systems

MDPI

Article

Operationalising Human-Centred AI Governance Under the EU AI Act: A Governance Framework for Human Oversight and Data Accountability

Hyun-Kyung Lee [1], Cheolhee Yoon [2] and Bong Gyou Lee [1,*]

[1] Graduate School of Information, Yonsei University, Seoul 03722, Republic of Korea; smhk77@yonsei.ac.kr
[2] Science and Technology Research Division, Police Science Institute, Korean National Police University, Asan-si 31539, Republic of Korea
* Correspondence: bglee@yonsei.ac.kr

Abstract

Artificial intelligence (AI) is increasingly embedded in high-stakes socio-technical systems, intensifying concerns about autonomy, accountability, data rights, and fundamental-rights protection. This article develops an exploratory, expert-informed Human-Centred AI (HCAI) pre-design governance framework that translates selected risk-based obligations of the EU Artificial Intelligence Act into early organisational decisions about human oversight, data accountability, documentation, and bounded algorithmic autonomy. Using a sequential mixed-methods design, the study combines an Analytic Hierarchy Process (AHP) survey of 28 experts with think-aloud interviews with 15 of those respondents. The AHP results show that, among the governance criteria included in the model, AI design objectives received the highest upper-level priority and human oversight and control received the highest global priority, followed by personal information protection, design ethics, intellectual property rights protection, and limits of algorithmic autonomy. The interviews explain these priorities by showing that experts framed trustworthy AI governance as a problem of controllability, responsibility allocation, traceable data use, rights protection, and verifiable human intervention rather than model performance alone. The study contributes by defining pre-design governance as a bounded initial consideration-stage decision structure, combining AHP-based priority evidence with qualitative justification logic, and proposing a preliminary governance package of decision points, minimum evidence artefacts, and illustrative operational check criteria. The package is not presented as a validated legal compliance model; instead, it provides an expert-informed translation pathway for future organisational, sector-specific, and empirical validation.

Keywords: Human-Centred AI (HCAI); EU AI Act; risk-based regulation; AHP; expert interviews; data accountability; AI governance

Check for updates

Academic Editor: Haris Alibašić

Received: 6 June 2026
Revised: 9 July 2026
Accepted: 13 July 2026
Published: 17 July 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

As artificial intelligence (AI) systems become increasingly embedded in critical infrastructures, public services, healthcare, education, finance, and organisational decision-making processes, governance challenges have grown alongside advances in technological capabilities [1]. Recent debates emphasise that trustworthy AI depends not only on model performance but also on governance mechanisms that ensure accountability, transparency, auditability, risk management, and meaningful human oversight throughout the AI lifecycle [2]. Consequently, organisations face growing pressure to translate broad ethical

principles and regulatory requirements into operational governance structures capable of supporting responsible AI deployment [3]. Despite this need, existing discussions frequently focus on compliance obligations, ethical principles, or technical safeguards in isolation, providing limited guidance on how governance decisions should be structured before AI systems are designed and implemented. A critical challenge therefore remains: how to operationalise AI governance as an early-stage decision-making process rather than as a post-deployment compliance activity.

This study addresses this challenge by examining whether Human-Centred AI (HCAI) can function as a governance and design mechanism during the initial decision-making stage of risk classification and responsibility allocation under the EU AI Act. HCAI proposes that AI should be designed to complement human capabilities under human judgement and responsibility rather than replace humans as autonomous decision-makers [4,5]. While AI extends human cognition and decision-making capabilities through learning from large-scale data [6], concerns regarding algorithmic bias, data concentration, privacy violations, and accountability gaps continue to threaten public trust and social legitimacy [7,8]. In response to these challenges, the European Union established the EU AI Act, which classifies AI systems according to risk levels and imposes obligations related to safety, transparency, accountability, and fundamental-rights protection [9]. Although existing HCAI and AI governance frameworks provide important ethical principles, they often remain at the level of normative recommendations and offer limited guidance regarding how governance responsibilities should be structured during the earliest planning and design stages [10,11]. HCAI distinguishes itself through a design-oriented perspective that seeks to embed human judgment, oversight, and accountability into decision-making processes from the outset [12]. Accordingly, this study conceptualises early-stage AI trust not as mere adherence to ethical principles, but as a governance structure in which the rationale, evidence, responsibilities, and decisions associated with risk classification and accountability can be documented, traced, and verified. However, empirical approaches for operationalising these principles within risk-based regulatory environments such as the EU AI Act remain underdeveloped [10,11,13,14].

Previous studies have generally addressed HCAI principles, the risk-based regulation of the EU AI Act, and decision-making analytical frameworks such as AHP independently. This study integrates these areas by examining human oversight and control and algorithmic autonomy as complementary governance concerns. It combines AHP with think-aloud interviews to present both governance priorities and their justifying rationale. Table 1 compares the approaches of prior research and the present study. The contributions of this study are threefold: (1) proposing an integrated pre-design governance mechanism, (2) deriving core governance priorities based on expert evaluation, and (3) suggesting policy and field application directions, including documentation, verification, auditability, and human oversight. The linkage between HCAI principles and the EU AI Act should therefore be understood as selective governance mapping rather than as a full article-by-article legal analysis.

Figure 1 illustrates that while previous studies addressed these areas separately, the present research integrates them through a mixed-methods design combining qualitative expert insights and quantitative AHP weights.

The linkage between HCAI principles and the EU AI Act is treated in this study as a selective governance mapping rather than as a full legal correspondence or doctrinal interpretation. To avoid presenting the EU AI Act as a complete compliance framework, this study uses selected obligations under Regulation (EU) 2024/1689 only as regulatory anchors for conceptual and methodological positioning [9]. Table 2 summarises how the main HCAI governance concerns examined in this study are translated into pre-design

governance tasks with reference to selected obligations related to risk management, data governance, technical documentation, record-keeping, transparency, human oversight, and post-market monitoring.

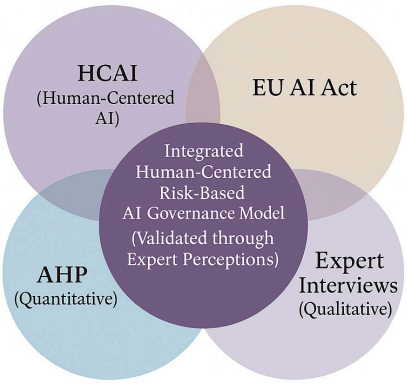


Figure 1. Comparative visualization of existing studies and present research.

Table 1. Comparison of Existing Studies and the Present Study.

Category	Existing Research (Example)	This Study
Research Focus	HCAI Ethics and Conceptual Discussions	Empirical Quantification of HCAI Regulatory Priorities
Regulatory Framework	Analysis of the EU AI Act from a Legal and Policy Perspective	Selective Governance Mapping of HCAI Principles to Relevant EU AI Act Provisions
Methodology	Qualitative and Conceptual Approach	Mixed Methods (Expert Interviews + AHP Weighting)
Data Sources	Focus on literature and policy materials	Expert Panel (N = 28) + Interviews (N = 15)
Deliverable Format	Conceptual or ethical recommendations	Quantified governance priorities + policy models

This study applies HCAI principles within the context of data capitalism and the risk-based regulatory framework of the EU AI Act. It proposes a requirements structure and a pre-design governance pathway intended to support decision-making during the initial consideration phase preceding design and development. To achieve this, it derives quantitative priorities (AHP) and qualitative justification logic (interviews) through expert judgment.

MRQ: How can the risk-based approach of the EU AI Act be structured for application to the initial consideration phase of AI systems from an HCAI perspective?

RQ1: How do experts perceive the relative importance of risk-based regulatory elements in the EU AI Act from an HCAI perspective?

RQ2: How does experts' social and policy awareness in a data capitalism environment influence the prioritization and interpretation of AI governance elements?

Accordingly, this study treats the EU AI Act not as complete article-by-article legal mapping, but as a selective regulatory reference for examining how risk-based requirements can inform pre-design governance decisions from an HCAI perspective. In this sense, risk-based regulation is considered not only as an ex-post compliance tool, but also as an operational governance mechanism for structuring accountability, oversight, documentation, and risk-management decisions during the earliest stages of AI system development. The

analysis is therefore not presented as a fully validated compliance model for all AI systems or regulatory contexts, but as an exploratory, expert-informed governance framework.

Table 2. Selective Governance Mapping between HCAI Concerns and EU AI Act Requirements.

HCAI Governance Concern Examined in This Study	Selected EU AI Act Obligation Used as A Regulatory Anchor	How this Study Translates into Pre-Design Governance
Risk classification and risk management	Risk-management system for high-risk AI systems under Regulation (EU) 2024/1689, Article 9	The study translates this obligation into pre-design risk classification, risk registration, and early identification of intended purpose, autonomy level, affected users, and potential fundamental-rights impacts.
Data responsibility and data accountability	Data governance and data quality requirements under Regulation (EU) 2024/1689, Article 10	The study translates this obligation into a data-accountability evidence chain covering data sources, rights, quality, bias, lineage, access control, and update/change history.
Documentation, logging, and auditability	Technical documentation and record-keeping requirements under Regulation (EU) 2024/1689, Articles 11–12	The study translates these obligations into minimum evidence artefacts, log policies, traceability records, and audit trails from the initial consideration phase onward.
Transparency and deployer-facing information	Transparency and provision of information to deployers under Regulation (EU) 2024/1689, Article 13	The study translates this obligation into explanation policies, limitation notices, appropriate-use information, and objection or dispute-response pathways.
Human oversight and intervention authority	Human oversight requirements under Regulation (EU) 2024/1689, Article 14	The study translates this obligation into assigned oversight roles, intervention thresholds, escalation rules, override authority, and termination or rollback conditions before implementation.
Corrective action and post-market feedback	Corrective action and post-market monitoring obligations under Regulation (EU) 2024/1689, Articles 20 and 72	The study uses these obligations only as forward-looking implications for pre-design planning, including monitoring indicators, incident response procedures, feedback loops, and redesign triggers.

Note. This table provides a selective governance mapping for conceptual and methodological positioning. It does not claim to provide a complete legal analysis, doctrinal interpretation, article-by-article compliance checklist, or validated compliance model. The EU AI Act is used only to anchor selected governance concerns examined in this study under Regulation (EU) 2024/1689.

Its contribution lies in identifying priority governance elements, explaining the reasoning behind these priorities, and proposing a preliminary set of actionable artefacts that can support future validation in organizational and sector-specific settings.

At the analytical level, the emphasis is not on the AHP ranking alone or the interview findings in isolation, but on translating expert priorities and think-aloud justification logic into an operational pre-design governance structure under the risk-based orientation of the EU AI Act. This is elaborated through three analytical steps: (1) framing pre-design governance as a bounded construct for the initial consideration phase, distinct from broader responsible AI design and compliance-oriented approaches [5,9,12,15,16] (2) identifies HCAI governance priorities through AHP and explains their rationale through think-aloud interviews [17–19] and (3) organizing these findings into a preliminary implementation package linking risk classification, human oversight, accountability, data responsibility, autonomy boundaries, documentation, monitoring, and auditability before detailed system design begins.

2. Literature Review

This section is guided by the premise that human control, accountability, and fundamental rights protection are key ethical principles of trustworthy AI governance and should be considered early-stage governance conditions before AI implementation [5,9,20]. To clarify the classification logic used in the literature review, this section reviews and synthesises four literature streams: HCAI design objectives, AI risk management and governance, data management measures, and algorithm regulation. Rather than summarising these areas separately, the review identifies the main points of consensus, unresolved issues, and remaining research gaps in each stream, thereby providing the conceptual basis for deriving the literature-based analytical categories used in the subsequent AHP and think-aloud interview analysis [17,18,21].

The term "pre-design governance" is used here to refer to the decision window before detailed AI system design, procurement, or development, in which organisations determine whether an AI use case should proceed, under what risk and autonomy conditions, and with which governance safeguards, including human oversight authority, data-accountability controls, documentation duties, and auditability requirements. It is narrower than responsible AI design, which covers broader design principles and human-AI interaction concerns across the lifecycle [5,12,15,16] and it is not equivalent to a full legal compliance programme. Rather, it functions as a decision-structuring lens that translates HCAI principles and selected risk-based requirements of the EU AI Act into actionable initial-stage governance judgements [9].

Previous research has addressed ethics, technology, law, and policy in a fragmented manner [4,22]. Ethics has focused on values and norms, technology on algorithm and data performance, and regulation on responsibility and risk classification [16,23]. However, this approach has limitations in comprehensively explaining the how human-centred values are recognised and problematised throughout the entire process, including pre-design stages of AI [12,24], and the demand for 'pre-emptive human control, accountability, and rights protection' has been discussed primarily as requirements/mechanisms for the design and operation phases [15,25]. Consequently, analysis of what is prioritised in the initial consideration phase has been relatively scarce [10,11]. Therefore, this literature review does not treat these elements as separate categories; instead, it adopts an analytical perspective that examines how responsibility and control are perceived and structured in the pre-design phase [12].

2.1. Human-Centred AI Design Objectives

HCAI adopts an ethical principle that AI should be designed not as an autonomous agent replacing humans, but to complement human judgment in a controllable and accountable manner [12,26,27]. This principle is justified because biased or erroneous data, if learned as rules, can lead to entrenched misjudgments, and biased outputs without human oversight can result in rights violations [23,28–30]. Therefore, the design goal of HCAI focuses on conditions that maintain human judgment, control, and responsibility during AI utilization, rather than performance improvement itself [4,5,20].

Meanwhile, the trend toward technology-centric development and user experience issues have expanded discussions on HCI and human–AI interaction [23,25,31], interaction optimization alone cannot sufficiently guarantee the protection of responsibilities and rights (especially regarding data processing, surveillance, and privacy) [12,24]. Subsequently, HCAI has emerged as an approach in which human-centredness, technological reliability, and Human–AI interaction are pursued as core values in AI design and are applied not only in design but also across the broader processes of development and operation [12,24]. The timing of HCAI design has expanded to become a critical factor in AI design objectives.

However, how human control and oversight should be implemented in AI systems remains contentious. The autonomy of AI algorithms can lead to rights violations through biased outputs without human oversight [23,28,29], making the balance between protecting legal interests and enhancing application value crucial [9]. This suggests that HCI-oriented approaches focused mainly on AI use and application value may be insufficient, thereby indicating the need for a shift in perspective toward HCAI [10].

A shared premise in the HCAI literature is that AI should support human judgement while maintaining controllability and responsibility rather than replace humans as autonomous decision-makers [4,12,20]. However, an unresolved issue remains regarding how these human-centred values should be balanced with autonomy, efficiency, and application value in early AI adoption decisions. Existing studies still provide limited guidance on how HCAI design objectives can be translated into concrete pre-design governance priorities, such as risk classification, responsibility allocation, and human oversight criteria [10,11,13,14].

2.2. Risk Management and AI Governance

This section presents a theoretical foundation to explain how risks and responsibilities are perceived and problematized during the pre-design phase providing the conceptual basis for why the expert interviews and AHP analysis reported in Section 4 are designed around 'human control' and 'responsibility'. To this end, this section adopts two ethical principles: (1) the need for effective human control and (2) the allocation and reinforcement of responsibility [5,11]. These principles are justified by the need to structurally ensure controllability and the attribution of responsibility from the pre-design stage to prevent the loss of control and responsibility gaps that highly autonomous AI could cause. Risk is defined as the possibility or consequence of uncertainty that could hinder goal achievement [32]. In such complex risk environments, collaborative governance involving government–industry–civil society is particularly crucial for responsibility sharing and trust building [33,34]. While the EU AI Act imposes documentation and record-keeping obligations on high-risk AI [9], changes to algorithms and data during operation are difficult to capture sufficiently through document-centric controls alone [12,14], and risks of shifting responsibility exist if rapid technological change outpaces regulatory revisions [35]. Therefore, sustainable risk management requires the interconnection of pre-emptive risk awareness, governance, technical documentation, and autonomy management, supported by the coordinated design of human control and technological autonomy [5,11]. The risk-management and AI-governance literature broadly indicates that trustworthy AI requires risk awareness, documentation, accountability, and organisational responsibility across the AI lifecycle [5,9,11,36]. Nevertheless, an unresolved issue remains whether documentation- and compliance-oriented controls are sufficient for managing dynamic AI risks that may emerge during development and operation [12,14]. A remaining gap is that existing studies provide limited explanation of how responsibility, authority, approval, and intervention criteria should be allocated at the initial consideration stage before AI development proceeds.

2.3. Data Management Measures

This section adopts data reliability, intellectual property rights protection, and personal information protection as core principles (sub-factors) to derive and interpret the perception factors of 'data management measures' from expert interviews and AHP as theoretical grounds [9,12,37]. This stems from the recognition that AI development has amplified issues associated with internet proliferation and data accumulation [6], ethical concerns, the right to self-determination over personal information, and the risk of copy-

right infringement [9,22]. The proliferation of AI-generated art (e.g., Obvious, The Next Rembrandt, Ai-Da) highlights concerns about intellectual property and privacy violations during data expansion, alongside limitations in quality control [38–41]. Furthermore, data capitalization transforms personal information into a commercial resource, reinforcing the mechanisms of surveillance capitalism [7,42] and expanding the predictive and formative power of platform companies [8]. However, it is also noted that such critiques risk overlooking the economic and social benefits of the technology [43]. Recent research warns that the expansion of surveillance capitalism threatens data sovereignty and privacy, posing raising concerns about data colonialism [44–46]. The data-management literature generally indicates that personal information protection, intellectual property rights, data reliability, and bias management are core conditions for trustworthy AI governance [7,9,22,42]. In addition, documentation-oriented approaches such as datasheets, model cards, and data cards show that data and model-related judgments can be made explicit through structured evidence and reporting formats [47–49]. However, an unresolved tension remains between the economic and operational value of data-driven AI and the risks of privacy infringement, copyright violation, bias, and data extraction [7,42,43]. The remaining gap is that data-related responsibilities are often discussed as separate obligations, rather than as an integrated data-accountability structure that can be prioritised together with human oversight and risk-based regulation during the initial consideration phase.

2.4. Algorithm Regulation

This section organizes 'regulatory standards, regulatory methods, and autonomy limits' as sub-factors of 'algorithm regulation' to examine how experts perceive the scope of algorithmic control and responsibility during the initial exploratory phase of AI system adoption, using interviews and AHP [12]. Furthermore, it adopts transparency, accountability, and the protection of fundamental rights as core ethical principles, grounded in the concern that expanding algorithmic autonomy could weaken control and responsibility [9,50]. AI is a system that automatically processes information through algorithms [23], and generative AI tends to increase its autonomy and performance through large-scale data learning [51], while also expanding its generalization capabilities and scope of application through big data utilization [52]. GANs also exhibit a structure where results are iteratively improved through generator–discriminator interaction [53]. While securing high-quality datasets is discussed as key to success [54,55], expanding data collection increases the risk of infringing individual rights. Consequently, the GDPR regulates transparency, while the AI Act requires fundamental rights safeguards for high-risk AI and restricts the use of remote biometric identification for law enforcement purposes, risk assessments of natural persons, and biometric profiling [9,56]. Increased autonomy makes ensuring transparency and accountability difficult, while opaque decision-making heightens risks of infringing individual autonomy and concealing bias and discrimination [57,58]. Accordingly, algorithm regulation is being increasingly implemented across domains ranging from consumer-rights protection to AI business autonomy [59]. Algorithm-regulation studies provide a basis for recognising transparency, accountability, and explainability as safeguards for limiting opaque or unaccountable automation [50,56,57]. Within the EU AI Act, these concerns are further institutionalised through risk classification, transparency obligations, human oversight, and requirements for technical documentation and record keeping [9]. However, a central unresolved issue concerns the degree of autonomy that should be permitted, under what risk conditions, and through which human or organisational control mechanisms [23,58]. The remaining gap is that algorithmic control is often treated as a separate legal or technical safeguard, without fully explaining how autonomy limits should

be coordinated with human oversight, data accountability, and organisational responsibility at the pre-design stage.

2.5. Synthesis and Analytical Framework

Taken together, the preceding review indicates that HCAI design objectives, AI risk management and governance, data management measures, and algorithm regulation should not be treated as separate or merely parallel domains. Although prior research has established important principles for aligning AI systems with human-centred values, risk-based governance, data accountability, and algorithmic control, these discussions are often developed as separate ethical, technical, legal, or organisational agendas [5,9,12,22,36]. During the initial consideration phase, these streams converge into concrete organisational judgements about who may authorise AI use, what degree of algorithmic autonomy is acceptable, which data sources and rights conditions are permissible, what evidence must be documented, and when human intervention or rollback is required. This convergence identifies the main gap addressed in this review and provides the conceptual basis for deriving the four literature-based analytical axes used in the subsequent AHP and think-aloud interview analysis [17,18,21].

3. Materials and Methods

3.1. Study Design: Literature

This study employed an exploratory mixed-methods design consisting of three sequential stages preceded by a literature-based framework-development phase [21]. First, the HCAI framework, AI governance literature, and relevant provisions of the EU Artificial Intelligence Act were reviewed to derive and operationalise evaluation factors related to human-centred governance, risk management, data accountability, and algorithmic control. These factors were then structured into an AHP hierarchy model comprising a goal, criteria, and subcriteria [9,22].

Second, an expert panel assessed the relative importance of the identified factors through an AHP pairwise-comparison survey using Saaty's 1–9 scale [17,60]. Individual judgments were aggregated using the geometric mean to derive group-level priorities [61], and consistency was assessed using the CR $\leq$ 0.10 threshold [17,60,62].

Third, semi-structured think-aloud interviews were conducted after the AHP analysis to explain and contextualize the reasoning underlying the quantitative priority structure [18,21]. Interview transcripts were coded and interpreted using thematic analysis [19].

The qualitative phase was not intended to statistically validate the AHP results, but to interpret and contextualize the AHP-derived priority judgments and the reasoning and contextual conditions underlying the AHP-derived priority judgments. Accordingly, the study is exploratory and interpretive in purpose while following a sequential explanatory mixed-methods structure in its empirical procedure [21]. The findings are not intended to represent the general population but to illuminate how experienced stakeholders structure early-stage AI governance judgments [63].

3.2. AHP Modeling and Analysis

3.2.1. Hierarchical Structure and Indicators

The AHP model, as shown in Figure 2, was constructed through a three-step procedure. First, candidate factors were extracted from the four literature streams reviewed in Section 2: HCAI design objectives, risk management and AI governance, data management measures, and algorithm regulation. Second, these factors were operationalised as AHP criteria and subcriteria when they met three conditions: conceptual consistency with HCAI principles of human controllability, accountability, and rights protection; rel-

evance to the EU AI Act's risk-based governance logic; and suitability for assessment as pre-design governance judgements rather than as purely technical performance indicators [4,5,9,12,36].Third, the draft hierarchy was reviewed by experts to examine whether the selected factors covered the intended HCAI governance construct and whether any major domain had been omitted, thereby supporting the content validity and comprehensiveness of the hierarchy [21,63]. The selected factors were also checked for conceptual distinctiveness and suitability for AHP pairwise comparison in accordance with the hierarchical decomposition logic of AHP [17,62]. During hierarchy construction, candidate factors that fell outside the scope of early-stage HCAI governance were not retained as separate AHP elements. For example, model accuracy, computational efficiency, latency, technical reliability, sector-specific compliance items, and post-deployment monitoring tasks were considered important for AI system development, but were excluded or consolidated into broader governance categories when they were too technical, context-specific, or lifecycle-specific to support meaningful pairwise comparison within the AHP structure [17,62]. Accordingly, the AHP results should be interpreted as relative priorities among the governance elements included in the final hierarchy, not as a comparison between governance and all possible technical-performance criteria. This screening process resulted in a final structure consisting of one goal, four criteria, and fourteen subcriteria, which were subsequently used as the framework for pairwise comparison and priority analysis.

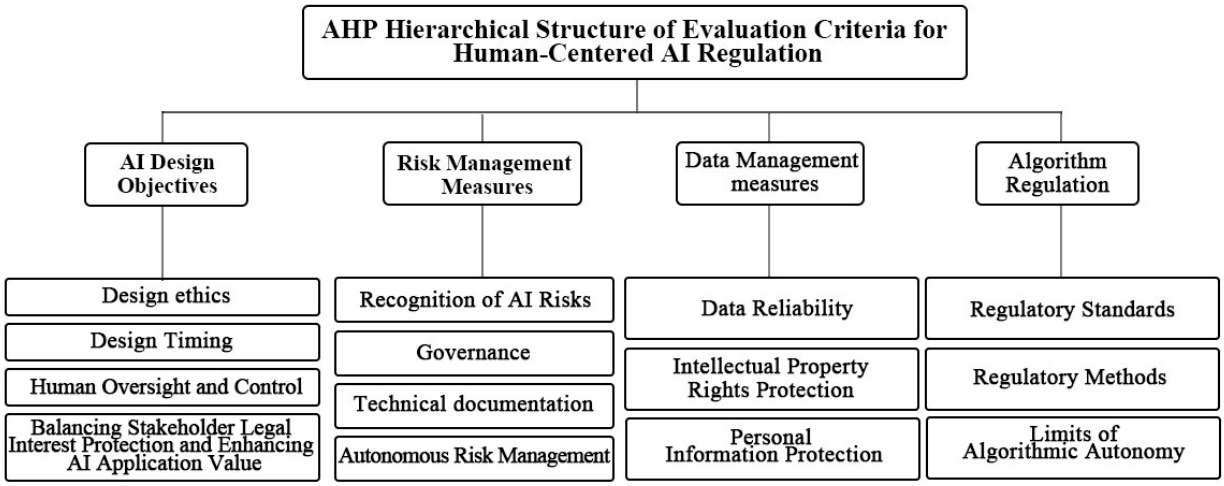


Figure 2. AHP Hierarchy Diagram.

Given this hierarchical structure, AHP was selected because the purpose of this study was to derive relative priorities among literature-based governance criteria within a clearly structured goal–criteria–subcriteria model. Compared with methods such as ANP, BWM, FUCOM, or DEMATEL, which are more suitable for modelling interdependencies, causal relationships, or alternative weighting procedures, AHP was appropriate for structuring expert judgements across upper- and lower-level criteria and for transparently calculating local and global priorities.

3.2.2. Sampling and Data Collection

This study assembled an expert panel through purposive sampling to examine role-based judgement structures in early-stage AI governance rather than to obtain a statistically representative sample [21,63]. AHP participants were recruited through professional networks, academic and expert communities, and referrals.

The AHP pairwise-comparison questionnaire, including the survey introduction, consent statement, Saaty 1–9 comparison scale, definitions of the evaluation criteria and

subcriteria, and pairwise-comparison items, is provided in Supplementary Material File S2, AHP Pairwise Comparison Questionnaire.

Eligible participants were adults with professional, research, policy-related, or practical experience relevant to AI governance or AI-related systems, including technology, education, law, design, media/content, public-sector practice, and industry applications. Participants were excluded if they lacked relevant AI-related experience, did not complete the AHP survey, or failed to meet the consistency requirement after one re-evaluation.

The final AHP panel comprised 28 experts (19 male, 9 female) from four role-based groups: artificial intelligence, education, law, and other AI-related fields, with an average career length of approximately 7.8 years. Expertise was assessed based on participants' self-reported role, years of experience, and relevance to AI governance or AI-related decision-making. Sample characteristics are presented in Table 3. This study is based on a Korea-based expert panel, and the results are applicable to domestic governance design and operational requirements. The EU AI Act was used as an analytical framework for structuring requirements, not as a prescriptive standard; additional validation is required for generalisation to other contexts.

Table 3. Demographic Characteristics of the Sample.

Category		Frequency (*n*)	Percentage (%)
Gender	Male	19	67.85
	Female	9	32.15
	Total	28	100.0
Work	Professional	12	42.86
	Staff	6	21.42
	Manager	10	35.72
	Total	28	100.0
Years of Experience	3 to 5 years	5	17.85
	5 to 7 years	7	25.00
	7 to 10 years	7	25.00
	10 years or more	9	32.14
	Total	28	100.0

Note: This table summarizes the general characteristics of the expert panel formed based on role criteria in the AI regulation decision-making process.

3.2.3. AHP Analysis Procedure and Consistency Verification

AHP analysis was conducted using Saaty's 1–9 pairwise comparison scale to derive the relative importance of the evaluation criteria [17]. Individual judgement matrices were aggregated into a group matrix using the element-wise geometric mean, and weights were derived using the principal eigenvector method [61]. Consistency was assessed using the consistency ratio (CR), with CR $\leq$ 0.10 applied as the acceptance threshold [17,62]. The AHP calculations were performed using Microsoft Excel for Microsoft 365 (Microsoft Corporation, Redmond, WA, USA). When any comparison matrix in an individual response exceeded this threshold, the participant was asked to complete one re-evaluation. In the initial consistency check, 32 responses were collected, of which six exceeded the CR threshold of 0.10 in at least one comparison matrix and were therefore subject to re-evaluation. After re-evaluation, two responses satisfied the CR $\leq$0.10 criterion and were retained, whereas four responses still exceeded the threshold and were excluded from the final analysis. Accordingly, the final AHP dataset consisted of 28 valid responses, all of which satisfied the CR $\leq$0.10 criterion. For the final aggregation, no statistical weighting was assigned to participants by professional background; professional diversity was used to capture role-based variation in expert judgement rather than to construct a representative population

sample. A comparison between the preliminary and final AHP results showed that the main priority ranking remained unchanged after screening, suggesting that the consistency-screening procedure did not materially alter the substantive interpretation of the priority structure. To improve reproducibility, the Supplementary Materials report the local weights, global weights, and consistency values for each comparison block.

For transparency, the analysis distinguishes between local weights, which indicate the priority of subcriteria within each upper-level criterion, and global weights, which are calculated by multiplying the upper-level criterion weight by the corresponding local subcriterion weight [62]. This distinction is essential because a factor may rank highly within its own category while receiving a lower overall priority when the weight of the upper-level category is considered. To further examine the robustness of the AHP priority structure, a one-at-a-time sensitivity check was conducted by varying each top-level criterion weight by ±10% and normalising the four top-level weights to sum to one. The adjusted top-level weights were then multiplied by the original local subcriterion weights to recalculate the global weights. This procedure was used as an interpretive robustness check rather than as a statistical significance test. The detailed sensitivity-check results are reported in Supplementary Table S8.

3.3. Interviews (Think-Aloud Protocol) Process and Analysis

3.3.1. Participants and Recruitment

For the qualitative analysis of this study, think-aloud-based in-depth interviews were conducted with 15 experts (N = 15). The 15 interview participants were selected from the 28 valid AHP respondents based on their agreement to participate in the follow-up qualitative phase, while considering role diversity across AI design, development, operation, oversight, education, law/policy, and applied AI or user-facing contexts. This sequential design allowed the qualitative phase to examine the reasoning underlying a subset of the quantitative AHP judgements. The interviews were used to interpret and contextualise selected AHP respondents' priority judgements, rather than to statistically validate the AHP results.

Participants were purposively selected to capture role-based judgement structures across AI governance decision-making contexts for the same regulatory agenda, rather than to achieve statistical or population representativeness [21,63]. To mitigate participant bias, role categories were defined broadly. General participant characteristics are presented in Table 4. Additional interview-level information, within limits that do not compromise participant anonymity, is provided in Supplementary Table S9.

Table 4. Summary Characteristics of Interview Participants.

Characteristic	Subcategory	Frequency (*n*)
Expert Role	Artificial Intelligence Expert	5
	AI-Related Expert	5
	Education Expert	3
	Legal and Policy Experts	2
Years of Experience	6–10 years	7
	11–20 years	7
	30 years or more	1
Total		15

3.3.2. Procedure

This study employed semi-structured interviews combined with think-aloud protocols to capture experts' cognitive judgements and interpretive contexts in real-time during the AI regulatory agenda evaluation process [18,21]. The interviews were structured in three analytical stages: (1) initial perception, (2) basis for judgement, and (3) post-guidance reflection after the EU AI Act-related regulatory framework was provided. To reduce potential priming effects, the analysis distinguished participants' initial perceptions and judgement rationales from their post-guidance reflections after the regulatory framework was provided. The post-guidance responses were treated as reflective interpretations shaped by the provided regulatory context, rather than as spontaneous baseline judgements. The EU AI Act-related framework was introduced only after participants had first provided their initial perceptions and judgement rationales. Participants were guided to freely verbalise their thought processes rather than provide definitive answers [18]. This approach complementarily collected judgment logic and contextual factors difficult to capture solely through AHP. The interview questions are summarised in Supplementary Table S11, and the full interview protocol and detailed topic-specific questions are provided in Supplementary Material File S3, "Expert Think-Aloud Interview." The entire procedure is summarised in Figure 3.

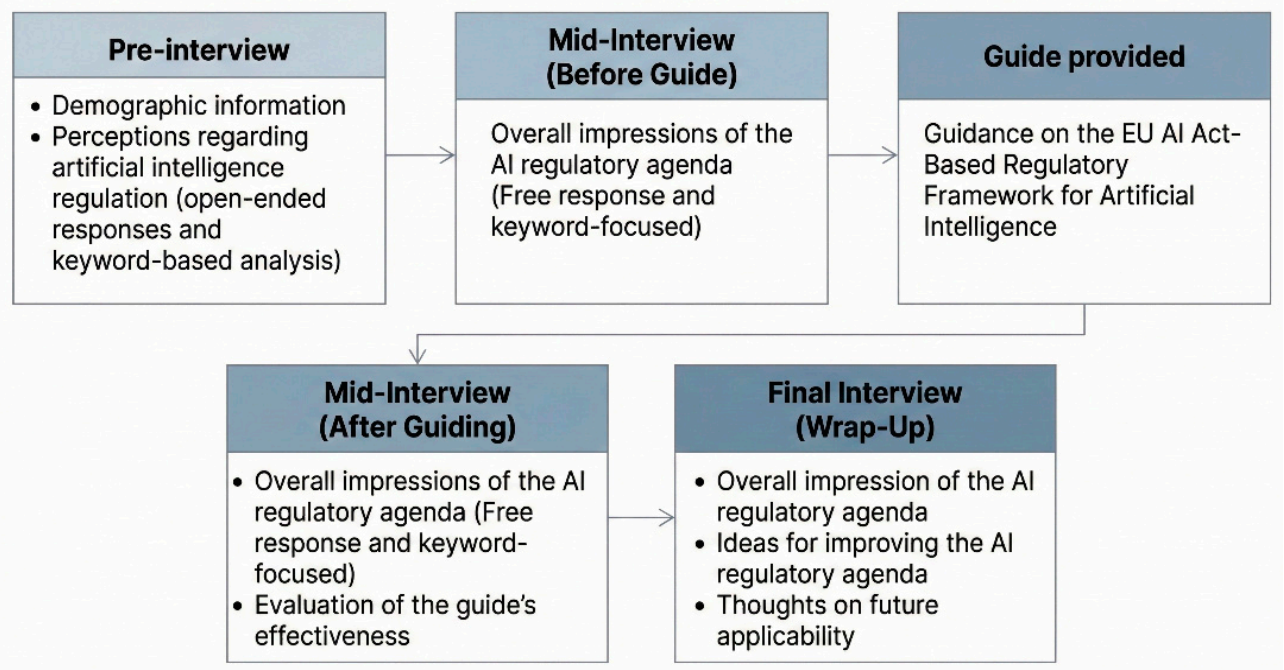


Figure 3. Research Flowchart.

3.3.3. Coding and Thematic Analysis

Interview data were analysed using Braun and Clarke's six-step thematic analysis [19]: familiarisation, coding, theme development, review, definition, and reporting. The transcripts were first segmented into meaning units, including full sentences and shorter utterance fragments. These meaning units were then assigned initial codes capturing regulatory criteria, judgement rationales, contextual conditions, and institutional or technical constraints. Related initial codes were grouped into sub-codes and subthemes, which were then organised into four higher-level themes corresponding to the AHP criteria: AI design objectives, risk management measures, data management measures, and algorithm regulation. Coding was conducted using Reframer (Optimal Workshop, web-based qualitative research software; official website: https://www.optimalworkshop.com/tag/reframer (ac-

cessed on 12 July 2026)) and iteratively refined through researcher discussion. To enhance reproducibility, two researchers independently coded approximately 20% of the transcripts (n = 3) and resolved disagreements through consensus coding. Saturation was defined as the point at which no new codes or subthemes emerged across two consecutive interviews. The resulting themes were interpreted alongside the AHP weights and priority rankings to explain the reasoning and contextual conditions underlying the quantitative priority structure, rather than to statistically validate the AHP results [19].

3.4. Research Ethics, Consent, and Data Handling

This study was conducted in accordance with institutional ethical requirements for research involving human participants. Ethical approval was obtained from the relevant institutional review board before data collection. All participants provided informed consent; interviews were recorded only with consent, and the data were pseudonymised, securely stored, and managed according to the approved research protocol. The data will be retained for five years after study completion and then destroyed in accordance with the approved data-management plan.

4. Results

4.1. AHP Results

The AHP results were interpreted at two levels: the upper-criterion level and the global subcriterion level. At the upper-criterion level, AI design objectives received the highest weight (0.311), followed by risk management measures (0.286), data management measures (0.227), and algorithm regulation (0.177). At the global subcriterion level, human oversight and control received the highest overall priority (0.1113), followed by personal information protection (0.0899), design ethics (0.0889), intellectual property rights protection (0.0849), and limits of algorithmic autonomy (0.0843). Limits of algorithmic autonomy ranked first within the algorithm-regulation category and fifth in the overall global-weight hierarchy (Table 5). These findings indicate the relative priority of governance factors within the selected HCAI pre-design hierarchy. They do not imply that human oversight, accountability, or rights protection are generally more important than technical performance, reliability, or accuracy in AI system development. Rather, within the governance-focused construct examined here, experts assigned the strongest global priority to human oversight and control, while also giving high priority to personal information protection, design ethics, intellectual property rights protection, and limits of algorithmic autonomy. This pattern suggests that experts viewed early-stage AI governance as a problem of controllability, accountable data use, and bounded autonomy before detailed system design begins [5,12,17]. The upper-level priorities, key local subcriteria, and consistency values are summarized in Table 5. The complete weights and CI/CR values are reported in the Supplementary Materials (Supplementary Tables S5 and S6).

Table 5. AHP Analysis Results: Top/Sub-Priority Outcomes.

Top Criteria	Importance	Priority	Key Sub-Factors Within Criteria (Top-2)	Sub-Criterion Comparison Matrix Consistency (CI/CR)
AI Design Objectives	0.311	1	Human Oversight and Control (0.358), Design ethics (0.286)	CI = 0.00/CR = 0.000 (n = 4, RI = 0.90)
Risk Management Measures	0.286	2	AI Risk Awareness (0.290), Autonomous Risk Management (0.266)	CI = 0.02/CR = 0.022 (n = 4, RI = 0.90)

Table 5. *Cont.*

Top Criteria	Importance	Priority	Key Sub-Factors Within Criteria (Top-2)	Sub-Criterion Comparison Matrix Consistency (CI/CR)
Data Management Measures	0.227	3	Personal Information Protection (0.396), Intellectual Property Rights Protection (0.374)	CI = 0.00/CR = 0.000 (n = 3, RI = 0.58)
Algorithm Regulation	0.177	4	Limits of Algorithmic Autonomy (0.476), Regulatory Methods (0.291)	CI = 0.02/CR = 0.034 (n = 3, RI = 0.58)

The consistency of the top-level comparison matrix (4 × 4) meets the criteria (CR $\leq$ 0.10) with CI = 0.01 and CR = 0.011 (RI = 0.90). The subsequent interview analysis was used to interpret the reasoning and contextual conditions underlying these priorities.

4.1.1. Final AHP Analysis Results

As shown in the AHP results in Table 6, experts assigned relatively high priority to governance factors related to control, accountability, and regulatory alignment within the set of criteria included in the AHP model [64,65]. This reflects a judgement that early-stage AI governance should establish acceptable starting conditions, including control, accountability, and regulatory alignment, before later-stage optimisation decisions are made. The significance of the results lies not in identifying universally superior factors, but in presenting the hierarchical structure of expert judgements among the governance factors included in the AHP model [64]. In particular, the relatively high ranking of personal information protection and intellectual property rights protection indicates that experts recognised privacy, data rights, and legal-security concerns as important early-stage governance issues, consistent with discussions that privacy and data-rights risks may intensify under surveillance capitalism [45].

Table 6. Final AHP Analysis Results for All Respondents.

Main Criterion	Lower Factor	Local Weight	Upper Weight	Global Weight	Rank
AI Design Objectives	Design ethics	0.286	0.311	0.0889	3
	Design Timing	0.152	0.311	0.0473	13
	Human Oversight and Control	0.358	0.311	0.1113	1
	Balancing Stakeholder Legal Interest Protection and Enhancing AI Application Value	0.204	0.311	0.0634	9
Risk Management Measures	Recognition of AI Risks	0.290	0.286	0.0829	6
	Governance	0.203	0.286	0.0581	10
	Technical documentation	0.242	0.286	0.0692	8
	Autonomous Risk Management	0.266	0.286	0.0761	7

Table 6. *Cont.*

Main Criterion	Lower Factor	Local Weight	Upper Weight	Global Weight	Rank
Data management measures	Data Reliability	0.230	0.227	0.0522	11
	Intellectual Property Rights Protection	0.374	0.227	0.0849	4
	Personal Information Protection	0.396	0.227	0.0899	2
Algorithm Regulation	Regulatory Standards	0.234	0.177	0.0414	14
	Regulatory Methods	0.291	0.177	0.0515	12
	Limits of Algorithmic Autonomy	0.476	0.177	0.0843	5

The upper-level criteria matrix (4 × 4) yielded CI = 0.01 and CR = 0.011. The CR values of the lower-level block matrices ranged from 0.000 to 0.034, all satisfying the CR ≤ 0.10 criterion. The one-at-a-time sensitivity check showed that Human Oversight and Control remained the highest-ranked global subcriterion under all ±10% top-level weight adjustment scenarios. This suggests that the top-ranked priority was stable under moderate changes in the upper-level criterion weights. The detailed sensitivity-check results are provided in Supplementary Table S8.

4.1.2. Global Weight (Upper × Local) as an Interpretive Cross-Check

To supplement the interpretation of the AHP results, we calculated a composite score (global weight) as the product of the upper- and lower-level weights (upper × lower) (Supplementary Table S7). This value is provided solely as an interpretive aid and should not be treated as a substitute for consistency verification based on the consistency ratio (CR) [62].

4.2. Interview Analysis Results (Think-Aloud Protocol)

4.2.1. Interpretive Analysis of Interview Results

This section uses the think-aloud interview data to interpret and contextualise the AHP priorities reported in Section 4.1 by examining the cognitive judgements and interpretive contexts underlying participants' evaluations [18]. The analysis did not focus on the frequency of particular expressions or on a simple listing of interview topics. Instead, thematic coding was used to identify recurring reasoning patterns through which participants explained, justified, and contextualised their priority judgements [19].

The analysis focused particularly on how experts understood the relationships among human oversight, risk management, data accountability, and algorithm regulation during the initial consideration phase. In doing so, the interview findings provided contextual insight into why selected AHP respondents attached higher or lower priority to certain governance factors.

Table 7 summarises sample responses, key judgement patterns, and their interpretive links to AHP priorities. The excerpts are presented not as exhaustive evidence or frequency counts, but as illustrative examples selected to show how participants connected human-centred design, risk management, data accountability, and algorithm regulation to early-stage governance judgements. In particular, the examples highlight dominant judgement patterns shared across several roles, boundary conditions under which a governance factor became salient, and cautionary or divergent views that limited overgeneralisation. Additional response examples, key tags, sub-codes, higher-level themes, participant

patterns, and interpretive links to the corresponding AHP subcriteria are provided in Supplementary Table S10.

Table 7. Sample Responses, Key Judgement Patterns, and Interpretive Links to AHP Priorities.

Top Criteria	Core Judgment Patterns	Representative Anonymised Excerpt	Interpretive Link to AHP Priority
AI Design Objectives	Proactively define the purpose and boundaries of AI utilization based on human control and responsibility.	"No matter how advanced AI becomes, I believe it is dangerous for it to replace human judgment or responsibility. Ultimately, I think we need human control to feel secure."	Helps interpret why human judgement, responsibility, and control were central to the high priority of Human Oversight and Control and Design Ethics.
AI Design Objectives	Determine human intervention points during the initial design stage rather than after implementation.	"When human intervention is needed isn't something to ponder later; it must be determined at the initial design stage."	Supports the interpretation that human oversight should be designed as an early-stage governance condition.
Risk Management Measures	Requires a documented, proceduralized prevention and ongoing management system based on stage-specific risks.	"I believe those operating artificial intelligence should not simply leave risks unaddressed but should proactively establish manuals or management procedures to conduct regular inspections."	Helps explain why risk management was interpreted as a continuous organizational process rather than a one-time compliance check.
Data Management Measures	Recognize data source, data trust, personal information, and intellectual property rights as core conditions for building trust.	"Can we trust data obtained through a commercial entity like Google?"	Supports the interpretation of data accountability as a prerequisite for trustworthy AI governance.
Algorithm Regulation	Establish standards and enforcement mechanisms aligned with risk levels.	"Since each algorithm's situation differs, I think simple cases might be fine with just proper disclosure, but cases significantly impacting people should be managed with greater caution."	Illustrates the risk-based interpretation of algorithm regulation and supports the need to apply differentiated controls according to impact level.
Algorithm Regulation	Set limits on autonomy when the scope of AI use becomes difficult to control.	"Artificial intelligence offers infinite possibilities, but its scope of use is so vast that it can feel overwhelming."	Helps explain why Limits of Algorithmic Autonomy ranked first within the algorithm-regulation category.

Note: Representative excerpts were anonymised by removing institutional names, project-specific references, and other potentially identifying contextual information where necessary. Additional response examples, corresponding AHP subcriteria, key tags and sub-codes, and participant identifiers are provided in the revised Supplementary Table S10.

4.2.2. Key Interpretive Findings from Expert Interviews

- AI Design objectives

In relation to AI design objectives, the think-aloud interviews qualitatively contextualized how participants interpreted AHP-related factors, grounded in the principles of human-centredness, accountability, and transparency. Experts acknowledged AI's utility and innovation while identifying unpredictability, accountability gaps, and risks of rights infringement as core judgment factors [66,67]. Unexpected outputs were understood as a dual phenomenon carrying both value and risk. One participant stated, "*No matter how advanced AI becomes, I believe it is dangerous for it to replace human judgment or responsibility. Ultimately, I think we need human control to feel secure.*" This excerpt illustrates that experts

interpreted AI design objectives not only as issues of usefulness or innovation, but also as conditions for maintaining human judgment, responsibility, and trust.

Consequently, experts assessed the value of AI design based on the incorporation of responsibility attribution and human control structures as early-stage governance conditions [22]. There was shared recognition that control should be proactively institutionalized in the initial design phase, not as an ex-post supplement during operation [9,68]. Another participant emphasized this timing by noting, "*When human intervention is needed isn't something to ponder later; it must be determined at the initial design stage*." Furthermore, human oversight was interpreted not as an ethical declaration but as a practical safeguard to mitigate potential harm in exceptional situations [9]. The interviews indicated that even highly beneficial AI systems may be difficult to accept if they threatened stakeholders' legal interests, such as personal data, intellectual property, or employment [45].

Therefore, the high priority given to human control and oversight, together with standards for protecting responsibility and rights, can be interpreted as reflecting a conditional view of AI acceptance: AI may be considered acceptable when uncertainty is managed through operational mechanisms such as control, responsibility attribution, and auditability, rather than being accepted unconditionally [22,69].

- Risk Management Measures

Participants recognized risks not merely as compliance checks but as complex dynamic risks that could amplify during operations, including rights violations, diminished information trust, data contamination, deepfakes, and copyright infringement [9,67,70]. One participant noted, "*I believe those operating artificial intelligence should not simply leave risks unaddressed but should proactively establish manuals or management procedures to conduct regular inspections.*" This excerpt illustrates how participants interpreted risk management not as a one-time compliance task, but as a continuous organisational process requiring regular inspection and operational responsibility. Participants also indicated that responses to these risks cannot be addressed through documentation alone. Instead, a "working system" must be designed from the outset, combining a governance structure—traceable technical documentation—with autonomous management during operation and a continuous inspection and improvement loop [3,9,71]. Another participant connected prevention to documentation by stating, "*I believe registering technical documentation that pre-screens datasets can prevent problems.*" This statement suggests that technical documentation was understood not merely as administrative evidence, but as a preventive mechanism for dataset review, traceability, and accountability. This recognition helps contextualise why risk management and rights-protection-related criteria were prioritised in the AHP results.

- Data Management Approaches

Participants perceived data management as a prerequisite for trustworthy and accountable AI governance, particularly in relation to data source, quality, bias, and update/change history. One participant asked, "*Can we trust data obtained through a commercial entity like Google?*" This excerpt illustrates how participants interpreted data management not merely as a technical issue of data quality, but as a governance condition involving data-source transparency, commercial data dependence, and trust. They suggested that trust from a commercial and regulatory perspective depends on the assurance of data source lineage, data quality, bias management, and update/change history, even when seemingly plausible outputs are produced [67,72,73]. Furthermore, copyright and personal information protection were understood not as secondary issues but as core conditions for maintaining the accountability and sustainability of AI use. This recognition helps contextualise why data reliability and rights-protection/accountability criteria received relatively high evaluations in the AHP results [3,67,74,75].

- Algorithm Regulation

Participants believed that algorithm regulation should be applied differentially based on risk levels (Article 6). They tended to recognise that declarations of principles alone are insufficient, necessitating enforceable regulations that combine risk management procedures (Article 9), verifiable documentation and data, and operational-stage monitoring, logging, and inspection systems (Article 26) [9,69]. In the recent context in which generative AI has caused significant real-world harm, such as privacy violations and voice phishing, it has become apparent that expanding autonomy does not necessarily lead to justified disclosure or use [76,77]. Consequently, participants viewed the issue not as one of rejecting autonomy itself, but as one of limiting the scope of autonomy and combining it with human oversight mechanisms [9]. Another participant described the perceived difficulty of uncontrolled autonomy by stating, "*Artificial intelligence offers infinite possibilities, but its scope of use is so vast that it can feel overwhelming*." This statement suggests that algorithmic autonomy was viewed as acceptable only under certain conditions, when bounded by risk classification, disclosure, monitoring, and human oversight.

Therefore, the relatively high evaluation of algorithm regulation standards/methods, together with human oversight and rights/data protection standards in the AHP results, can be interpreted as reflecting a conditional acceptance logic: autonomy may be accepted when it is combined with differentiated regulation, verifiability, and operational checks [9,69].

5. Discussion

This discussion synthesises the empirical findings into three connected governance implications: operationalising human oversight, structuring data accountability, and translating the findings into a preliminary pre-design governance package. These implications are limited to the governance criteria examined in this study and should be understood as an expert-informed governance translation requiring further organisational and sector-specific validation.

5.1. AI Regulation Regarding Human Oversight and Algorithmic Autonomy from an HCAI Perspective

The combined AHP and think-aloud findings suggest that the expert priorities reflected not only numerical rankings but also a broader governance logic. In the AHP analysis, "human oversight and control" emerged as the element requiring the highest priority review in the initial consideration phase, and similar reasoning patterns were identified in the interviews. Together, these findings suggest that algorithmic autonomy becomes acceptable only when it is bounded by clearly assigned human authority, intervention thresholds, and accountable decision procedures [23,78]. This interpretation is consistent with the HCAI principle that AI should complement, rather than replace, responsible human judgment.

From a procedural legitimacy perspective, human oversight should therefore be specified as a decision structure that defines who may intervene, under what authority, at which threshold, and with what record [79].

Article 14 of the EU AI Act supports this interpretation by requiring meaningful human oversight for high-risk AI systems, including the ability to understand, intervene in, interrupt, or disable system operation [9]. In this study, this requirement is interpreted as a pre-design governance task: assigning oversight roles, authority, intervention criteria, and record-keeping conditions before system development proceeds.

Figure 4 summarises this operational logic through five minimum elements—roles, authority, thresholds, actions, and records—which are later translated into the implementation package in Table 8.



Figure 4. Operationalization of Human oversight.

Table 8. Preliminary HCAI Pre-Design Governance Translation Derived from Literature Review, AHP Priorities, and Think-Aloud Interview Findings.

Governance Focus (Derived Priority)	Practical Action (Decision Point)	Minimum Deliverables (Evidence)	Illustrative Operational Check Criteria
1. Human oversight and Control	Establish intervention criteria, approval, and termination/rollback conditions	Summary of Oversight/Control Design (1 page)	Rollback P95 (minutes), Approval-omission incidents = 0
2. Scope of Automated Execution and Exceptions	Confirmation of automated/human intervention scope and exception criteria	Summary of Scope/Exception Criteria (1p)	Exceptions (per 1000), Human-intervention transition rate (%)
3. Explainability	Establishment of Explanation Level and Dispute Response Criteria	Explanation Policy (1p)	Explanation comprehension score (5-point scale; illustrative), Processing SLA (days)
4. Data Accountability	Establish principles for source, rights (IP), quality, bias, lineage/version	Data Summary Register	Source Identification Rate (%), Inter-Group Gap Δ
5. Risk·Approval Process	High-Risk Scenarios, Mitigation, and Responsibility Confirmation	Risk Register (Summary)	Mitigation Implementation Rate (%), Unmitigated high-risk scenarios = 0
6. Monitoring and Audit	Indicators, Alerts, Regular Audit Loop Confirmed	Dashboard/Audit Summary	MTTD/MTTR (minutes), Audit item completion rate (%) or regular-report issuance rate (%)
7. Incident Response	Establish reporting, action, and recurrence prevention procedures	Incident Runbook (Summary)	24 h Reporting (%), Recurrence Rate (%)
8. User Rights	Establishment of Notification, Explanation Request, Objection, and Reevaluation Pathways	Remedial Process	Process Lead Time (Days), Satisfaction (5-point scale)

Note. This package should be read as a preliminary, expert-informed governance translation rather than as a legally validated compliance checklist or a field-tested implementation model. The governance focus areas are derived from the highest-ranked AHP priorities and interview-based judgement themes. The minimum deliverables indicate evidence artefacts that may support documentation, accountability, monitoring, and auditability, while the check criteria are illustrative and require further organisational and sector-specific validation.

5.2. Data Accountability as a Structural Dimension of Human-Centred AI Governance

The interview findings suggest that data accountability is not limited to privacy protection, but functions as a structural condition for trustworthy AI governance by linking data source, rights, quality, bias, lineage, and change history to responsible

decision-making [5,79] Accordingly, data accountability should be operationalised through data-quality assessment, source and rights documentation, lineage and version records, and corrective-action procedures, rather than being treated as a general ethical statement [7,9,41,69].

5.3. Application Value of AI in Data Capitalism

In the interviews, experts evaluated the application value of AI through both economic efficiency and human value. In this study, data capitalism is therefore used only as an interpretive lens to explain why experts linked AI governance to data concentration, surveillance risks, rights protection, and accountability, rather than as an object of independent empirical testing [8,42,80].

This interpretation supports the need to evaluate AI application value not only by productivity or service efficiency, but also by fairness, rights protection, and responsibility as interconnected governance dimensions [23].

5.4. Policy and Industrial Implications for AI Regulation from an HCAI Perspective

The findings suggest that AI regulation should be treated as a design and institutional mechanism rather than as an ex-post review. From an HCAI perspective, the initial consideration phase should therefore define risk thresholds, data-accountability conditions, and human-oversight decision points before documentation, explanation, and audit tasks are treated as later-stage additions.

The practical implication is to translate these findings into a preliminary HCAI predesign implementation package that links each priority to a decision point, a minimum deliverable, and an operational check criterion [32,78,81,82].

Table 8 presents this preliminary package. Its structure was derived through three steps. First, the literature review identified four theoretical governance axes: human oversight and control, risk management, data accountability, and algorithmic autonomy boundaries [9,12,78]. Second, the AHP results identified priority governance elements, particularly human oversight and control, personal information protection, design ethics, intellectual property rights protection, and limits of algorithmic autonomy. Third, the think-aloud interviews specified how these priorities can be operationalized through intervention authority, automation boundaries, data-accountability documentation, monitoring, and incident response. The resulting decision points, minimum deliverables, and operational KPIs/checklists were therefore not introduced as independent recommendations, but as a governance translation of the priorities and interpretive themes identified in the empirical analysis. In this sense, Table 8 should be read as a preliminary governance translation of the empirical findings, not as a fully validated compliance checklist.

The corresponding practitioner-oriented checklist, implementation package, and templates are provided in Supplementary Tables S1–S4.

6. Answers to Research Questions

This section answers the research questions by consolidating the findings reported above without restating the full discussion. Figure 5 summarises the study's core argument: the EU AI Act's risk-based approach can be interpreted as an HCAI-based pre-design decision structure rather than as an ex-post compliance checklist.

6.1. Answer to the Main Research Question

The MRQ is addressed by showing that the EU AI Act's risk-based approach can be operationalised in the initial consideration phase through four linked decision axes: human oversight, bounded autonomy, ethical design values, and data accountability [9,83].

6.2. Answer to RQ1

Experts perceived human oversight and control as the most important governance factor during the initial consideration phase. At the global-weight level, human oversight and control ranked first, followed by personal information protection, design ethics, intellectual property rights protection, and limits of algorithmic autonomy (Table 6). Within the algorithm-regulation category, however, limits of algorithmic autonomy received the highest local priority. This distinction indicates that, within the governance factors included in the AHP model, experts assigned relatively high priority to controllability, rights protection, ethical design, and accountability. The findings therefore indicate relative priorities among the examined governance factors rather than a direct comparison with technical performance.

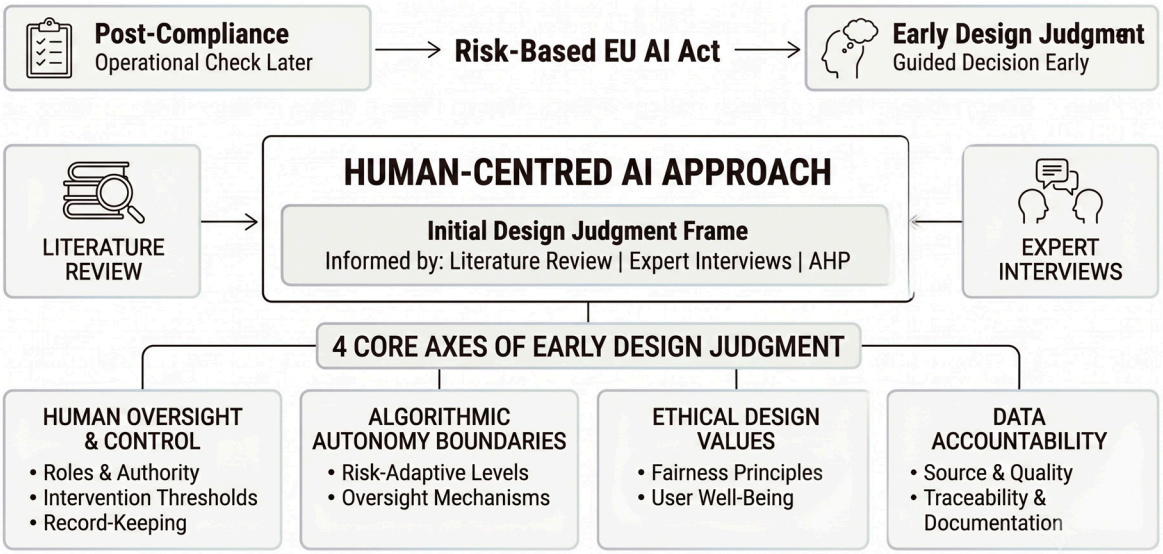


Figure 5. From Ex-Post Compliance to Early Design Judgement: An HCAI-Based. Pre-Design Decision Frame for the EU AI Act.

The interview findings provide interpretive context for this priority by showing that human oversight was interpreted in terms of assignable authority, intervention criteria, documentation, and monitoring procedures, rather than as a general ethical preference [9,15].

6.3. Answer to RQ2

RQ2 is addressed by showing that experts interpreted governance priorities within a data-intensive socio-economic context, especially privacy, IP rights, data quality, bias, and traceability (Tables 6 and 7; Section 5.2). These priorities indicate that data accountability was understood not merely as a technical data-management issue but as an important governance condition for rights protection and legitimate AI use [7,8,84].

However, data capitalism is used here as an interpretive lens rather than as an independently tested theory. The findings therefore support that data accountability should be treated as a pre-design governance condition for rights protection and legitimate AI use.

7. Conclusions

7.1. Summary and Key Contributions

This study examined how HCAI principles may inform early-stage AI governance decisions under the risk-based orientation of the EU Artificial Intelligence Act. By combining AHP with think-aloud interviews, the study derived expert priorities and the interpretive reasoning underlying them. The AHP results identified human oversight and control as

the highest-priority governance factor. At the global-weight level, this was followed by personal information protection, design ethics, intellectual property rights protection, and limits of algorithmic autonomy. The findings suggest that trustworthy AI governance can be more effectively structured during the pre-design phase, when roles, responsibilities, data conditions, intervention criteria, documentation, monitoring, and accountability structures are initially defined. The proposed framework should therefore be understood as an exploratory, expert-informed structure that provides a basis for further organisational and sector-specific validation.

7.2. Theoretical Contribution

This study contributes by positioning pre-design governance as an operational bridge that translates HCAI principles and selected risk-based EU AI Act requirements into early-stage organisational decisions on oversight, accountability, data responsibility, autonomy boundaries, and evidence artefacts [5,9,12]. Although prior HCAI studies emphasise reliable, safe, and trustworthy AI aligned with human judgement [4,12,78], the present study shows how these principles can be translated into early-stage governance priorities involving roles, authority, intervention criteria, documentation, and accountability.

Second, the study connects the EU AI Act's risk-based orientation with HCAI governance. Rather than treating regulatory obligations solely as post-development compliance requirements, the findings suggest that risk classification, human oversight, documentation, and data accountability can guide governance decisions before system development begins [9,36,59]. The study also clarifies that human oversight, and algorithmic autonomy should be understood as complementary governance conditions. Autonomy may remain acceptable when it is bounded by clearly defined intervention rights, monitoring, documentation, and accountability mechanisms [23,50,57].

Third, the study conceptualises data accountability as a structural dimension connecting ethical legitimacy and regulatory responsibility. Data accountability extends beyond personal information protection to include data origin, rights, quality, bias, lineage, version management, and traceability [7,8,40,41]. Methodologically, the combination of AHP and think-aloud interviews further illustrates how quantitative priorities and the interpretive reasoning underlying them can be examined together [17–19].

7.3. Practical Implication

The findings offer practical implications for organisations, practitioners, policymakers, and regulators seeking to operationalise responsible AI governance. First, AI governance should be incorporated into the earliest planning and design decisions rather than treated primarily as a late-stage compliance activity conducted after key design decisions have already been made. Organisations should define oversight responsibilities, intervention criteria, approval authority, documentation requirements, and risk thresholds before technical development begins. This approach may improve organisational preparedness and reduce the difficulty of adding governance controls after key system decisions have already been made.

Second, human oversight and data accountability should be translated into operational procedures. Effective oversight requires organisations to specify who may intervene, under what conditions intervention is required, which actions are permitted, and how decisions and outcomes are recorded. Similarly, data accountability requires systematic documentation of data sources, rights, quality, bias, lineage, versions, and changes. These measures support transparency, traceability, risk management, and auditability across the AI lifecycle.

Third, the preliminary implementation package proposed in this study provides an actionable structure connecting human oversight, automated-execution boundaries, explainability, data accountability, risk approval, monitoring, incident response, and user rights, as summarised in Table 8. The package should not be interpreted as a fully validated compliance model, but as a preliminary governance tool that organisations may adapt and further validate in sector-specific contexts. Policymakers and regulators may also support implementation by providing practical templates, documentation guidance, oversight standards, and audit procedures that help organisations translate regulatory requirements into operational practice.

7.4. Limitations and Future Research Directions

Several limitations should be noted. First, the expert panel was purposively selected and relatively small, which limits the generalisability of the findings. The results should therefore be interpreted as expert-informed exploratory priorities rather than representative population-level judgements. Future studies should examine the proposed framework using larger and more diverse samples.

Second, the proposed governance package should be understood as a preliminary governance translation rather than as a field-tested implementation model. Although the package translates literature-based governance domains, AHP-derived priorities, and think-aloud interview interpretations into decision points, minimum deliverables, and check criteria, it has not yet been tested through organisational case studies, pilot applications, external expert validation, or implementation within a real AI system. Therefore, the package should not be interpreted as a fully validated compliance or implementation model. Future research should examine its applicability through sector-specific case studies, organisational pilots, external validation with practitioners and regulators, and longitudinal evaluation of whether the proposed artefacts and check criteria improve governance traceability, accountability, and human oversight in practice.

Third, the EU Artificial Intelligence Act was used as an analytical reference point rather than as a complete legal framework for doctrinal analysis. Future legal research should further examine how individual obligations under the Act apply across different AI system types, provider roles, deployer responsibilities, and sector-specific regulatory contexts. Future studies should also examine whether the priority structure changes when technical-performance criteria, such as accuracy, robustness, latency, and reliability, are included alongside governance criteria. In addition, because AI regulation and implementation guidance continue to evolve, the selective EU AI Act mapping used in this study should be periodically updated to reflect authoritative regulatory guidance and implementation instruments.

Fourth, the qualitative analysis relies on expert interpretation and could be strengthened through additional anonymised excerpts, further inter-coder reliability procedures, and more explicit links between codes, themes, and AHP results. Future studies should address these limitations through larger samples, cross-national comparisons, sector-specific validation, and longitudinal evaluation of governance outcomes.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/systems14070849/s1.

Author Contributions: Conceptualization, H.-K.L.; methodology, H.-K.L.; formal analysis, H.-K.L.; investigation, H.-K.L.; data curation, H.-K.L.; writing—original draft preparation, H.-K.L.; writing—review and editing, H.-K.L., C.Y. and B.G.L.; supervision, B.G.L.; project administration, H.-K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Institute for Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Ministry of Science and ICT (MSIT, Korea, RS-2026-25525363, Agentic Experts for Generative-AI Inspection Solution).

Institutional Review Board Statement: The study involving human participants was conducted in accordance with the Declaration of Helsinki and approved by the Public Institutional Review Board Designated by Ministry of Health and Welfare (IRB File No. 2025-0938-005; Approval No. P01-202509-01-008; approval date: 1 September 2025).

Informed Consent Statement: Written informed consent was obtained from all participants involved in the study. Consent for audio recording was also obtained prior to the interviews. No identifiable participant information is disclosed in this article.

Data Availability Statement: The aggregated data, AHP summary tables, coding summaries, and supplementary analysis materials supporting the findings of this study are available in Supplementary Materials. Specifically, Supplementary Material File S1 includes the HCAI implementation and operation toolkit, AHP details, and think-aloud interview materials; Supplementary Material File S2 provides the AHP pairwise-comparison questionnaire; and Supplementary Material File S3 provides the expert think-aloud interview guide/protocol. Raw interview transcripts, audio recordings, and individual-level AHP response data are not publicly available because they contain information that could compromise participant confidentiality and are subject to ethical restrictions.

Conflicts of Interest: The authors declare no 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. Gama, F.; Magistretti, S. Artificial intelligence in innovation management: A review of innovation capabilities and a taxonomy of AI applications. *J. Prod. Innov. Manag.* **2025**, *42*, 76–111.
2. Ahmed, F.A.; Gul, S.; Shahzad, S. Ensuring accountability and transparency in AI-driven corporate governance. *Int. J. Soc. Sci. Bull.* **2025**, *3*, 330–341. [CrossRef]
3. Camilleri, M.A. Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert Syst.* **2024**, *41*, e13406.
4. Shneiderman, B. Human-centered artificial intelligence: Reliable, safe & trustworthy. *Int. J. Hum.–Comput. Interact.* **2020**, *36*, 495–504. [CrossRef]
5. Shneiderman, B. Bridging the gap between ethics and practice: Guidelines for reliable, safe, and trustworthy human-centered AI systems. *ACM Trans. Interact. Intell. Syst. (TiiS)* **2020**, *10*, 1–31. [CrossRef]
6. McAfee, A.; Brynjolfsson, E. *Machine, Platform, Crowd: Harnessing Our Digital Future*; WW Norton & Company: New York, NY, USA, 2017.
7. Zuboff, S. *The Age of Surveillance Capitalism: The Fight for a Human Future at the New Frontier of Power*, 1st ed.; PublicAffairs: New York, NY, USA, 2019.
8. Srnicek, N. *Platform Capitalism*; Polity: Cambridge, UK, 2016.
9. European Parliament and the Council. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Available online: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng (accessed on 12 July 2026).
10. Bevilacqua, R.; Bailoni, T.; Maranesi, E.; Amabili, G.; Barbarossa, F.; Ponzano, M.; Virgolesi, M.; Rea, T.; Illario, M.; Piras, E.M. Framing the human-centered artificial intelligence concepts and methods: Scoping review. *JMIR Hum. Factors* **2025**, *12*, e67350. [CrossRef] [PubMed]
11. Schmager, S.; Pappas, I.O.; Vassilakopoulou, P. Understanding Human-Centred AI: A review of its defining elements and a research agenda. *Behav. Inf. Technol.* **2025**, *44*, 3771–3810. [CrossRef]
12. Xu, W.; Dainoff, M. Enabling human-centered AI: A new junction and shared journey between AI and HCI communities. *Interactions* **2023**, *30*, 42–47. [CrossRef]
13. Balcioğlu, Y.S.; Çelik, A.A.; Altindağ, E. A turning point in AI: Europe's human-centric approach to technology regulation. *J. Responsible Technol.* **2025**, *23*, 100128. [CrossRef]
14. Calvano, M.; Curci, A.; Desolda, G.; Esposito, A.; Lanzilotti, R.; Piccinno, A. Building Symbiotic Artificial Intelligence: Reviewing the AI Act for a Human-Centred, Principle-Based Framework. *Minds Mach.* **2026**, *36*, 1. [CrossRef]

15. Amershi, S.; Weld, D.; Vorvoreanu, M.; Fourney, A.; Nushi, B.; Collisson, P.; Suh, J.; Iqbal, S.; Bennett, P.N.; Inkpen, K. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland, 4–9 May 2019*; ACM: New York, NY, USA, 2019; pp. 1–13.
16. Friedman, B.; Hendry, D.G. *Value Sensitive Design: Shaping Technology with Moral Imagination*; Mit Press: Cambridge, MA, USA, 2019.
17. Saaty, T.L. *The Analytic Hierarchy Process*; Springer: New York, NY, USA, 1980.
18. van Someren, M.W.; Barnard, Y.F.; Sandberg, J. *The Think Aloud Method: A Practical Guide to Modelling Cognitive Processes*; Academic Press: Cambridge, MA, USA, 1994.
19. Braun, V.; Clarke, V. Using thematic analysis in psychology. *Qual. Res. Psychol.* **2006**, *3*, 77–101. [CrossRef]
20. Shneiderman, B. Design lessons from AI's two grand goals: Human emulation and useful applications. *IEEE Trans. Technol. Soc.* **2020**, *1*, 73–82. [CrossRef]
21. Bryman, A. *Social Research Methods*; Oxford University Press: Oxford, UK, 2016.
22. Floridi, L.; Cowls, J. A unified framework of five principles for AI in society. *Mach. Learn. City Appl. Archit. Urban Des.* **2022**, 535–545. [CrossRef]
23. Rahwan, I.; Cebrian, M.; Obradovich, N.; Bongard, J.; Bonnefon, J.-F.; Breazeal, C.; Crandall, J.W.; Christakis, N.A.; Couzin, I.D.; Jackson, M.O. Machine behaviour. *Nature* **2019**, *568*, 477–486. [CrossRef] [PubMed]
24. Mazarakis, A.; Bernhard-Skala, C.; Braun, M.; Peters, I. What is critical for human-centered AI at work?—Toward an interdisciplinary theory. *Front. Artif. Intell.* **2023**, *6*, 1257057. [PubMed]
25. Yang, Q.; Steinfeld, A.; Rosé, C.; Zimmerman, J. Re-examining whether, why, and how human-AI interaction is uniquely difficult to design. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020*; ACM: New York, NY, USA, 2020; pp. 1–13.
26. Shneiderman, B.; Plaisant, C.; Cohen, M.; Jacobs, S.; Elmqvist, N.; Diakopoulos, N. Grand challenges for HCI researchers. *Interactions* **2016**, *23*, 24–25. [CrossRef]
27. Donahoe, E. Human Centered AI: Building trust, democracy and human rights by design. An Overview of Stanford's global digital policy incubator and the XPRIZE Foundation's June 11th Event. *Stanf. Glob. Digit. Policy Incubator (GDPi)* **2018**.
28. Kaber, D.B. A conceptual framework of autonomous and automated agents. *Theor. Issues Ergon. Sci.* **2018**, *19*, 406–430.
29. Xu, W. From automation to autonomy and autonomous vehicles: Challenges and opportunities for human-computer interaction. *Interactions* **2021**, *28*, 48–53.
30. McGregor, S. AI Incident Database. Partnership on AI. Available online: https://incidentdatabase.ai (accessed on 12 July 2026).
31. Baecker, R.M. *Readings in Human-Computer Interaction: Toward the Year 2000*; Elsevier: Amsterdam, The Netherlands, 2014.
32. *ISO 31000:2018*; Risk Management—Guidelines. International Organization for Standardization (ISO.): Vernier, Switzerland. Available online: https://www.iso.org/standard/65694.html (accessed on 12 July 2026).
33. Rhodes, R.A. *Understanding Governance: Policy Networks, Governance, Reflexivity and Accountability*; Open University: Milton Keynes, UK, 1997.
34. Agranoff, R.; McGuire, M. Governance Networks. In *Encyclopedia of Public Administration and Public Policy*; Routledge: London, UK, 2003; pp. 552–557.
35. Yampolskiy, R.V. Unexplainability and Incomprehensibility of AI. *J. Artif. Intell. Conscious.* **2020**, *7*, 277–291. [CrossRef]
36. Wirtz, B.W.; Weyerer, J.C.; Kehl, I. Governance of artificial intelligence: A risk and guideline-based integrative framework. *Gov. Inf. Q.* **2022**, *39*, 101685. [CrossRef]
37. Personal Data Protection Commission. Model Artificial Intelligence Governance Framework (2nd ed). Available online: https://www.pdpc.gov.sg/-/media/Files/PDPC/PDF-Files/Resource-for-Organisation/AI/SGModelAIGovFramework2.pdf (accessed on 12 July 2026).
38. Hwang, Y. A study on the use of a generative artificial intelligence design platform for BI development: Focusing on the possibility of using Midjourney. *J. Korean Des. Cult.* **2023**, *29*, 527–541. [CrossRef]
39. Son, H.-R.; Lee, C.-G. Strategies for utilizing artificial intelligence (AI) in the design field: Focusing on in-depth interviews. *J. Digit. Contents Soc.* **2024**, *25*, 2035–2044. [CrossRef]
40. Thatcher, J.; O'Sullivan, D.; Mahmoudi, D. Data colonialism through accumulation by dispossession: New metaphors for daily data. *Environ. Plan. D. Soc. Space* **2016**, *34*, 990–1006. [CrossRef]
41. Van Dijck, J. Datafication, dataism and dataveillance: Big Data between scientific paradigm and ideology. *Surveill. Soc.* **2014**, *12*, 197–208. [CrossRef]
42. Zuboff, S. Big other: Surveillance capitalism and the prospects of an information civilization. *J. Inf. Technol.* **2015**, *30*, 75–89. [CrossRef]
43. Königs, P. In defense of 'surveillance capitalism'. *Philos. Technol.* **2024**, *37*, 122. [CrossRef]
44. Makanadar, A. Digital surveillance capitalism and cities: Data, democracy and activism. *Humanit. Soc. Sci. Commun.* **2024**, *11*, 11533. [CrossRef]

45. Blease, C. Open AI meets open notes: Surveillance capitalism, patient privacy and online record access. *J. Med. Ethics* **2024**, *50*, 84–89. [PubMed]
46. Alshamy, Y.; Coyne, C.J.; Hall, A.R.; Owens, M.A. Surveillance capitalism and the surveillance state: A comparative institutional analysis. *Const. Political Econ.* **2024**, 1–23. [CrossRef]
47. Gebru, T.; Morgenstern, J.; Vecchione, B.; Vaughan, J.W.; Wallach, H.; Iii, H.D.; Crawford, K. Datasheets for datasets. *Commun. ACM* **2021**, *64*, 86–92. [CrossRef]
48. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 29–31 January 2019; ACM: New York, NY, USA, 2019; pp. 220–229.
49. Pushkarna, M.; Zaldivar, A.; Kjartansson, O. Data cards: Purposeful and transparent dataset documentation for responsible ai. In Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, Seoul, Republic of Korea, 21–24 June 2022; ACM: New York, NY, USA, 2019; pp. 1776–1826.
50. Mittelstadt, B.D.; Allo, P.; Taddeo, M.; Wachter, S.; Floridi, L. The ethics of algorithms: Mapping the debate. *Big Data Soc.* **2016**, *3*, 2053951716679679. [CrossRef]
51. Wang, R.; Chen, Z.S. Large-scale foundation models and generative AI for BigData neuroscience. *Neurosci. Res.* **2025**, *215*, 3–14. [CrossRef] [PubMed]
52. Zhang, Z.; Zhang, J.; Zhang, X.; Mai, W. A comprehensive overview of Generative AI (GAI): Technologies, applications, and challenges. *Neurocomputing* **2025**, *632*, 129645. [CrossRef]
53. Arifin, M.M.; Ahmed, M.S.; Ghosh, T.K.; Udoy, I.A.; Zhuang, J.; Yeh, J.-h. A survey on the application of generative adversarial networks in cybersecurity: Prospective, direction and open research scopes. *arXiv* **2024**, arXiv:2407.08839.
54. Tidhar, N.; Li, Y.; Baduge, K.K.; Mendis, P.; Thilakarathna, P.; Chandrathilaka, E. A Slack approach to optimised modularisation of prefabricated structures using a multi-variable modularisation index. *Structures* **2021**, *33*, 1235–1251. [CrossRef]
55. Jia, H.; Rao, H.; Wen, C.; Mirjalili, S. Crayfish optimization algorithm. *Artif. Intell. Rev.* **2023**, *56*, 1919–1979. [CrossRef]
56. Wachter, S.; Mittelstadt, B.; Floridi, L. Why a right to explanation of automated decision-making does not exist in the general data protection regulation. *Int. Data Priv. Law.* **2017**, *7*, 76–99. [CrossRef]
57. Yeung, K. 'Hypernudge': Big Data as a mode of regulation by design. In *The Social Power of Algorithms*; Routledge: London, UK, 2019; pp. 118–136.
58. Burr, C.; Cristianini, N.; Ladyman, J. An Analysis of the Interaction Between Intelligent Software Agents and Human Users. *Minds Mach.* **2018**, *28*, 735–774. [CrossRef]
59. Cajueiro, D.O.; Celestino, V.R.R. A comprehensive review of Artificial Intelligence regulation: Weighing ethical principles and innovation. *J. Econ. Technol.* **2025**, *4*, 77–91.
60. Kim, D.-G.; Park, Y.-W.; Lee, S.-M. A study on the evaluation of tourism resource development projects using the Analytic Hierarchy Process (AHP): Focusing on the planning stage. *J. Tour. Stud.* **2007**, *21*, 5–18. [CrossRef]
61. Kim, S.-C.; Eo, H.-J. A method of aggregating multiple expert opinions in AHP weight determination. *J. Korean Oper. Res. Manag. Sci. Soc.* **1994**, *19*, 41–51.
62. Saaty, T.L. How to make a decision: The analytic hierarchy process. *Eur. J. Oper. Res.* **1990**, *48*, 9–26. [CrossRef]
63. Palinkas, L.A.; Horwitz, S.M.; Green, C.A.; Wisdom, J.P.; Duan, N.; Hoagwood, K. Purposeful sampling for qualitative data collection and analysis in mixed method implementation research. *Adm. Policy Ment. Health Ment. Health Serv. Res.* **2015**, *42*, 533–544.
64. WHO. *Ethics and Governance of Artificial Intelligence for Health*; World Health Organization: Geneva, Switzerland. Available online: https://www.who.int/publications/i/item/9789240029200 (accessed on 12 July 2026).
65. Global Partnership on AI. Global Artificial Intelligence (AI) Governance, Trust, and Human Rights: A framework. Available online: https://verixiv.org/articles/2-187/pdf (accessed on 12 July 2026).
66. Floridi, L.; Cowls, J.; Beltrametti, M.; Chatila, R.; Chazerand, P.; Dignum, V.; Luetge, C.; Madelin, R.; Pagallo, U.; Rossi, F. AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds Mach.* **2018**, *28*, 689–707. [CrossRef]
67. UNESCO. *Recommendation on the Ethics of Artificial Intelligence: Asia-Pacific regional Review.*; United Nations Educational, Scientific and Cultural Organization: Paris, France. Available online: https://unesdoc.unesco.org (accessed on 12 July 2026).
68. BearingPoint. The AI Act Requires Human Oversight. Bearingpoint Insight Report. Available online: https://www.bearingpoint.com/en/insights-events/insights/the-ai-act-requires-human-oversight/ (accessed on 12 July 2026).
69. Papagiannidis, E.; Mikalef, P.; Conboy, K. Responsible artificial intelligence governance: A review and research framework. *J. Strateg. Inf. Syst.* **2025**, *34*, 101885. [CrossRef]

70. UNESCO. Synthetic Content and Its Implications for AI Policy: A Primer. Available online: https://www.researchgate.net/profile/Mariagrazia-Squicciarini/publication/386581930_Synthetic_Content_and_its_Implications_for_AI_Policy_A_Primer/links/6757dee710214e7cc4aa8064/Synthetic-Content-and-its-Implications-for-AI-Policy-A-Primer.pdf (accessed on 12 July 2026).
71. Watkins, L. EU AI Act: Obligations for Deployers of High-Risk AI Systems. Available online: https://www.lw.com/en/insights/eu-ai-act-obligations-for-deployers-of-high-risk-ai-systems (accessed on 12 July 2026).
72. Zendata. Why Data Lineage is Essential for Effective AI Governance. Available online: https://www.zendata.dev/post/why-data-lineage-is-essential-for-effective-ai-governance (accessed on 12 July 2026).
73. EthosAI. Diving Into AI Data Lifecycle Governance. Available online: https://www.ethos-ai.org/p/ai-data-lifecycle-governance (accessed on 12 July 2026).
74. Trigyn Technologies. Intellectual Property Issues in AI: Navigating a Complex Landscape. Available online: https://www.trigyn.com/insights/intellectual-property-issues-ai-navigating-complex-landscape (accessed on 12 July 2026).
75. GDPR Local. AI and Intellectual Property: Legal Challenges and Opportunities. Available online: https://gdprlocal.com/ai-and-intellectual-property-legal-challenges-and-opportunities/ (accessed on 12 July 2026).
76. Government Digital Service. Ethics, Transparency and Accountability Framework for Automated Decision-Making. UK Government. Available online: https://www.gov.uk/government/publications/ethics-transparency-and-accountability-framework-for-automated-decision-making (accessed on 12 July 2026).
77. AI Now Institute. Algorithmic Accountability: Moving Beyond Audits. Available online: https://ainowinstitute.org/publications/algorithmic-accountability (accessed on 12 July 2026).
78. Shneiderman, B. *Human-Centered AI*; Oxford University Press: Oxford, UK, 2022.
79. Habermas, J. *Moral Consciousness and Communicative Action*; MIT Press: Cambridge, MA, USA, 1990.
80. Fuchs, C. Labor in Informational Capitalism and on the Internet. *Inf. Soc.* **2010**, *26*, 179–196. [CrossRef]
81. Endsley, M.R. From here to autonomy: Lessons learned from human–automation research. *Hum. Factors* **2017**, *59*, 5–27. [PubMed]
82. Rahwan, I. Society-in-the-loop: Programming the algorithmic social contract. *Ethics Inf. Technol.* **2018**, *20*, 5–14.
83. Sigfrids, A.; Leikas, J.; Salo-Pöntinen, H.; Koskimies, E. Human-centricity in AI governance: A systemic approach. *Front. Artif. Intell.* **2023**, *6*, 976887. [CrossRef] [PubMed]
84. Owen, R.; Bessant, J.; Heintz, M. *Responsible Innovation; Managing the Responsible Emergence of Science and Innovation in Society*; Wiley: Chichester, UK, 2013.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.