ISSN:1738-6764, eISSN:2093-7504

Research Ethics of Artificial Intelligence in the Digital Transformation Era: Fairness, Accountability, and Human-Centered Design

Usmanov Doniyor [1], Sungwoong Hong [2,*] and Jinhee Kim [3]

[1] National Advisor on Investment and Technology Promotion in Uzbekistan, United Nations Industrial Development Organization (UNIDO); d.usmanov@unido.org
[2] Department of Software Convergence, Cheongju University; leoking@cju.ac.kr
[3] Department of Biomedical Laboratory Science, Cheongju University; jinheekim23273@gmail.com
* Correspondence

https://doi.org/10.5392/IJoC.2025.21.4.010

Manuscript Received 4 December 2025; Received 25 December 2025; Accepted 25 December 2025

Abstract: *The rapid expansion of digital transformation across industry, government, education, healthcare, and finance has positioned Artificial Intelligence (AI) as a central driving force of societal change. AI technologies are being integrated into diverse domains, restructuring industrial systems, innovating public services, and enhancing personalized quality of life. They significantly improve efficiency and productivity while creating new markets and employment opportunities. In particular, the rise of generative AI marks a qualitative shift in the technological paradigm, extending its influence into high-level human activities such as information creation, creative work, and decision support. However, these technological advancements and societal diffusion simultaneously generate a range of ethical and social risks. Algorithmic bias, opacity, privacy violations, workforce displacement, and unclear responsibility attribution undermine trust and societal acceptance of AI. Generative AI introduces additional concerns, including the veracity of generated content, copyright violations, and the proliferation of deepfakes, while philosophical and legal debates about accountability for AI-generated outcomes remain unresolved. In response to these challenges, this study re-examines the core principles and responsibility structures required for AI research ethics in the digital transformation era. It investigates an integrated research ethics framework grounded in three pillars: fairness, accountability, and human-centered design (HCD). Reflecting the expanding role of generative AI and evolving Human–Computer Interaction (HCI) dynamics, the study proposes a practical, operational evaluation model that overcomes the limitations of principle-centric ethical discourse. Specifically, it introduces the Integrated FAH–AIL Evaluation Framework (IFAEF), which combines the FAH-75 index for ethical compliance with the AIL-5 index that assesses AI intelligence and risk levels. As AI evolves from a tool to a "partner-like entity" influencing societal decision-making and human activity, mere technical performance is insufficient. Ethical frameworks must prioritize human dignity, societal context, and value-based judgment. Accordingly, AI research ethics must function not as a passive regulatory mechanism but as a strategic foundation for the sustainable coexistence of technology and society. This requires a multidimensional collaborative structure involving government, industry, academia, and civil society. This paper aims to contribute to this goal by examining the ethical conditions necessary throughout the AI development and deployment lifecycle, and by presenting actionable directions and quantitative evaluation standards for researcher ethics suited to the digital transformation era. In doing so, it offers a concrete foundation for strengthening the accountability, fairness, and human-centeredness of AI technologies.*

Keywords: Artificial Intelligence; Digital Transformation; Research Ethics; Generative AI; Accountability; Fairness; Human-Centered Design

1. Introduction

Digital Transformation has triggered structural changes across all sectors of society—including industry, public administration, education, healthcare, and finance—with artificial intelligence (AI) positioned as the core driving force behind this shift [1]. AI has expanded data-driven decision-making systems and enabled improvements in efficiency, cost reduction, automation, and large-scale information processing capabilities [1]. In particular, the recent proliferation of generative AI demonstrates that AI has entered a stage in which it directly intervenes in human knowledge production, creation, analysis, prediction, and judgment, signaling a transition in which AI is integrated as a constitutive component of social systems rather than functioning merely as a tool [2, 3]. However, the accelerated pace of digital transformation also introduces new ethical and social risks [4]. One of the most prominent concerns is the ambiguity of responsibility arising when AI systems learn and make decisions autonomously, a problem defined as the responsibility gap [2]. When an AI system makes an erroneous decision, unresolved questions remain regarding who—developers, operators, or users—should bear responsibility, and to what extent [5]. Furthermore, if the data used for AI training reflects or amplifies existing social inequalities, this may lead to algorithmic bias and fairness issues [1], [4]. Generative AI has also brought fundamental changes to the research environment. While traditional research ethics (RCR) emphasized principles such as honesty, data integrity, and the prevention of plagiarism, contemporary research ethics increasingly require the disclosure of AI usage, verification of AI-generated content, identification of data sources, and prevention of misuse [6-8]. International scholarly communities now demand that researchers independently validate the authenticity of AI-generated outputs and transparently disclose the extent of AI involvement in their work [3], [6]. At the same time, AI ethics discourse has shifted from technology-oriented rules to human-centered values. The OECD, EU, and UNESCO position human dignity, human agency, social trust, and inclusiveness as the core values of AI ethics [1], [9, 10], reinforcing the necessity of constructing a research ethics framework based on Human-Centered Design (HCD) [11-13]. Therefore, AI research ethics in the digital transformation era must be reconstructed around three central pillars: Fairness, Accountability, and Human-Centered Design (HCD) [2], [12, 13]. Recognizing the limitations of existing research—which often exhibits fragmented approaches and a lack of actionable models [4], [14, 15]—this study proposes a new AI research ethics evaluation methodology that integrates these three elements while additionally incorporating AI intelligence levels [1], [13], [16].

2. Related Research

Research on AI research ethics in the digital transformation era has developed around five major pillars:

(1) international norms and principles,
(2) the responsibility gap,
(3) fairness and bias,
(4) human-centered design, and
(5) the re-establishment of research ethics in response to the rise of generative AI [1].

While international guidelines have established a shared set of ethical values for AI, practical and integrated evaluation models usable in real research and development contexts remain insufficient [2], [4].

2.1 International AI Ethics Norms and Principles

The AI Principles established by the OECD in 2019 defined human-centered values, fairness, transparency, safety, and accountability as core components of AI ethics and laid the foundation for international discussions [1]. In the same year, the EU published its Ethics Guidelines for Trustworthy AI, specifying requirements for "trustworthy AI" [9], and UNESCO strengthened AI ethics to a global normative level through its 2021 Recommendation [10]. However, Floridi and Cowls argue that although international guidelines use similar language, they embed heterogeneous ethical concepts and require a unified metaframework [2]. Jobin et al. found that many global AI ethics guidelines repeat the same values but lack concrete mechanisms for implementation and evaluation [4]. Thus, while international norms are abundant, their application through empirical or operational models remains limited [4].

2.2 Responsibility Gap

Accountability is one of the oldest and most fundamental issues in AI ethics. Matthias defined the situation in which learning AI systems autonomously make decisions that no longer allow clear attribution of responsibility as the "responsibility gap" [5]. The responsibility gap has since become a central challenge in norms, law, and policy, particularly in domains such as medical AI misdiagnosis, autonomous vehicle accidents, and automated credit scoring [2], [5]. However, research on the responsibility gap remains largely rooted in moral philosophy, creating significant gaps with real-world operational contexts, and lacks clear legal or institutional criteria for responsibility allocation [2], [5].

2.3 Fairness and Bias

Fairness is among the ethical values most strongly emphasized by the OECD, EU, and UNESCO [1], [9, 10]. Numerous studies highlight that AI systems, which rely on data-driven learning, may internalize or reinforce societal biases [4]. Fairness research has expanded across three layers—data-level bias, algorithmic structural bias, and outcome bias [1, 2], [12]—but much of the work remains measurement-oriented, with insufficient development of actionable mitigation frameworks [2], [4].

2.4 Human-Centered Design (HCD/HCAI)

Recent AI ethics discourse has shifted from technical rule-based approaches toward human-centered perspectives [11]. Schmager et al. identified key components of human-centered AI design, including human controllability, explainability, predictability, trustworthiness, ethical safety, and social acceptability [12]. Bevilacqua defined human-centered AI as a design philosophy that ensures human involvement, control, psychological safety, and societal trust [13]. Nevertheless, Wilkens and colleagues found that much of the human-centered AI literature remains conceptual or declarative, with limited empirical validation [14, 15].

2.5 Generative AI and the Re-establishment of Research Ethics

Since the rise of generative AI, "Generative AI + Research Ethics" has become one of the most rapidly expanding areas of inquiry [3], [6]. Nature warned that generative AI could destabilize or collapse scholarly production systems and argued for stronger researcher-led oversight mechanisms [3]. The ERA Forum's guidelines established explicit obligations for researchers, including AI usage disclosure, verification responsibilities, provenance transparency, privacy considerations, risk evaluation, and misuse prevention [6]. Yet these guidelines remain mostly principle-driven, leaving a persistent gap between ethical norms and practical implementation [7], [17].

2.6 Limitations of Existing Research and Contributions of This Study Despite the richness of international AI ethics principles, existing research shares several limitations:

(1) lack of executable evaluation models,
(2) fragmented treatment of fairness, accountability, and HCD,
(3) insufficient incorporation of risks associated with post–generative AI environments,
(4) lack of empirical or practice-based validation, and
(5) limited reflection of sector-specific characteristics, particularly in the content industry [1], [3, 4], [13, 14].

To address these gaps, this study integrates the three core ethical components and incorporates AI intelligence levels (AIL-5) to propose a practical and quantifiable evaluation methodology for AI research ethics.

3. Integrated Ethical Framework for AI Research

This section analyzes AI research ethics in the digital transformation era in terms of three core pillars—fairness, accountability, and human-centered design—and further integrates generative AI, HCI considerations, and intelligence levels (AIL-5) to propose an integrated evaluation methodology.

3.1 Research Ethics from the Perspective of Fairness

Fairness is a foundational value that determines the ethical acceptability of AI technologies [1]. Because AI systems learn and make decisions based on data, insufficient representativeness or embedded historical bias can result in structural disadvantages for particular groups [4]. As AI increasingly mediates resource allocation, evaluation, and recommendation processes within digital transformation environments, fairness concerns evolve from technical imperfections to systemic social risks [18]. Accordingly, fairness must be institutionalized as a lifecycle-based procedure encompassing (1) minimizing sampling bias at the data collection stage, (2) bias detection and mitigation at the algorithmic design stage, and (3) verifying that outcomes do not reproduce or exacerbate inequality at the deployment stage [1], [12], [19].

3.2 Research Ethics from the Perspective of Accountability

Accountability is the most fundamental concern in AI research ethics, and the responsibility gap remains an area without complete institutional solutions [5]. In digital transformation environments, accountability extends beyond responsibility for automated decisions to include obligations such as the researcher's duty to verify AI-generated outputs, responsibility for data governance, and mechanisms for preventing research misconduct [2]. In particular, the rise of generative AI has strengthened norms requiring researchers to (1) take ultimate responsibility for verifying AI-generated results and (2) disclose the extent of AI involvement in their work [3], [6, 7]. Thus, accountability must be re-established as a proactive ethical element integrated into research design—shifting from post-hoc regulation to anticipatory responsibility frameworks that identify risks and specify responsibility structures in advance [7], [17].

3.3 Research Ethics from the Perspective of Human-Centered Design (HCD)

Human-centered design reflects the ethical demand that AI systems must not violate but rather enhance human values, rights, autonomy, and welfare [10, 11], [13]. Key components of HCD include human controllability, explainability, risk awareness, and psychological and social acceptability [12, 13]. If HCD is not institutionalized during the research and development process, fairness and accountability risk remaining at the level of technical declarations rather than becoming actionable ethical practices [14]. Therefore, research ethics must integrate HCD as the foundational philosophical frame that supports fairness and accountability and continuously monitor how well human values are embedded throughout the AI lifecycle [13], [20].

3.4 Ethical Issues of Generative AI from an HCI Perspective and Corresponding Responses

Generative AI intervenes in problem-solving, information acquisition, and content creation processes, fundamentally altering human cognition and interaction. From an HCI perspective, the central ethical issues can be summarized as:

(1) weakened user control,
(2) instability of trust experiences,
(3) increased cognitive dependency,
(4) erosion of authorship and creative agency, and
(5) deterioration of human-to-human interaction.

To address these issues, the following strategies are required:

— Human-in-the-Loop design,
— construction of Explainable UX,
— designing AI for human cognitive augmentation rather than replacement,
— transparent attribution and provenance systems,
— and interface designs that preserve human-human interaction [9-14].

These strategies provide concrete methods for restructuring generative AI systems within a human-centered design paradigm.

4. Design of IFAEF(Integrated FAH–AIL Evaluation Framework)

4.1 Design of the FAH-75

The FAH-75 model quantitatively evaluates compliance with AI research ethics based on the three pillars of fairness (F), accountability (A), and human-centered design (H). It draws on the OECD AI Principles [1], the EU Trustworthy AI Guidelines [9], UNESCO's Recommendation [10], Floridi's ethical design principles [2], [19], and the responsibility gap literature [5]. FAH-75 comprises 15 items (5 per pillar), each scored on a 1–5 scale, yielding a maximum of 75 points. The model aims to move beyond principle-based ethical discourse toward actionable and measurable procedures [4], [10, 11]. It also incorporates a three-stage evaluation process—self-assessment, expert review, and stakeholder verification—in line with the heightened research accountability norms of the generative AI era [3], [6, 7].

4.2 Structure and Evaluation Criteria of the FAH-75

The FAH-75 (Fairness–Accountability–Human-Centered Design) evaluation framework proposed in this study is designed to quantitatively assess key ethical dimensions of AI research. The framework consists of a total of 15 evaluation items, each scored on a five-point scale ranging from 1 (very insufficient) to 5 (excellent), yielding a maximum score of 75 points. The items are organized into three core ethical dimensions: Fairness (F), Accountability (A), and Human-Centered Design (H).

(1) Fairness (25 points)

The fairness dimension assesses whether an AI system is designed and operated in a manner that prevents the reproduction or amplification of social biases. The evaluation focuses on the following five criteria:

Data Representativeness: Whether training data adequately reflect demographic and contextual diversity.

Bias Detection and Mitigation Mechanisms: The existence of systematic procedures to identify and reduce algorithmic bias.

Disparate Impact Analysis: Assessment of whether algorithmic outcomes produce discriminatory effects across social groups.

Pre-assessment of Social Inequality Risks: Evaluation of potential risks related to reinforcing structural inequalities.

Continuous Fairness Monitoring: Presence of mechanisms for ongoing fairness evaluation after deployment.

Each item is scored based on the transparency, consistency, and verifiability of fairness-related controls and processes.

(2) Accountability (25 points)

The accountability dimension evaluates whether responsibility for AI system behavior is clearly defined and enforceable across the system lifecycle.

Attribution of Responsibility: Clarity regarding who is accountable (developers, operators, institutions).

Disclosure of AI Use and Intervention Scope: Transparency regarding when and how AI systems influence decisions.

Verification and Error-Correction Mechanisms: Existence of procedures for validating outputs and correcting errors.

Data Governance and Security Responsibility: Structures for managing data integrity, privacy, and security.

Risk Response and Liability Mechanisms: Defined protocols for managing failures, harm, and accountability allocation.

Scores are assigned based on the clarity of responsibility structures, documentation practices, and the feasibility of external auditing.

(3) Human-Centered Design (HCD) (25 points)

The human-centered design dimension evaluates whether AI systems enhance human autonomy, trust, and well-being rather than undermine them.

Human-in-the-Loop Control: The extent to which humans can intervene, override, or supervise system decisions.

Explainability and Interpretability: The degree to which system behavior can be understood by users.

Psychological and Cognitive Safety: Consideration of user well-being and avoidance of manipulation or overreliance.

Human Capability Augmentation: Whether AI is designed to support and enhance human judgment rather than replace it.

Social Acceptability and Ethical Impact Awareness: Assessment of broader societal implications and ethical acceptability.

(4) Evaluation emphasizes user experience, social consequences, and long-term ethical implications rather than purely technical performance. Scoring Methodology (1–5 Scale)

Each evaluation item is assessed according to the following criteria:

Table 1. Scoring Methodology (1–5 Scale)

Score	Description
1	Not considered or presents serious ethical risks
2	Minimal consideration; limited to formal or superficial measures
3	Partial implementation with limited verification
4	Systematic implementation meeting most requirements
5	Exemplary level with verifiable and continuously improving practices

These criteria are designed not to eliminate qualitative judgment, but to ensure consistency and reproducibility across evaluators by providing a semi-quantitative assessment framework.

(5) Multi-Stage Evaluation Structure and Conflict Resolution Mechanism

The FAH-75 framework adopts a three-stage evaluation structure:

1) Self-Assessment
Developers or researchers conduct an internal assessment to identify potential ethical risks and compliance gaps at an early stage.

2) Expert Review
Domain experts in AI ethics, law and policy, and human–computer interaction (HCI) review and validate the assessment results, providing professional judgment and corrective feedback.

3) Stakeholder Verification
Stakeholders—including end users, civil society representatives, and domain specialists—evaluate the system from a societal impact perspective.

When discrepancies arise among the three evaluation stages, the precautionary principle is applied: the lowest score is adopted as the final value, and the rationale for this decision must be explicitly documented. This mechanism is designed to prevent the underestimation of ethical risks and to ensure conservative, responsibility-oriented decision-making.

4.3 The AI Intelligence Level Framework (AIL-5)

Ethical requirements for AI must intensify as the “intelligence level” of AI systems increases. Reflecting advancements in cognitive science, HCI, and AGI research, this study defines the AIL-5 (AI Intelligence Level-5) framework as follows:

AIL-1: Reactive intelligence (rule-based or simple statistical systems)
AIL-2: Pattern-learning intelligence (ML/DL prediction models)
AIL-3: Language and reasoning intelligence (LLM-based inference and generation)
AIL-4: Autonomous agent intelligence (goal formulation, planning, and action)
AIL-5: General adaptive intelligence (AGI-oriented, highest-risk systems)
AIL-5 is not a mere performance taxonomy.

It functions as an ethical-risk indicator that integrates autonomy, generality, societal impact, high-risk decision involvement, and sensitivity to data. Therefore, as AIL levels increase, AI systems require proportionally stronger safeguards in fairness, accountability, and human-centered design.

Table 2. AIL-1 ~ AIL-5 Intelligence Levels

AIL Level	Intelligence Category	Core Characteristics (Cognitive Perspective)	Representative Examples	Ethical / Research Ethics Implications
AIL-1	Reactive Intelligence	Rule-based or simple statistical responses; no context, memory, or self-adaptation	Rule-based systems, simple classifiers	Very low risk. Ethics focus on basic data quality, error checking, and verification.
AIL-2	Pattern-Learning Intelligence	ML/DL-based pattern learning and prediction; limited semantic understanding	Conventional deep learning models, image/speech recognition	Data bias becomes a central issue. Fairness (F) requirements increase significantly.
AIL-3	Cognitive Reasoning Intelligence	LLM-driven language understanding/generation; capable of chain-of-thought reasoning; hallucination and reliability problems	GPT-family, multimodal LLMs	Increased need for trust, transparency, and verification. Accountability (A) and HCD requirements rise sharply.
AIL-4	Agentic Intelligence	Goal setting, planning, tool use, and autonomous action execution	AutoGPT-type systems, workflow/automation agents	High risk of responsibility gaps and weakened human control. Very strong A and H requirements.
AIL-5	General Adaptive Intelligence (AGI-oriented)	Domain-unbounded generality; environmental adaptation; self-extension; highest-level autonomy	(Current research and hypothetical stage)	Highest risk level. AIL-5 functions as an ethical risk indicator integrating autonomy, generality, societal impact, high-risk decision involvement, and data sensitivity → Fairness (F), Accountability (A), and Human-Centered Design (HCD) demands increase proportionally with AIL level.

3.7 FAH-75 × AIL-5 Integrated Evaluation Methodology (IFAEF)

FAH-75 operates in conjunction with the AIL-5 intelligence level framework. As AI systems exhibit higher levels of autonomy, generality, and social impact, the required ethical threshold increases accordingly. Thus, FAH scores are not evaluated in isolation but are interpreted relative to the system's AIL level.

This design ensures that:

Low-risk systems are not over-regulated, and High-risk or agentic systems are subject to proportionally stronger ethical requirements.

Existing AI ethics assessments remain largely at the level of declarative principles and fail to adequately reflect varying levels of AI risk. To address these limitations, this study proposes an integrated evaluation model—IFAEF (Integrated FAH–AIL Evaluation Framework).

The framework consists of four components:

1. FAH-75 Ethical Compliance Assessment (0–75 points)

Evaluates fairness (F), accountability (A), and human-centered design (H) across 15 items, each rated on a 1–5 scale.

2. AIL-5 Intelligence and Risk Level Assessment (5–25 points)

Scores the AI system based on autonomy, generality, societal impact, high-risk decision involvement, and data sensitivity.

3. Level-Based Ethical Threshold Adjustment

Ethical requirements are strengthened proportionally to the AI's intelligence/risk score (L).

A higher L score requires a higher minimum FAH score to pass the evaluation.

4. Final Score Calculation

ESS = FAH + L (Total: 100 points)

The system is deemed ethically acceptable only when the FAH score meets or exceeds the threshold corresponding to its L score.

Table 3. Structure of the Integrated FAH–AIL Evaluation Framework (IFAEF)

Component	Detailed Description	Score
FAH-75 (Ethical Compliance Assessment)	Fairness (F), Accountability (A), Human-Centered Design (H), 25 points each	0–75 points
AIL-5 (Intelligence & Risk Assessment)	Autonomy, Generality, Social Impact, High-Risk Intervention, Data Sensitivity (1–5 points each)	5–25 points
ESS (Final Ethical Suitability Score)	ESS = FAH + L	Total 100 points
Ethical Passing Criterion	FAH ≥ Threshold(L)	PASS / FAIL

Table 4. Minimum FAH Passing Thresholds by AIL (L-Score) Category

AIL L-Score Range	AI Intelligence / Risk Level	Required Minimum FAH Score (out of 75)	Characteristics
5–9 points	AIL-1~2 (Low intelligence / Low risk)	≥ 45	Traditional models, rule-based / ML-based systems
10–14 points	AIL-3 (Language / reasoning intelligence)	≥ 52	LLM- and reasoning-based systems; increased risks of hallucination and bias
15–19 points	AIL-3.5~4 (High intelligence / High exposure)	≥ 60	High-performance LLMs, tool-use capability, partial autonomy
20–25 points	AIL-4~5 (Autonomous agents / AGI-oriented)	≥ 68	Goal-setting, autonomous task execution, potential high-risk interventions

Table 5. IFAEF Evaluation Result Calculation

Item	Calculation Method	Description
ESS (Final Score)	ESS = FAH + L	Total score out of 100
Pass/Fail Decision	FAH ≥ Threshold(L)	Higher AI intelligence levels require stricter ethical thresholds

The characteristics of the IFAEF integrated evaluation model are distinguished from existing AI ethics assessment frameworks in four major aspects.

First, the model adopts a level-adjusted ethical threshold, meaning that the required levels of fairness, accountability, and human-centered design automatically increase as the AI's intelligence level and associated risks rise. This transforms the abstract guidance of international AI ethics principles into concrete, risk-responsive regulatory criteria that correspond to the actual capabilities of AI models [1, 2], [12].

Second, IFAEF enables a quantified and operational ethics assessment, moving beyond the declarative and principle-driven evaluations that dominate existing approaches. By assigning numerical values (25 points each) to fairness (F), accountability (A), and human-centered design (H), and by evaluating autonomy, generality, and high-risk action potential through the AIL index, the framework allows researchers and institutions to diagnose the ethical status of AI systems quantitatively. This advances the field by overcoming the ambiguity of traditional ethical discourse and providing a comparable and verifiable evaluation methodology [4], [16].

Third, the model has multi-dimensional applicability across research, development, and policy domains. FAH-75 covers both research ethics (RCR) and practical development ethics, while AIL-5 is designed to support policy-oriented risk and impact assessments. Accordingly, IFAEF is not only a tool for researchers but can also be applied to service operation standards, administrative decision-making systems, and public-sector AI impact assessment frameworks [3], [6], [9].

Fourth, IFAEF surpasses existing models by providing an ethical computation framework capable of addressing agentic AI and AGI-level systems. AI at AIL-4 to AIL-5 possesses high autonomy—such as goal-setting, action execution, and tool use—and consequently introduces dramatically heightened risks, including responsibility gaps, weakened human control, and amplified societal impacts. By incorporating strengthened accountability checks and HCD requirements for these high-risk systems, the framework offers an ethical preparedness model suitable for the AGI-oriented technological era [3], [19].

In conclusion, IFAEF quantitatively reflects risk levels associated with AI development stages and incorporates an adaptive structure that automatically adjusts ethical requirements. Thus, it can be regarded as a practical, predictive, and integrated evaluation framework that supports AI research ethics in the digital transformation era.

5. Conclusion and Recommendations

This study integrated three core pillars of AI research ethics in the digital transformation era—fairness, accountability, and human-centered design—and proposed a combined evaluation methodology (IFAEF) that incorporates generative AI, HCI perspectives, and AI intelligence levels (AIL-5). Fairness requires institutionalizing procedures that proactively embed bias mitigation across all stages of research—data, algorithms, and outcomes [1], [4], [12]. Accountability must address the responsibility gap by adopting shared responsibility structures and by ensuring that researchers uphold verification duties throughout the AI lifecycle [3], [5, 6]. Human-centered design should be embedded as the foundational ethical philosophy that ensures human control, autonomy, and social trust beyond mere technical stability [10], [12, 13].

Based on these findings, this study offers the following practical recommendations:

First, fairness should be reinforced by mandating data diversity validation and algorithmic impact assessments (AIA) during the research process.

Second, in studies utilizing generative AI, norms for AI use disclosure, source attribution, and result verification must be institutionalized within the national scholarly ecosystem.

Third, an HCD-based ethical standard tailored to the characteristics of the content industry should be developed to safeguard creators' rights and strengthen user trust.

Fourth, the combined FAH-75 × AIL-5 evaluation framework must be applied across various real-world contexts, and subsequent empirical studies should validate its practical effectiveness.

In conclusion, AI research ethics in the digital transformation era must evolve from a passive framework that merely controls technological harm to a strategic foundation enabling the sustainable use of AI for humanity and society. The integrated framework and evaluation methodology proposed in this study provide a practical basis for achieving this goal.

Acknowledgments: This research was supported by the MSIT(Ministry of Science ICT), Korea, under the National Program for Excellence in SW, supervised by the IITP(Institute of Information & Communications Technology Planning & Evaluation) (No. 2025-0-00056). This research was supported by the Regional Innovation System & Education (RISE) program through the Chungbuk Regional Innovation System & Education Center, funded by the Ministry of Education (MOE) and the Chungcheongbuk-do, Republic of Korea (2025-RISE-11-013-03).

Conflicts of Interest: The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

[1] OECD, OECD Principles on Artificial Intelligence: Human-Centred Values & Fairness; Transparency & Explainability; Robustness, Security & Safety; Accountability, OECD Council Recommendation on Artificial Intelligence, 2019.

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

[3] C. L. Bockting, E. A. M. van Dis, R. van Rooij, W. Zuidema, and J. Bollen, "Living Guidelines for Generative AI — Why Scientists Must Oversee Its Use," Nature, vol. 622, no. 7984, pp. 693–696, 2023, doi: https://doi.org/10.1038/d41586-023-03266-1.

[4] A. Jobin, M. Ienca, and E. Vayena, "The Global Landscape of AI Ethics Guidelines," Nature Machine Intelligence, vol. 1, pp. 389–399, 2019, doi: https://doi.org/10.1038/s42256-019-0088-2.

[5] A. Matthias, "The Responsibility Gap: Ascribing responsibility for the actions of learning automata," Ethics and Information Technology, vol. 6, pp. 175–183, 2004, doi: https://doi.org/10.1007/s10676-004-3422-1.

[6] European Research Area Forum, Living Guidelines on the Responsible Use of Generative AI in Research, European Commission, 2024.

[7] A. da Veiga, "Ethical Guidelines for the Use of Generative Artificial Intelligence and Artificial Intelligence-Assisted Tools in Scholarly Publishing: A Thematic Analysis," Science Editing, vol. 12, no. 1, pp. 28–34, 2025, doi: https://doi.org/10.6087/kcse.352.

[8] S. H. Lee, "Issues and Trends Related to Artificial Intelligence in Research Ethics," Health Policy and Management, vol. 34, no. 2, pp. 103–105, Jun. 30, 2024, doi: https://doi.org/10.4332/KJHPA.2024.34.2.103.

[9] High-Level Expert Group on Artificial Intelligence, Ethics Guidelines for Trustworthy AI, European Commission, 2019.

[10] UNESCO, Recommendation on the Ethics of Artificial Intelligence, 2021.

[11] H. Yoo, B. Yoon, and H. Choi, AI ethics: Strategies for harmonious coexistence between humans and artificial intelligence, KISTI Issue Brief, no. 68, 2024.

[12] S. Schmager, I. Pappas, and P. Vassilakopoulou, "Defining Human-Centered AI: A Comprehensive Review of HCAI Literature," MCIS 2023 Proceedings, p. 13, 2023. [Online] Available: https://aisel.aisnet.org/mcis2023/13

[13] R. Bevilacqua, T. Bailoni, E. Maranesi, G. Amabili, F. Barbarossa, M. Ponzano, … G. Sakellariou, "Framing the Human-Centered Artificial Intelligence Concepts and Methods: Scoping Review," JMIR Human Factors, vol. 12, e67350, 2025, doi: https://doi.org/10.2196/67350.

[14] U. Wilkens, C. Cost Reyes, T. Treude, and A. Kluge, Understandings and Perspectives of Human-Centered AI – A Transdisciplinary Literature Review, GfA Frühjahrskongress, 2021.

[15] A. Mazarakis, C. Bernhard-Skala, M. Braun and I. Peters, "What is critical for human-centered AI at work? – Toward an interdisciplinary theory," Front. Artif. Intell, vol. 6, 1257057, 2023, doi: https://doi.org/10.3389/frai.2023.1257057.

[16] N. Díaz-Rodríguez, J. Del Ser, M. Coeckelbergh, M. López de Prado, E. Herrera-Viedma, and F. Herrera, "Connecting the Dots in Trustworthy Artificial Intelligence: From AI Principles, Ethics, and Key Requirements to Responsible AI Systems and Regulation," arXiv preprint arXiv:2305.02231, 2023.

[17] S. Kapania, R. Wang, T. J. J. Li, T. Li, and H. Shen, ""I'm Categorizing LLM as a Productivity Tool": Examining Ethics of LLM Use in HCI Research Practices," Proceedings of the ACM on Human-Computer Interaction, 2025.

[18] O. Ozmen Garibay, B. Winslow, S. Andolina, M. Antonaci, et al., "Six Human-Centered Artificial Intelligence Grand Challenges," International Journal of Human–Computer Interaction, vol. 39, no. 3, pp. 371-437, 2023, doi: https://doi.org/10.1080/10447318.2022.2153320.

[19] L. Floridi, The Ethics of Artificial Intelligence: Principles, Challenges, and Opportunities, Oxford University Press, 2023.

[20] U. Wilkens, D. Lupp, and V. Langholf, "Configurations of human-centered AI at work: seven actor-structure engagements in organizations," Front. Artif. Intell, vol. 6, 1272159, 2023, doi: https://doi.org/10.3389/frai.2023.1272159.