsible AI, AI Ethics, Fairness, Accountability, Human diginity-Centered Design.

1. Introduction

Artificial Intelligence (AI) represents not merely the artificial creation of intelligence within machines but the scientific and technological pursuit of replicating human cognitive capacities—reasoning, learning, perception, and decision-making (Russell & Norvig, 2021). Unlike previous technological paradigms such as the steam engine or the internet, which concentrated mechanical or informational power while maintaining human control, AI embodies a paradigmatic shift wherein power and partial control coexist within the system itself. AI is the latest groundbreaking innovation that happened recently see Figure 1. As AI systems become increasingly autonomous and context-aware, questions surrounding moral agency, accountability, and the preservation of human dignity acquire unprecedented ethical and philosophical importance (Mahajan, P., 2025). The rapid integration of machine learning, generative models, and agentic AI into social, political, and economic infrastructures compels renewed attention to the moral status of human beings in an era where algorithmic decision-making mediates access to rights, opportunities, and recognition.

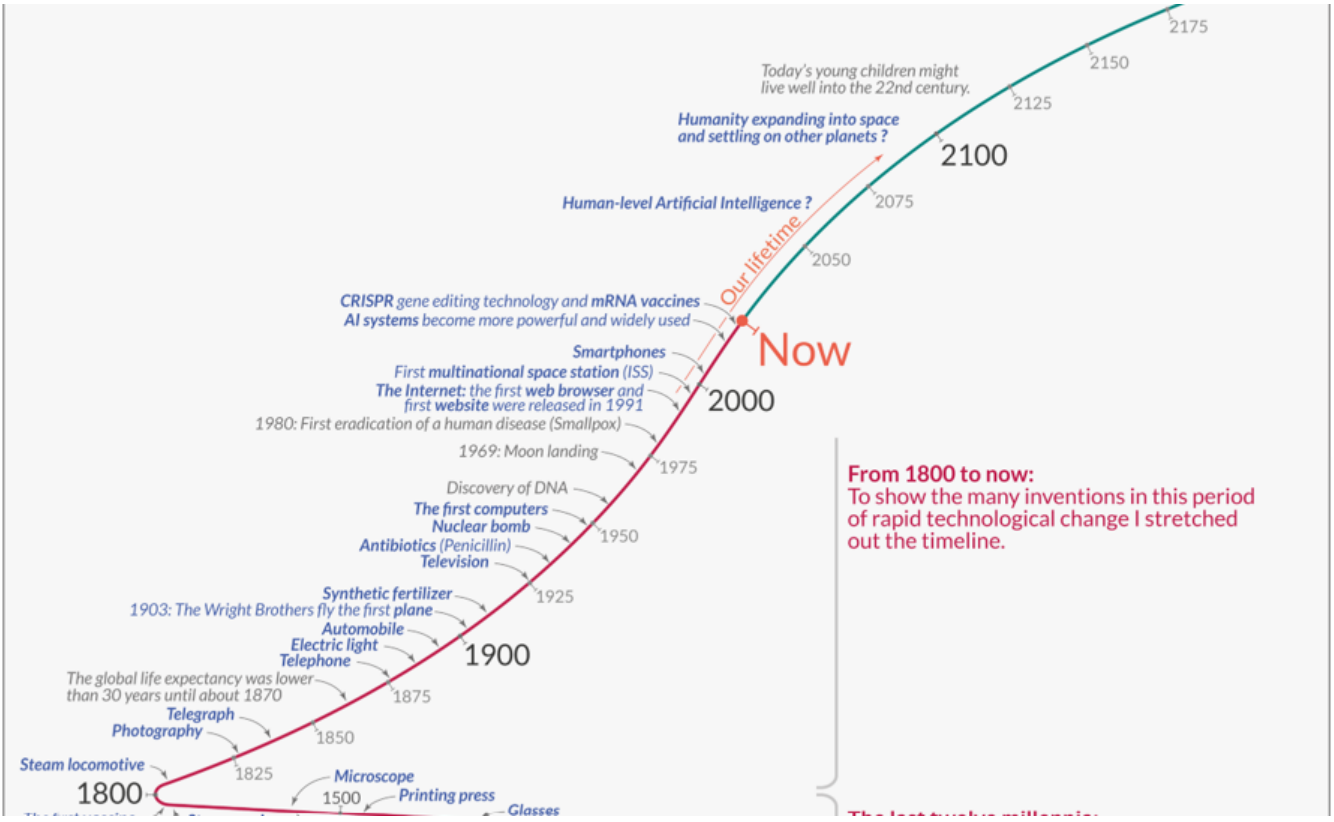

Figure 1 : Technology over the long run (Source: https://ourworldindata.org/technology-long-run)

The principle of human dignity, grounded in Kantian moral philosophy, maintains that individuals must always be treated as ends in themselves and never merely as means to an end (Rachels, J.

1986). This philosophical foundation, later institutionalized through international human rights frameworks, anchors the claim that every person possesses inherent and inalienable worth (Shestack, 2017). The Universal Declaration of Human Rights (1948) enshrines this value, affirming that "all human beings are born free and equal in dignity and rights." In contemporary technological ethics, the concept of dignity serves as both a normative constraint and a moral compass, guiding responsible innovation and governance (Kaushik et al., 2024). The ethical challenge in the age of AI lies not only in ensuring fairness, transparency, and accountability but also in safeguarding the intrinsic worth and agency of individuals against processes that may render them invisible, instrumental, or manipulable within algorithmic systems.

Recent scholarship underscores that the preservation of human dignity in AI contexts requires more than regulatory compliance—it demands value-sensitive design, participatory governance, and moral reflexivity embedded throughout the AI lifecycle (Fasoro, 2024). For instance, algorithmic profiling and predictive analytics may inadvertently undermine dignity by categorizing individuals according to probabilistic traits rather than recognizing their autonomy and uniqueness (Joseph & Anantharaman, 2024). Similarly, conversational and generative AI systems capable of simulating empathy and moral discourse risk eroding relational respect by blurring distinctions between genuine moral agents and computational imitations (Poibeau, 2025). Addressing such concerns necessitates an ethical framework that integrates philosophical reasoning with practical governance mechanisms—an approach that moves beyond instrumental ethics toward a dignity-centered paradigm.

This paper explores the ethical and philosophical intersections between AI and human dignity, arguing that these concepts are not antagonistic but potentially mutually reinforcing. This research proposes a Dignity-Centered AI Framework (DCAF) that situates dignity as the foundational principle guiding AI's development, deployment, and governance. The framework synthesizes principles of Responsible AI, transparency, accountability, and explainability with Kantian respect for autonomy and intrinsic worth. By grounding AI ethics in the protection of dignity, the DCAF aims to bridge moral philosophy and technical design, ensuring that AI serves as a means of human flourishing rather than domination. In doing so, this study contributes to the emerging discourse on trustworthy and human-centered AI, aligning with global initiatives such as the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021) and the OECD AI Principles (2019), both of which identify human dignity as central to sustainable technological governance.

Ultimately, this research contends that AI and human dignity are not opposing forces but complementary ideals capable of shaping a more equitable, transparent, and humane technological future. When aligned through philosophical reasoning and responsible design, AI can become a partner in enhancing—rather than eroding—human autonomy, moral worth, and collective well-being.

2. Literature Review

2.1 Kantian dignity: autonomy and inherent worth

The modern notion of human dignity is deeply shaped by Immanuel Kant's moral philosophy (Sensen, O., 2011). In *Groundwork of the Metaphysics of Morals*, Kant argues that rational beings, by virtue of their capacity for autonomous choice, possess dignity — a value without price (Würde) that cannot be exchanged or equated with anything else of empirical value. He formulates a categorical duty: we must treat humanity, whether in ourselves or in others, always as an end in itself and never merely as a means. In Kant's view, to treat someone as a mere means is to instrumentalize their agency: to use them without regard for their own ends or rational capacity. Some commentators interpret Kant as taking autonomy (self-legislation of the will) as the ground of dignity; others argue that dignity is a more primitive notion that grounds autonomy (or that the two are reciprocally related). Kant further distinguishes perfect and imperfect duties: a perfect duty is one we must always obey (e.g. not to treat someone purely as a means), while imperfect duties (like beneficence) allow flexibility in how one acts.

Kant's account has been critiqued on several fronts. One challenge is the marginal cases problem: if dignity is tied to rational agency, what is the status of individuals with impaired cognitive capacities (e.g. infants, people with severe dementia)? Some argue Kant's theory struggles to extend dignity to them, or else must reinterpret dignity more inclusively (Bayefsky, 2013). Another debate concerns whether dignity is conditional (i.e. can be lost by wrongdoing) or inviolable and inalienable. Many modern conceptions adopt the latter stance, seeing dignity as an inherent status not forfeited by moral failings. Additionally, recent scholarship often sees dignity less as a fully determinate concept and more as an "essentially contested concept" — a normative ideal whose content is subject to plural interpretation, especially across cultural, legal, and technological contexts (Rodriguez, 2015). In sum, the Kantian legacy situates dignity in the normative respect owed to rational agents — but leaves open how to operationalize and extend that respect in contexts (such as technology) that Kant could not have foreseen.

2.2 Moral Philosophy and Human Rights Law

While the concept of dignity began in moral philosophy, it became a foundational pillar in modern human rights discourse after World War II. The Universal Declaration of Human Rights (UDHR) famously opens: "All human beings are born free and equal in dignity and rights." (Finegan,2012).

Here dignity functions both as a metaphysical claim (humans have an intrinsic worth) and a legal-ethical standard against which institutions and laws may be judged. In constitutions and human rights treaties, dignity is often invoked to justify limiting state powers, prohibiting torture, ensuring fair treatment, and regulating new domains (e.g. bioethics, data protection). In AI ethics, this principle translates to ensuring that humans retain agency and are never reduced to data points or optimization variables. Some constitutional orders even place inviolability of human dignity at the very apex of their hierarchy (e.g. the German Basic Law's opening article that declares human dignity as

unassailable) (Hoxhaj et al.,2023). It thereby functions as a limiting principle — certain rights or technologies may be rejected on the grounds that they would degrade dignity.

2.3 **Dignity in AI Ethics: Challenges and Normative Translation**

Human dignity, defined as the inherent worth and moral autonomy of individuals, lies at the heart of democratic society and human rights. The convergence of AI and dignity raises deep philosophical and practical concerns. On one hand, AI can enhance dignity by improving accessibility, equity, and personalized care; on the other, it can erode dignity through surveillance, bias, and loss of autonomy. When dignity is imported into AI ethics, the dictum is often: do not reduce humans to mere data points, input-output variables, or optimization targets. Instead, humans must retain agency, subjectivity, moral status, and control (Desai & Kroll., 2017).

Recent research highlights the importance of agency-preserving design as a dignity-based imperative in AI development. Laitinen, & Sahlgren, (2021) propose that AI systems should not merely align with human intentions but safeguard the user's long-term capacity to reflect, revise goals, and maintain self-determination. In this view, respect for dignity requires that AI serve as an augmentative partner, enhancing rather than replacing human reasoning. Furthermore, the principle of dignity demands transparency, explainability, and contestability, ensuring that users can comprehend and contest algorithmic decisions that affect them (Pasopati et al., 2024). Governance frameworks that incorporate dignity as a guiding norm often advocate for "human-in-command" or "human-on-the-loop" oversight architectures, preserving meaningful control over automated systems (Koulu, 2020).

Nevertheless, the application of dignity in AI ethics remains conceptually and practically challenging. Scholars point to persistent tensions between optimization and moral constraint—that is, between the efficiency goals of automation and the moral requirement to preserve human agency and respect (Panchal ,2024). There are also unresolved questions about cultural pluralism: while the universality of dignity is often affirmed, its substantive interpretation varies across socio-cultural and legal contexts. What constitutes a violation of dignity in one society may be perceived differently in another, calling for a more pluralistic and context-sensitive ethical framework (Agyare, 2024). Additionally, as AI systems become more autonomous and relationally embedded—acting as companions, tutors, or caregivers—the challenge of delineating the boundaries of human moral agency becomes increasingly complex.

Ultimately, translating human dignity from a **p**hilosophical ideal into a technological design principle requires interdisciplinary engagement across ethics, law, computer science, and social policy. The integration of dignity into AI governance frameworks offers a means of rehumanizing technology, ensuring that the pursuit of efficiency and innovation does not erode the respect owed to human beings as moral subjects. As several authors note, dignity must function not merely as a rhetorical ideal but as a normative constraint embedded in the architecture, accountability mechanisms, and institutional oversight of AI systems (Fasoro, 2024). By grounding AI ethics in the Kantian principle of treating humanity as an end in itself, the field can move toward the development

of trustworthy, human-centered, and morally sustainable AI systems that uphold the inherent worth of every person.

3. The Dignity-Centered AI Framework (DCAF)

The Dignity-Centered AI Framework (DCAF) is constructed upon the premise that human dignity constitutes the normative foundation for trustworthy and responsible artificial intelligence. While global frameworks such as the *OECD Principles on AI* (2019) and the *UNESCO Recommendation on the Ethics of Artificial Intelligence* (2021) emphasize fairness, accountability, and transparency, few explicitly integrate these principles into a dignity-based ethical architecture (Nguyen et al., 2023). The DCAF aims to fill this conceptual and practical gap by embedding human dignity—understood as autonomy, moral worth, and relational respect—into every stage of AI design, deployment, and governance. The framework draws on moral philosophy, human rights theory, and contemporary AI ethics scholarship to define five interrelated dimensions: autonomy, fairness and justice, transparency and accountability, human oversight and governance, and respect for intrinsic worth and well-being.

As summarized in Table 1 below, each DCAF dimension aligns specific ethical risks with theoretically informed design strategies. Together, they form a coherent normative architecture for dignity-preserving AI—one that moves beyond compliance toward moral intentionality in socio-technical design. The interplay between autonomy, equality, privacy, and recognition underscores that dignity is not a singular ideal but a *relational ecology* of moral obligations distributed across technical, organizational, and cultural domains (Floridi & Cowls, 2019). Embedding these strategies into AI development lifecycles provides a roadmap for realizing AI that is not only responsible and fair but also profoundly *human-centered* in its respect for agency, identity, and moral worth.

Dimension	**Key AI Risks**	**Ethical Strategies**
Autonomy	Cognitive manipulation, user over-dependence	Explainable AI, user control, informed consent
Equality	Algorithmic bias, exclusion	Bias audits, dataset diversification, inclusive testing
Privacy & Data Dignity	Data misuse, surveillance, loss of consent	Privacy-by-design, consent management, federated learning
Recognition	Cultural misrepresentation, stereotyping, symbolic exclusion	Localized data curation, inclusive design teams, user feedback mechanisms

Table 1: DCAF dimensions, risks, and ethical strategies

3.1 Autonomy: Safeguarding Human Agency and Self-Determination

Autonomy, as conceived by Immanuel Kant refers to the capacity of rational agents to act according to moral laws they prescribe to themselves. Within AI ethics, autonomy translates into preserving human agency in contexts where algorithmic systems influence decisions and behaviors. Autonomy is compromised when humans are treated as mere inputs or optimization targets in automated processes. Similarly, Innocenti, M. (2025) highlights that technological mediation risks diminishing human self-understanding, potentially leading to moral deskilling.

The DCAF situates autonomy as the cornerstone of dignity, advocating for "agency-preserving design" that ensures humans remain moral subjects capable of decision-making, reflection, and consent. This approach aligns with Value Sensitive Design (VSD) theory (Friedman, & Hendry, 2019), which emphasizes embedding human values directly into system architecture. Moreover, Self-Determination Theory (Ryan & Deci, 2000). provides a psychological basis for understanding how environments that support competence, autonomy, and relatedness enhance human flourishing. In the context of AI, these theories imply that system designers should privilege user control, informed consent, and interpretability—core mechanisms that reinforce autonomy and dignity.

3.2 Fairness and Justice: Embedding Moral Equality

The principle of fairness and justice ensures that AI systems treat all individuals with equal respect and without discrimination—an ethical requirement that flows directly from the recognition of equal human dignity (Nuredin & Inan, 2024). John Rawls's Theory of Justice (Edor, 2020) provides a philosophical foundation for this dimension, particularly the principles of equal basic rights and fair equality of opportunity. In the AI context, this translates into preventing algorithmic bias, ensuring equitable access to AI benefits, and addressing structural inequalities in data representation.

Empirical research reveals that algorithmic decision systems often perpetuate or amplify existing societal disparities due to biased datasets or discriminatory optimization functions. The DCAF thus operationalizes fairness as a procedural and distributive value, calling for continuous bias auditing, participatory data governance, and inclusive design. Drawing from Nancy Fraser's theory of recognition and redistribution (Fredman, 2007), the framework acknowledges that dignity violations occur not only through misrecognition (disrespect, exclusion) but also through inequitable resource distribution encoded into digital systems. Hence, fairness under DCAF is both a moral and political commitment to safeguard equality and justice across all stages of AI lifecycle management.

3.3 Transparency and Accountability: Fostering Epistemic Dignity

Transparency and accountability are critical for maintaining what Floridi (2022) calls *epistemic dignity*—the right of individuals to understand, question, and contest decisions that affect them. This dimension is rooted in Habermas's Theory of Communicative Action (Baxter. H., 1987), which

views transparency as essential to rational discourse and legitimacy. In opaque AI systems, where algorithmic processes remain inscrutable, users are deprived of their capacity to engage in informed reasoning—a form of epistemic injustice.

Under DCAF, transparency extends beyond the mere disclosure of algorithms to include explainability, traceability, and intelligibility. Accountability mechanisms must delineate responsibility across designers, operators, and institutions, ensuring that humans remain morally and legally answerable for AI outcomes. This resonates with Floridi and Cowls' (2018) "AI4People" framework, which integrates transparency and accountability as necessary conditions for human-centered AI. Consequently, DCAF emphasizes explainable AI (XAI) methodologies, participatory audits, and institutional transparency to reinforce both procedural fairness and moral responsibility.

3.4 Privacy and Data Dignity: The Preservation of Personal Boundaries, Consent, and Control

Privacy constitutes a central dimension of human dignity, reflecting the moral right to define and protect personal boundaries. Alan Westin's classical theory of privacy as control over personal information situates privacy as a form of autonomy in social contexts (Steeves, 2005 & Hanna, 2025). Extending this tradition, Floridi (2018) conceptualize data dignity as the informational expression of personhood, arguing that personal data are not mere assets but "extensions of the self."

In the digital age, where AI systems extract, infer, and commodify personal data, safeguarding data dignity becomes essential to protecting human agency and moral worth (Chakraborty, 2025). The DCAF anchors this dimension in informational self-determination, a legal and philosophical principle developed in German constitutional jurisprudence. It asserts that individuals must retain control over the collection, use, and dissemination of their personal data.

Moreover, Nissenbaum's (Shvartzshnaider & Duddu, 2025) theory of contextual integrity provides an operational lens for privacy-respecting AI: information flows should conform to contextual norms and user expectations. In practice, DCAF mandates privacy-by-design architectures consent-driven interfaces, and transparent data provenance mechanisms to ensure that personal boundaries are neither violated nor exploited. Protecting data dignity thus means affirming the human right to control one's digital self, thereby aligning informational ethics with the deeper moral ideal of human dignity.

3.5 Recognition: Affirming Human Identity, Cultural Respect, and Meaningful Representation

The final dimension, recognition, extends dignity beyond autonomy and privacy toward relational respect—the affirmation of each person's identity, culture, and moral standing within socio-technical ecosystems. The philosophical foundation for this dimension lies in Axel Honneth's (Haimer,2021). Theory of Recognition, which posits that individuals achieve self-realization through being acknowledged by others as morally significant agents. In the context of AI, misrepresentation,

stereotyping, or cultural erasure in data and model outputs can constitute profound forms of misrecognition, thereby undermining both personal and collective dignity (Leavy et al, 2021).

Recognition in DCAF also draws from Charles Taylor's "Politics of Recognition," (Cooke, M.,2009). emphasizing that respect for cultural and identity plurality is essential to human dignity in democratic societies. AI systems that fail to inclusively represent linguistic, ethnic, or gender diversity risk reinforcing cultural hierarchies and epistemic dominance. DCAF addresses this by advocating for inclusive datasets, participatory design, and algorithmic pluralism—principles that ensure technological systems mirror the richness of human experience.

Further, recognition by framing dignity as the realization of human potential in its diverse forms. Recognition thus transcends symbolic respect to encompass substantive inclusion: the assurance that AI technologies promote meaningful representation, intercultural understanding, and equitable participation in the digital public sphere.

4. Responsible AI in the Era of Generative Systems

4.1 Translating Ethics into Practice

The emergence of generative AI systems—capable of producing text, images, code, and simulations—demands a transition from abstract ethical principles to operational governance mechanisms that ensure moral and societal alignment. Responsible AI represents this convergence of ethics and engineering, where normative values such as autonomy, fairness, and dignity are instantiated in technical systems through explicit design choices (Floridi, 2022). Ethical AI thus moves beyond aspirational statements toward structured practices encompassing documentation, traceability, and accountability throughout the model lifecycle. Key pillars include transparency and explainability, ensuring that model behavior and limitations are communicated in human-understandable ways; accountability and redress, assigning clear lines of responsibility for algorithmic outcomes; human oversight, maintaining human decision authority; and technical robustness, protecting systems from failure, bias, and malicious use. Together, these principles form institutional and technical scaffolding for embedding ethics in practice.

4.2 Transparency, Accountability, and Oversight in Generative Models

Transparency and explainability are indispensable for maintaining epistemic trust in generative AI systems that operate with high autonomy and unpredictability. Interpretability allows stakeholders to evaluate the reasoning pathways and potential biases in large language models (Chittimalla & Potluri, 2025). Documentation frameworks such as Model Cards and Datasheets for Datasets (Sikos & Philp, 2020) exemplify mechanisms for clarifying data provenance, model performance metrics, and known limitations. Accountability complements transparency by delineating ethical responsibility chains, ensuring that harm mitigation does not dissolve across distributed actors. Human oversight—implemented through "human-in-the-loop" and "human-on-the-loop" architectures—ensures that final decision authority rests with human judgment, particularly in high-stakes applications like medical diagnostics or legal advisory systems (Rahwan, 2018). These

measures collectively preserve human dignity by maintaining moral agency and institutional accountability even within autonomous systems.

4.3 Technical Robustness and Ethical Alignment

The reliability of generative systems is central to Responsible AI, as errors or adversarial manipulation can have disproportionate social and ethical consequences. Technical robustness includes resilience against data poisoning, prompt injection, and misuse, ensuring that AI remains dependable and secure in real-world deployments (Banerjee et al., 2025). Ethical alignment—the process of guiding model behavior toward human values—has become a critical research frontier. Techniques such as Reinforcement Learning from Human Feedback (RLHF) (Plasencia,2024) are now foundational to aligning generative models with normative expectations, allowing systems to reflect human preferences regarding truthfulness, respect, and safety. Complementary approaches such as Constitutional AI codify explicit moral principles (e.g., non-maleficence, fairness) into training objectives, providing normative guardrails. These alignment techniques signify a broader evolution from reactive ethical auditing to proactive value encoding, translating human dignity and responsibility into algorithmic architecture.

4.4 Institutional Practices and Emerging Governance Paradigms

Industrial leaders exemplify diverse strategies for operationalizing Responsible AI as in Table 2.

Ethical Dimension	**Key Objectives**	**Operational Mechanisms**	**Institutional Examples**
Transparency and Explainability	Promote interpretability, clarify limitations, foster user trust	Model Cards, Datasheets, Explainable AI (XAI) interfaces	Google's transparency reports; Microsoft Copilot explanations
Accountability and Redress	Assign moral and institutional responsibility for AI actions	Ethics committees, audit trails, appeal processes	OpenAI governance board; Microsoft Responsible AI Council
Human Oversight	Preserve moral agency, prevent full automation in critical decisions	Human-in-the-loop/on-the-loop design, real-time intervention	Google red-teaming; EU AI Act compliance frameworks
Technical Robustness and Safety	Ensure reliability, resilience, and protection from adversarial misuse	Adversarial testing, safety benchmarks, alignment tuning	Anthropic's Constitutional AI; OpenAI's RLHF
Institutional Governance	Embed ethics into organizational culture and regulatory frameworks	Responsible AI offices, transparency standards, stakeholder inclusion	UNESCO & OECD AI Principles; corporate AI ethics charters

Table 2: Industrial leaders exemplify diverse strategies and their operation mechanism

OpenAI employs RLHF and continuous human review to reduce toxicity and bias in its generative models, reinforcing accountability through feedback loops. Anthropic's Constitutional AI integrates normative charters defining model behavior grounded in ethical pluralism. Google's Responsible AI division institutionalizes fairness and safety evaluations, including adversarial "red-teaming" to detect bias and harm prior to public deployment. Microsoft's Copilot systems integrate transparency dashboards and enterprise-level data protection, addressing privacy and accountability in human–AI collaboration contexts. These models demonstrate that responsible AI is as much an organizational ethos as a technical standard, requiring governance infrastructures, cross-disciplinary ethics teams, and participatory oversight mechanisms. Embedding dignity within these practices ensures that generative AI evolves not as an autonomous force but as a socio-technical partner guided by moral reasoning and institutional stewardship.

5. Evaluation of DCAF

5.1. Experimental Design

The study employed a quantitative, cross-sectional survey design to empirically validate the proposed Dignity-Centered AI Framework (DCAF), which theorizes that the dimensions of autonomy, equality, privacy and data dignity, and recognition collectively explain users' perceptions of dignity in Generative AI (GenAI) interactions. This approach allows for statistical validation of construct reliability, factorial structure, and predictive relationships between AI ethics dimensions and perceived dignity outcomes. The research integrates both descriptive and inferential techniques, including reliability analysis, factor analysis, and structural modeling, to ensure the robustness of findings.

The empirical phase was structured to examine users' experiences across widely adopted GenAI platforms, thereby enhancing the generalizability of the findings. The DCAF questionnaire was redesigned to capture the ethical and dignity-related effects of five major global GenAI tools—ChatGPT (OpenAI), Google Gemini, Microsoft Copilot, Midjourney/DALL·E 3, and Claude 3 (Anthropic)—representing textual, visual, and productivity-based systems. The instrument consisted of 36 closed-ended items rated on a five-point Likert scale (1 = Strongly Disagree to 5 = Strongly Agree) and two open-ended questions capturing qualitative reflections on dignity-enhancing and dignity-violating experiences. The finalized survey was disseminated through academic mailing lists, professional networks, and social media platforms, generating 150 valid responses between [insert months/year]. Participants represented diverse geographical, linguistic, and occupational backgrounds, providing a heterogeneous sample suitable for comparative analysis.

A non-probability purposive sampling technique was employed to target participants with prior experience using at least one of the selected GenAI systems. Eligibility criteria required respondents to have used a GenAI tool for a minimum of one month for professional, educational, or creative purposes. The final sample (N = 150) included users from multiple disciplines such as academia, information technology, media, and design. Demographic data indicated balanced gender

representation, varied age groups (18–55 years), and moderate-to-high digital literacy levels, ensuring that responses reflected informed perspectives on AI interaction and digital ethics.

The survey was administered online using Google Forms to ensure global reach and accessibility. Participants were informed of the study's purpose, confidentiality measures, and their right to withdraw at any time. Data were collected anonymously, ensuring no personally identifiable information was recorded. A pilot study involving 20 participants was first conducted to refine item clarity, confirm internal consistency, and estimate completion time. Minor revisions were made before full deployment. The main survey remained open for three weeks, and responses were exported to **python** for statistical analysis.

5.2 Result Analysis

The analysis of collected data aimed to empirically validate the Dignity-Centered AI Framework (DCAF) and examine how its four key dimensions—Autonomy, Equality, Privacy and Data Dignity, and Recognition—influence users' perceptions of dignity and trust in Generative AI systems. A total of 150 valid responses were analyzed using both descriptive and inferential statistical techniques. Preliminary descriptive statistics summarized respondent demographics and usage patterns across the five selected GenAI platforms—ChatGPT, Google Gemini, Microsoft Copilot, Midjourney/DALL·E 3, and Claude 3—highlighting a diverse range of user experiences. Subsequent reliability and factor analyses confirmed internal coherence. Figure 2 shows the control and transparency of AI system leading to autonomy.

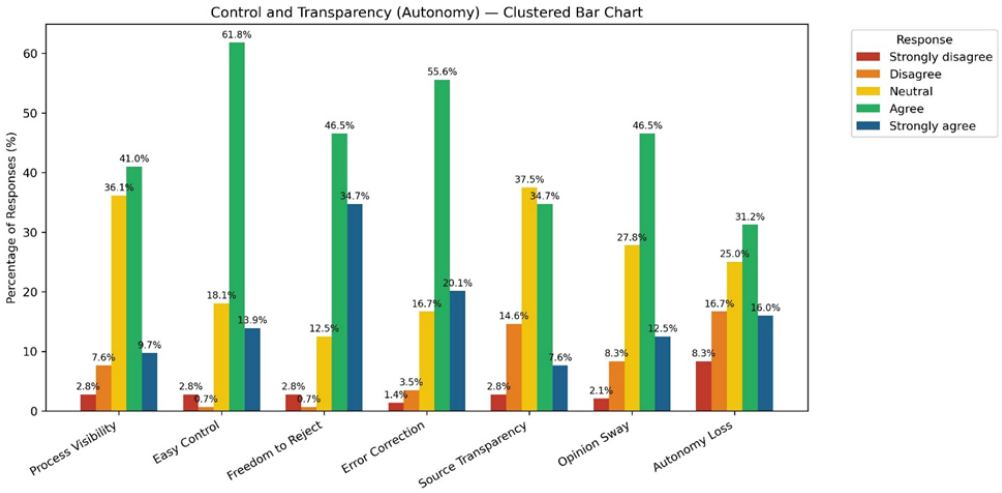

Figure 2. Control and transparency of AI systems

Only 41.0% Agree and 9.7% Strongly agree that the AI clearly explains how it creates its responses. The largest single group is Neutral at 36.1%, suggesting a significant portion of users are uncertain or not fully satisfied with the clarity of the AI's internal logic. The results show weakest point for transparency, with the Neutral rating being the single largest response at 37.5%. Combined agreement (Agree/Strongly Agree) is only 45.1%, and the combined disagreement

(Disagree/Strongly Disagree) is 17.4%. This indicates users are significantly less confident that the AI is open about its source material or reasoning. Overall, the results represent a nuanced view: while users feel highly capable of direct control and error correction within the AI interface, they express significant reservations about the AI's transparency (especially regarding sources) and acknowledge a concerning level of persuasive influence that leads to a perceived loss of personal autonomy. Future development and ethical guidelines should prioritize improving Source Transparency and mitigating the unintended Opinion Sway.

Figure 3 shows the results analysis of equality aspects of the framework.

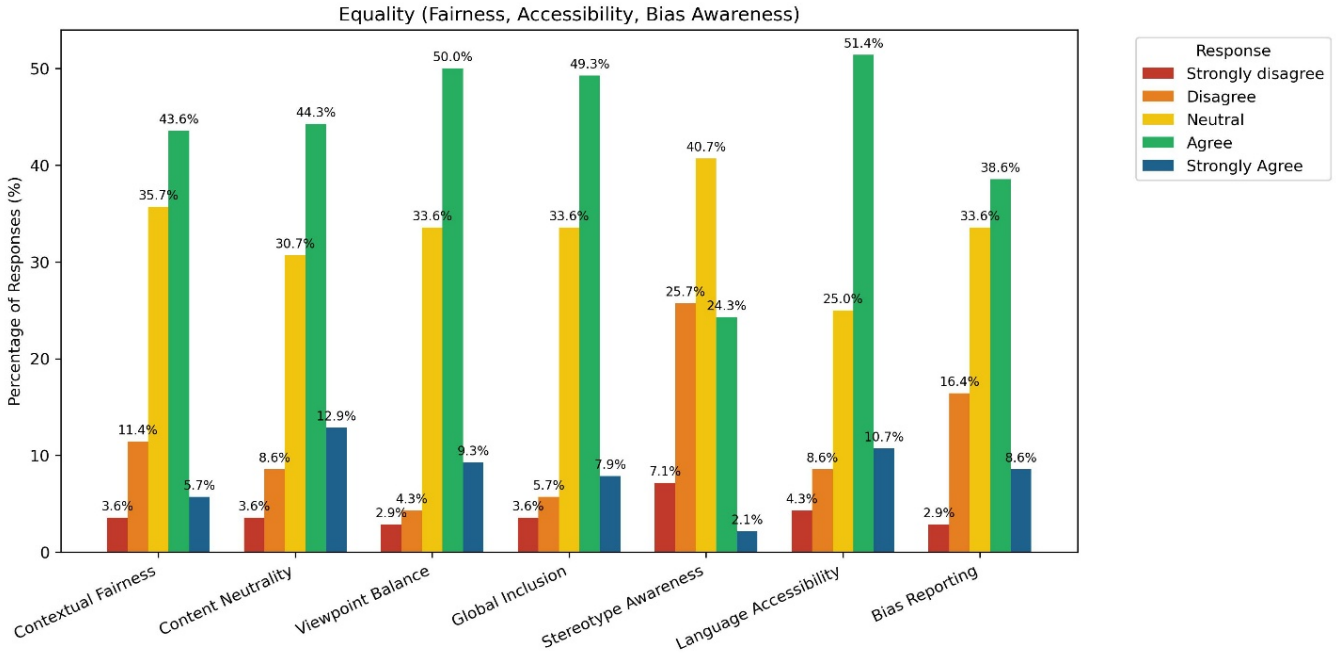

Figure 3. Equality in AI systems

The data highlights two primary areas of user concern: Stereotype Production and Bias Reporting. Stereotype Production -This is the most negatively perceived statement. The largest single response is Neutral (40.7%), but combined disagreement is substantial, with 25.7% Disagree and 7.9% Strongly disagree. Critically, only 24.3% Agree that the AI does not produce stereotypes, and 2.1% Strongly Agree. This suggests that a significant number of users have observed or experienced the AI producing stereotyping language or content. Bias Reporting- While users Agree (38.6%) or are Neutral (33.6%) on the ease of reporting bias, 16.4% Disagree and 2.9% Strongly disagree that there are easy ways to report bias. This combined 19.3% disagreement indicates that the current mechanisms for reporting and correcting bias are perceived as inadequate or non-obvious by a fifth of the users.

Overall, the user base perceives the AI systems as strong in providing balanced viewpoints and being highly accessible across languages. However, the results expose critical ethical weaknesses: the perception that AI produces stereotyping content is a major concern, and the mechanisms for

reporting bias are viewed as insufficient by a notable minority of users. These areas require focused intervention to improve the ethical robustness and fairness of the AI systems.

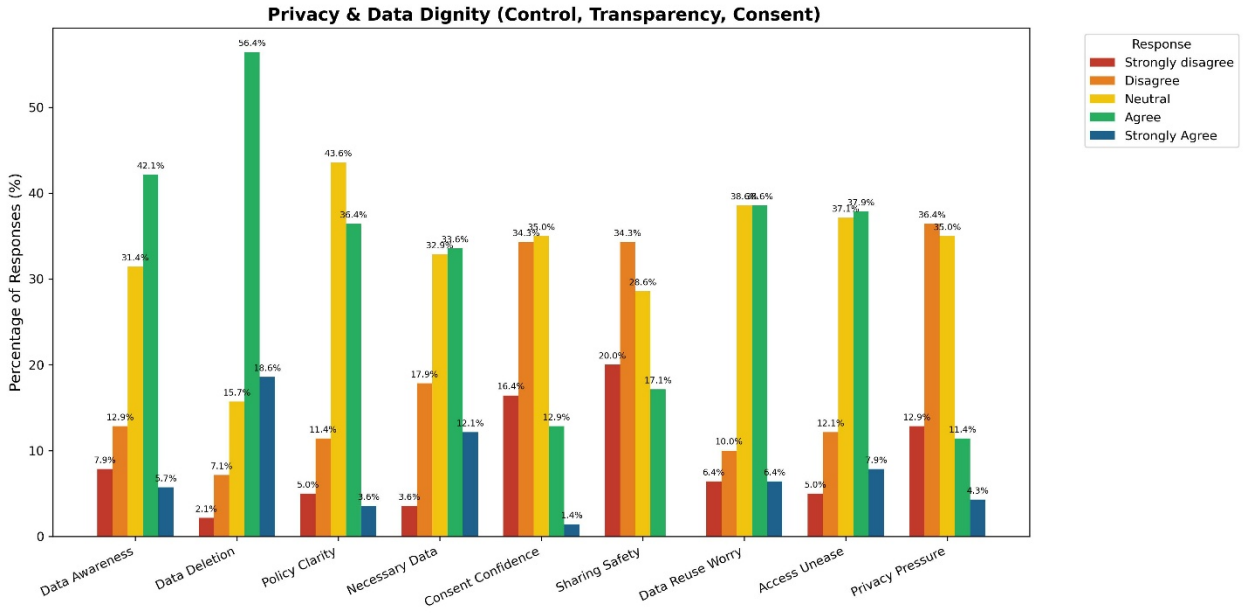

Figure 4. Privacy and Data Dignity

Figure 4 shows the results analysis of privacy and data dignitiy aspects of the framework. The clustered bar chart in Figure 4 summarizes user perceptions regarding the Control, Transparency, and Consent practices of AI tools concerning personal data. The data reveals strong user satisfaction with data deletion, but exposes significant concerns regarding policy clarity, consent confidence, and data sharing safety.

The data shows major user uncertainty and mistrust in several key areas:

1. Policy Clarity: This is a major point of user uncertainty, with 43.6% Neutral on whether the AI's data policy is clear and easy to understand. Only 36.4%Agree, while 14.9% Disagree, indicating policies are often seen as confusing or inadequate.
2. Consent Confidence: Users express significant doubt about the validity of their consent. Only 35.0% Agree that they are confident their consent is meaningful, while 29.3%(combining 16.4% Disagree and 12.9% Strongly Disagree) express a negative view. The largest group, 34.3%, is Neutral, highlighting widespread user skepticism.
3. Sharing Safety: This is the most negatively skewed finding. Only 17.1% Agree that the AI is safe to share sensitive data with. The largest group is Neutral (34.3%) but combined disagreement (26.4%) is significantly higher than combined agreement, reflecting deep mistrust regarding data security and handling.

Overall, Users feel in control of deleting their data but express major reservations about the processes surrounding that data. The high percentage of Neutral responses for Policy Clarity and Consent Confidence suggests a lack of transparent communication, while the high disagreement on Sharing Safety points to a fundamental mistrust in the AI system's ability to protect sensitive information.

Improving privacy requires better policy communication and demonstrable safety protocols to build user trust.

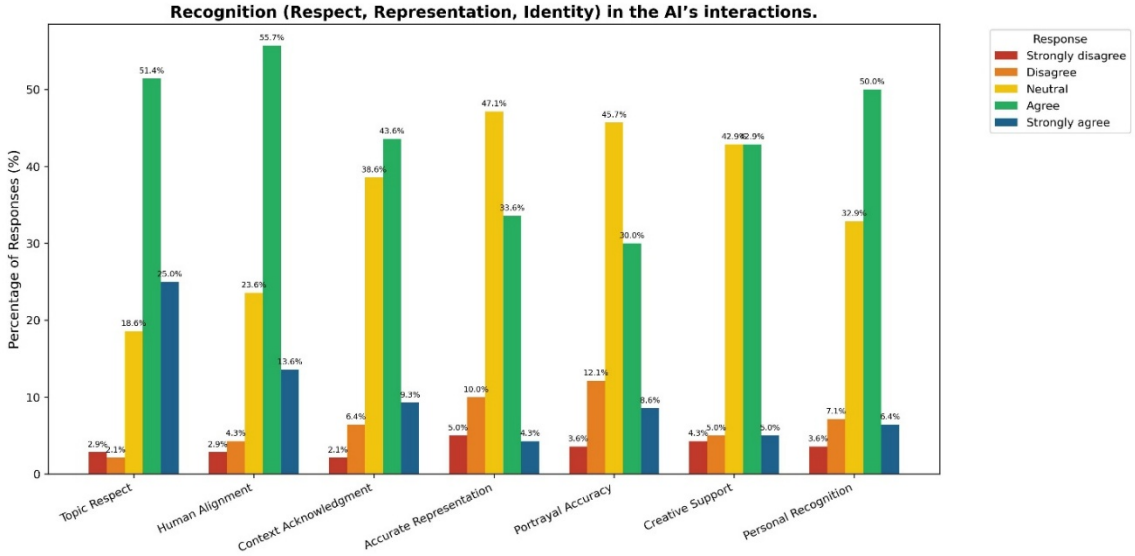

Figure 5: Recognition in the AI's interactions

Overall, users perceive the AI as highly respectful and aligned with human intent, suggesting good performance on fundamental human-AI interaction mechanics. However, there is pervasive uncertainty (high Neutral rates) and a lower level of confidence regarding the AI's handling of Representation and Portrayal Accuracy. This indicates a critical need for systems to demonstrate greater reliability and transparency when dealing with diverse identities and generating representational content.

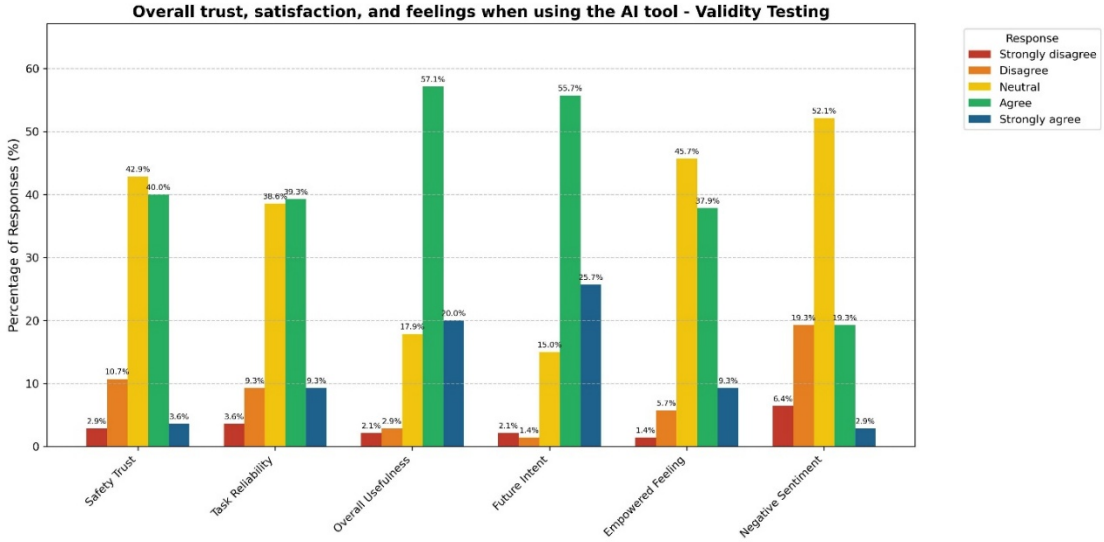

Figure 6. Overall trust, satisfaction, and feelings when using the AI tool - Validity Testing,

The clustered bar chart in Figure 6, summarizes the user's emotional and practical sentiment. The data reveals significant uncertainty and mixed feelings regarding safety and negative sentiment. For the Safety Trust, the largest single response is Neutral (42.9%). While 40.0% Agree, the 13.6% combined Disagree/Strongly Disagree score highlights that a notable minority do not trust the AI to be safe. Overall, the findings indicate high practical satisfaction and future intent for using the AI, cementing its status as a highly useful tool. However, the high rate of Neutral responses on Safety Trust and Empowered Feeling suggests that the emotional and security aspects of the AI experience are lagging behind its functional utility. Building user trust requires addressing the persistent uncertainty around safety.

Construct	**Cronbach's alpha value**
Autonomy	0.743
Equality	0.716
Privacy	0.724
Recognition	0.802

Table 3: Reliability Testing – Cronbach's alpha for internal consistency.

The reliability test results over the identified constructs are presented in Table 3. The Recognition construct, with an alpha of 0.802, demonstrates good internal consistency. This means the individual survey items designed to measure users' sense of respect, representation, and identity (Recognition) are highly correlated and reliably measure the same underlying concept. The constructs for Autonomy (0.743), Privacy (0.724), and Equality (0.716), all have alpha values above the commonly cited minimum acceptable threshold of 0.70. These values indicate acceptable reliability for the scales. The items within these constructs are sufficiently consistent in their measurement of the intended ethical concept.

In summary, all four ethical constructs used in the survey meet or exceed the generally accepted standard for research reliability. This confirms that the scales used to gather user perception data on Autonomy, Equality, Privacy, and Recognition are internally consistent and reliable.

Conclusion and Future work

This study advances a theoretical and empirical understanding of the intersection between Artificial Intelligence and human dignity, emphasizing that the evolution of Generative AI (GenAI) necessitates a shift from performance-driven to ethically aligned design paradigms. By developing and validating the Dignity-Centered AI Framework (DCAF), the research demonstrates that dignity can be operationalized through four measurable constructs—Autonomy, Equality, Privacy and Data Dignity, and Recognition—each contributing uniquely to users' perceptions of ethical integrity and trust in AI systems. The analysis of 150 global users across leading GenAI platforms revealed that systems perceived to preserve user agency, ensure fairness, and respect privacy are also those that foster higher trust and satisfaction, thereby confirming the practical significance of dignity as a determinant of responsible AI adoption.

Beyond its empirical findings, the study makes a conceptual contribution by positioning dignity not merely as a moral aspiration but as an evaluative principle that can inform both AI system design

and governance frameworks. The DCAF provides researchers, policymakers, and developers with a structured, evidence-based approach to assessing and improving human–AI interaction quality. In doing so, it bridges philosophical ethics with computational accountability, enabling dignity to serve as both a normative anchor and a measurable construct in AI ethics research.

Future research can extend this work in several directions. First, longitudinal studies could examine how perceptions of dignity evolve as AI systems become more autonomous and integrated into daily life. Second, cross-cultural validations are necessary to explore how varying cultural conceptions of dignity influence AI ethics perceptions and trust. Third, extending the DCAF to specific domains—such as healthcare, education, creative industries, and public administration—would strengthen its contextual applicability and policy relevance. Finally, technological implementations of the DCAF, such as "dignity compliance dashboards" or ethical audit tools, could provide real-time assessment mechanisms for AI developers and regulators.

In conclusion, this research underscores that the progress of artificial intelligence should not be measured solely by its cognitive capabilities or generative sophistication but by its capacity to preserve and enhance the moral worth of the human person. Embedding dignity into AI design and governance is not an abstract ideal, it is a practical imperative for ensuring that technological intelligence remains in service of human flourishing and societal good.

References

1. Agyare, P. (2024). Contextualizing human rights in multicultural environments. Research in Social Sciences and Technology, 9(3), 10-46303.
2. Banerjee, S., Thomas, M., Chavan, V., Mangla, U., & Tummalapenta, S. (2025, May). Securing the future of AI: a holistic approach to trust and robustness. In Assurance and Security for AI-enabled Systems 2025 (Vol. 13476, pp. 121-142). SPIE.
3. Baxter, H. (1987). System and lifeworld in Habermas's" theory of communicative action". Theory and Society, 39-86.
4. Bayefsky, R. (2013). Dignity, honour, and human rights: Kant's perspective. Political Theory, 41(6), 809-837.
5. Chakraborty, S. (Ed.). (2025). Human Values, Ethics, and Dignity in the Age of Artificial Intelligence. IGI Global.
6. Chittimalla, S. K., & Potluri, L. K. M. (2025, March). Explainable AI Frameworks for Large Language Models in High-Stakes Decision-Making. In 2025 International Conference on Advanced Computing Technologies (ICoACT) (pp. 1-6). IEEE.
7. Cooke, M. (2009). Beyond dignity and difference: revisiting the politics of recognition. European Journal of Political Theory, 8(1), 76-95.
8. Desai, D. R., & Kroll, J. A. (2017). Trust but verify: A guide to algorithms and the law. Harv. JL & Tech., 31, 1.
9. Edor, E. J. (2020). John Rawls's Concept of Justice as Fairness. PINISI Discretion Review, 4(1), 179-190.

10. Fasoro, A. (2024). Cultivating Dignity in Intelligent Systems. Philosophies, 9(2), 46.
11. Finegan, T. (2012). Conceptual foundations of the universal declaration of human rights: human rights, human dignity and personhood. Austl. J. Leg. Phil., 37, 182.
12. Floridi, L. (2022, December). The green and the blue: a new political ontology for a mature information society. In The Green and the Blue (pp. 9-52). Verlag Karl Alber.
13. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. Minds and machines, 28(4), 689-707.
14. Fredman, S. (2007). Redistribution and recognition: Reconciling inequalities. South African Journal on Human Rights, 23(2), 214-234.
15. Friedman, B., & Hendry, D. G. (2019). Value sensitive design: Shaping technology with moral imagination. Mit Press.
16. Haimer, R. (2021). Recognition and Reification in the Philosophy of Axel Honneth.
17. Hoxhaj, O., Halilaj, B., & Harizi, A. (2023). Ethical implications and human rights violations in the age of artificial intelligence. Balkan Social Science Review, 22(22), 153-171.
18. Innocenti, M. (2025). Value Pluralism and Autonomy in Meaningful Work: Rethinking the Role of AI Moral Advisors in Innovation. Available at SSRN 5386124.
19. Joseph, S., & Anantharaman, V. (2024). Algorithmic Bias and Human Rights: Exploring the Intersection and Implications. Issue 1 Int'l JL Mgmt. & Human., 7, 1526.
20. Koulu, R. (2020). Human control over automation: EU policy and AI ethics. Eur. J. Legal Stud., 12, 9.
21. Kaushik, N., Khilwani, M., & Bajaj, N. (2024). Responsible Tech Innovation for Humanity: Navigating Ethical Challenges and Opportunities through the Case Study of Fairphone. Lloyd Business Review, 1-28.
22. Laitinen, A., & Sahlgren, O. (2021). AI systems and respect for human autonomy. Frontiers in artificial intelligence, 4, 705164.
23. Leavy, S., Siapera, E., & O'Sullivan, B. (2021, July). Ethical data curation for AI: An approach based on feminist epistemology and critical theories of race. In Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society (pp. 695-703).
24. Mahajan, P. (2025). The Soul of the AI: Governance, Ethics, and the Future of Human–AI Integration
25. Nguyen, A., Ngo, H. N., Hong, Y., Dang, B., & Nguyen, B. P. T. (2023). Ethical principles for artificial intelligence in education. Education and information technologies, 28(4), 4221-4241.
26. Nuredin, A., & Inan, T. C. (2024). The impact of AI-based decision-making system on justice and equality. International Scientific Journal Vision, 9.
27. Panchal, J. (2024). Ethics and artificial intelligence in the age of automation: reexamining moral frameworks in techno-ethical dilemmas. ShodhKosh Journal of Visual and Performing Arts. https://doi. org/10.29121/shodhkosh. 5.

28. Pasopati, R. U., Bethari, C. P., Nurdin, D. S. F., Camila, M. S., & Hidayat, S. A. (2024, March). Ethical Consequentialism in Values and Principles of UNESCO's Recommendation on the Ethics of Artificial Intelligence. In Proceeding International Conference on Religion, Science and Education (Vol. 3, pp. 567-579).
29. Poibeau, T. (2025). Foundations of Conversational AI.
30. Plasencia, M. M. (2024). Reinforcement Learning from Human Feedback for Ethically Robust Ai Decision-Making.
31. Rachels, J. (1986). Kantian theory: the idea of human dignity. The elements of moral philosophy, 114-117.
32. Rodriguez, P. A. (2015). Human dignity as an essentially contested concept. Cambridge Review of International Affairs, 28(4), 743-756.
33. Russell, S., Norvig, P., Popineau, F., Miclet, L., & Cadet, C. (2021). Intelligence artificielle: une approche moderne (4e édition). Pearson France.
34. Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. American psychologist, 55(1), 68.
35. Sensen, O. (2011). Kant on human dignity (Vol. 166). Walter de Gruyter.
36. Shestack, J. J. (2017). The philosophic foundations of human rights. In Human rights (pp. 3-36). Routledge.
37. Shvartzshnaider, Y., & Duddu, V. (2025). Position: Contextual Integrity is Inadequately Applied to Language Models. arXiv preprint arXiv:2501.19173.
38. Sikos, L. F., & Philp, D. (2020). Provenance-aware knowledge representation: A survey of data models and contextualized knowledge graphs. Data Science and Engineering, 5(3), 293-316.
39. Steeves, V. M. (2005). Beyond data protection: applying Mead's symbolic interactionism and Habermas's communicative action to Westin's theory of privacy (Doctoral dissertation, Carleton University).
40. Hanna, M. (2025). A sociological theory of privacy as a communicative right. Revista Direito Mackenzie, 19(1).