IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Who Decides When AI Advises? Human-AI Cognitive Governance in Higher Education: An Empirical Documentary Framework and Multi-Criteria Decision Simulation

Dr. Kuldeep Kaur
Academic Researcher and Educator
Ph.D., Department of Education, University of Allahabad, Prayagraj, India

Prof. Pramendra Singh Pundir
Professor and Head
Department of Statistics, University of Allahabad, Prayagraj, India

ABSTRACT

The rapid diffusion of generative artificial intelligence (GenAI) in higher education is transforming teaching, learning, assessment, research, academic administration and professional practice. In response, universities have increasingly developed policies concerning permissible AI use, academic integrity, disclosure, privacy, verification and responsible adoption. However, these provisions do not necessarily answer a more fundamental governance question: who retains cognitive and decision-making authority when AI participates in academic judgement? This study develops and empirically grounds Human–AI Cognitive Governance (HACG), an authority-centred framework for examining how higher education institutions structure the relationship between AI assistance and human academic judgement. The study combines documentary policy analysis with a Multi-Criteria Decision-Making (MCDM) simulation. The documentary corpus comprises official AI/GenAI policies, assessment guidance, academic-integrity regulations, research guidance and related institutional documents from 34 universities across the United Kingdom, Australia, the United States, Canada and India. The analysis identifies eight recurring governance dimensions: human decision authority, AI delegation boundaries, verification, epistemic accountability, high-stakes governance, disclosure and traceability, professional accountability, and policy adaptability. These dimensions constitute the proposed Human–AI Cognitive Governance Index (HACGI). Because a fully independently coded 34 × 8 institutional matrix and expert pairwise-comparison data were unavailable, no institutional scores or university rankings are fabricated. Instead, 34 synthetic governance profiles are used to examine the computational behaviour of the HACGI architecture. Equal criterion weights provide a transparent baseline, while TOPSIS is applied to the simulated matrix. A 10,000-replication Monte Carlo sensitivity analysis using Dirichlet-distributed alternative weights assesses ranking stability. HACGI and TOPSIS closeness coefficients show an almost perfect positive association (Pearson's $r = .998$; Spearman's $\rho = .996$; both $p < .001$), while the mean Kendall rank association across weight perturbations is approximately $\tau = .932$, indicating substantial ranking stability under moderate weighting variation. These findings establish computational coherence and methodological robustness, not external criterion validity. Documentary evidence demonstrates that the proposed governance dimensions are observable in contemporary institutional policies. The study

contributes an operational framework for shifting higher-education AI governance from permission-oriented regulation toward authority-oriented governance, with practical relevance for universities, educators, students and policymakers seeking clearer boundaries for AI delegation, verification, human authority and accountability. It also establishes a reproducible pathway for future institutional validation through independent coding, inter-rater reliability, expert content validation and empirically grounded criterion weighting.

Keywords: Generative Artificial Intelligence; Higher Education; AI Governance; Human-AI Cognitive Governance; Authority Specification Gap; HACGI; Human Authority

Introduction

Artificial intelligence (AI) has evolved from a specialised computational technology into an increasingly embedded component of educational, administrative and professional activity. The emergence of generative artificial intelligence (GenAI) has accelerated this transformation by enabling systems to generate text, code, images, summaries, explanations, recommendations, feedback and other forms of knowledge-related output. As these capabilities become increasingly accessible across institutional settings, higher education institutions must reconsider not only how AI can be adopted responsibly but also how human responsibility, professional judgement and decision authority should be organised when AI becomes integrated into academic and administrative processes. In higher education, GenAI is increasingly relevant to teaching, learning, assessment, research, student support, curriculum development, academic administration and professional practice. Recent systematic reviews indicate that institutional responses to GenAI increasingly encompass academic integrity, privacy, responsible use, transparency, equitable access, AI literacy, institutional preparedness, human oversight and accountability (Alfiras et al., 2026; Bacalso et al., 2026; Crompton et al., 2026; Jiang & Abdullah, 2026). UNESCO's guidance similarly places human agency, ethical safeguards, privacy, institutional preparedness and human capacity at the centre of responsible GenAI adoption in education and research (Miao & Holmes, 2023). The OECD (2026) likewise emphasises systematic and responsible approaches to the adoption and governance of GenAI in higher education. Collectively, this literature demonstrates that responsible AI adoption is not solely a technological concern; it is also an institutional and governance challenge. Despite the growing body of research on responsible AI, AI readiness and institutional governance, an important question remains insufficiently specified: what happens to human decision authority when AI becomes a participant in processes of academic and professional judgement? This question is particularly significant given the increasing influence of AI systems on decisions without their necessarily becoming formal decision-makers.

Consider several increasingly common situations. An instructor may use an AI system to generate feedback on a student's assignment; a university may employ an AI-supported system to identify students who may require academic intervention; a researcher may use AI to prioritise literature or classify evidence; or an administrative system may generate a recommendation concerning student eligibility. In each case, a human actor may remain formally involved in the process. However, formal human involvement does not necessarily imply meaningful human authority. For example, an AI-supported assessment system may generate a recommendation regarding a student's academic performance, progression or eligibility while the responsible academic may lack the expertise, time, information, institutional authority or procedural protection required to critically evaluate and challenge that recommendation. Similarly, an institutional policy may require "human oversight" without specifying whether the responsible academic is expected to independently evaluate the AI output, whether the individual has the authority to modify or reject it, who is authorised to override the recommendation or who retains accountability when an AI-supported recommendation is accepted. Thus, the presence of a human in a decision process should not automatically be interpreted as evidence of meaningful human control. The central distinction, therefore, is that the mere presence of human involvement does not, in itself, ensure meaningful human authority, particularly when the human actor lacks the capacity or institutional power to evaluate, challenge, modify or override AI-generated outputs. This distinction provides the conceptual starting point for Human-AI Cognitive Governance (HACG). HACG focuses on the institutional structuring of cognitive, evaluative and decision-making authority between human

academic actors and AI systems. It complements existing responsible-AI governance and AI-readiness approaches by addressing a different analytical layer: not merely whether an institution possesses the capacity to use AI responsibly but who has the right, responsibility and practical capacity to decide when AI advises, influences, recommends or contributes to consequential academic and institutional actions.

Existing technology-adoption research provides an important foundation for understanding AI acceptance and use. The Technology Acceptance Model (TAM) explains technology acceptance primarily through perceived usefulness and perceived ease of use (Davis, 1989) while the Unified Theory of Acceptance and Use of Technology (UTAUT) incorporates performance expectancy, effort expectancy, social influence and facilitating conditions (Venkatesh et al., 2003). AI-literacy research further identifies competencies required to understand, evaluate and appropriately interact with AI systems (Long & Magerko, 2020). Although these approaches provide valuable insights into adoption, acceptance and individual capability, they do not directly specify institutional decision rights, authority boundaries or accountability arrangements in AI-mediated decision processes. Related research on human interaction with automation provides another important foundation. Studies of automation have examined levels of automation, human interaction, supervisory control and calibrated reliance on automated systems (Parasuraman et al., 2000; Lee and See, 2004). However, the institutional governance of AI-mediated cognitive activity requires a broader perspective. HACG extends this line of inquiry from human–automation interaction to institutional governance by examining whether human participation is accompanied by meaningful decision authority, verification capacity, intervention and override rights and clearly assigned accountability.

Recent work by Kaur and Pundir (2026a, 2026b) conceptualises responsible AI readiness in higher education through the interconnected dimensions of faculty capability, professional judgement and institutional conditions. The present framework addresses a related but analytically distinct problem: how authority should be structured once AI becomes embedded within cognitive, evaluative and professional activity. Whereas AI readiness concerns an institution's capacity to engage with AI responsibly, HACG focuses on the governance of authority during AI-supported activity. In this sense, readiness provides an enabling condition for responsible AI use, whereas cognitive governance addresses the allocation and protection of human decision authority within that use. To address this gap, the present study develops HACG as an authority-centred framework for analysing responsible AI governance in higher education. A central concept within the framework is the Authority Specification Gap, which refers to situations in which institutional policies regulate or encourage AI use without sufficiently specifying decision rights, delegation boundaries, verification obligations, intervention or override rights and accountability. The existence of such a gap may create a discrepancy between the formal presence of human oversight and the actual capacity of human actors to exercise meaningful authority. To further conceptualise this relationship, the study proposes a Human Authority Gradient, whereby the requirement for meaningful human authority should increase as the potential consequence and irreversibility of an AI-influenced decision increase and as the opportunities for contestation decrease. This principle shifts attention from the binary question of whether "human oversight" exists to the more substantive question of how much authority, verification capacity and intervention capability the human actor actually possesses. The framework is operationalised through the eight-dimensional Human-AI Cognitive Governance Index (HACGI), which provides a structured basis for evaluating institutional arrangements governing AI-mediated cognitive and decision-making activity. To strengthen the methodological treatment of HACGI, the study further incorporates a multi-criteria decision-making (MCDM) simulation using the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and Monte Carlo sensitivity analysis. Importantly, the study distinguishes simulated results from empirical observations and does not present simulation outputs as evidence of actual institutional performance. Instead, the simulation is used to examine the methodological behaviour and robustness of the proposed index under alternative weighting conditions.

Accordingly, the study makes five principal contributions. First, it develops HACG as an authority-centred framework for responsible AI governance in higher education. Second, it conceptualises the Authority Specification Gap as a distinct governance problem arising when institutional AI policies fail to adequately define decision rights, delegation boundaries, verification obligations, override rights and accountability. Third, it introduces the Human Authority Gradient to explain why the strength of human authority should

correspond to the consequence, irreversibility and contestability of AI-influenced decisions. Fourth, it operationalises the framework through the eight-dimensional HACGI. Fifth, it strengthens the methodological foundation of the framework by applying TOPSIS and Monte Carlo sensitivity analysis while maintaining a clear distinction between simulated evidence and observed institutional evidence. Based on these conceptual and methodological objectives, the study addresses the following research questions:

RQ1. How can human decision authority in AI-mediated higher education be operationalised as a documentary governance construct?

RQ2. Can HACGI provide a transparent, multidimensional representation of institutional AI governance?

RQ3. Can HACGI be incorporated into an MCDM framework without introducing unsupported expert weights?

RQ4. How robust is HACGI-based institutional differentiation under alternative criterion-weight configurations?

RQ5. What does contemporary institutional policy reveal about the emerging Authority Specification Gap?

Literature Review

From Technology Adoption to AI Governance

Research on information technology adoption has traditionally examined the factors that explain why individuals accept and use technological systems. Davis's (1989) Technology Acceptance Model (TAM) established perceived usefulness and perceived ease of use as key determinants of technology acceptance while Venkatesh et al. (2003) subsequently integrated major technology-adoption perspectives through the Unified Theory of Acceptance and Use of Technology (UTAUT). These models remain relevant to higher education as they help explain the adoption of AI systems by students, faculty, researchers and administrators. However, adoption and governance address fundamentally different questions. Adoption concerns why individuals use AI; governance concerns the conditions, safeguards, responsibilities and institutional boundaries under which AI should be used. HACG extends this distinction by asking a further question: when AI participates in judgement, who retains authority over the resulting decision?

This distinction becomes critical when AI-generated outputs influence assessment, academic progression, student support, research evaluation or institutional decision-making. AI literacy provides an important complementary dimension. Long and Magerko (2020) conceptualise AI literacy as the competencies required to understand, evaluate and effectively interact with AI systems. Such competencies are essential for meaningful human oversight, as effective verification depends on understanding AI capabilities, limitations and contextual relevance. Nevertheless, competence does not itself confer authority. An institution may have highly AI-literate personnel while leaving decision rights and override responsibilities ambiguous. Conversely, formal responsibility may be assigned to faculty without providing adequate time, training, information or institutional support to exercise it effectively. HACG therefore treats capability and authority as related but analytically distinct dimensions of AI governance.

Human Interaction with Automation

Research on automation has long examined how humans interact with increasingly autonomous systems. Parasuraman et al. (2000) distinguish different levels and functions of human interaction with automation while Lee and See (2004) emphasise appropriate reliance and calibrated trust in automated systems. These perspectives are increasingly relevant to GenAI for contemporary systems can produce fluent, plausible and apparently authoritative outputs despite substantial variation in reliability across tasks and contexts.

Accordingly, the central governance issue is not merely whether a human remains *in the loop* but whether that human possesses sufficient information, competence, discretion and institutional authority to exercise independent judgement. A nominal human reviewer may therefore provide procedural oversight without possessing substantive control over the decision. HACG extends the human–automation literature by reframing *human-in-the-loop* as *human-in-authority*, emphasising substantive human capacity to evaluate, challenge, modify, reject or override AI-supported outputs.

Responsible AI Governance in Higher Education

The rapid expansion of GenAI has prompted higher education institutions to develop increasingly comprehensive governance arrangements. Recent systematic reviews identify academic integrity, responsible and ethical use, privacy, equitable access, AI literacy, integration strategy, institutional support, human oversight and accountability as recurring governance priorities (Alfiras et al., 2026; Bacalso et al., 2026; Crompton et al., 2026; Jiang & Abdullah, 2026). Academic integrity has received particular attention **as** GenAI challenges established assumptions concerning authorship, originality, independent student work, and assessment validity (Bittle & El-Gayar, 2025). Broader governance concerns additionally encompass privacy, intellectual property, fairness, transparency, security, responsible adoption and professional accountability.

UNESCO's *Guidance for Generative AI in Education and Research* advocates a human-centred approach in which AI serves human purposes without displacing human agency (Miao & Holmes, 2023). Similarly, the OECD (2026) emphasises systematic institutional governance rather than technology deployment alone. Crompton et al. (2026), drawing on a global Delphi study, identify interconnected governance domains including academic integrity, responsible use, privacy and protection, equitable access, GenAI literacy, integration strategy, human oversight and accountability and institutional support. Qian (2026) likewise highlights instructor-level governance, disclosure, attribution, privacy safeguards, AI literacy and caution regarding automated detection.

This literature establishes a substantial foundation for responsible AI governance in higher education, yet the allocation and protection of human decision authority remain insufficiently specified. Institutional policies may require transparency without identifying who holds final authority; mandate verification without establishing who may reject an AI output; or prescribe human oversight without defining the level of discretion, intervention and accountability that oversight entails. Thus, the existence of governance provisions does not necessarily demonstrate the existence of meaningful human authority.

HACG addresses this unresolved governance layer by examining how decision rights, verification responsibilities, intervention and override capacities and accountability are institutionally specified when AI becomes involved in cognitive and professional judgement.

The Authority Specification Gap

The central conceptual problem addressed by this study is the Authority Specification Gap. It arises when an institutional governance policy defines what constitutes acceptable or unacceptable AI use but does not adequately specify how decision-making authority is allocated between AI systems and human actors. In other words, a policy may regulate whether AI can be used without establishing who has the authority and responsibility to determine what ultimately happens when AI contributes to a decision. For example, a policy may permit AI to support the generation of academic feedback. Such a provision establishes primarily a permission rule: it defines an allowable use of AI. A stronger governance provision would specify that AI-generated feedback may be used only as an advisory input, that the responsible educator must independently evaluate its relevance and accuracy and that the educator retains final authority over the academic judgement. This constitutes an authority rule, as it clearly delineates the boundaries of AI contribution while preserving human decision responsibility. The distinction is therefore fundamental. Permission determines whether AI may be used, whereas delegation determines how decision authority is distributed. Similarly, human oversight concerns human involvement in an AI-mediated process, whereas human authority concerns the human capacity to independently evaluate, influence and ultimately control the decision outcome.

The Authority Specification Gap becomes particularly consequential when AI outputs inform decisions affecting assessment, academic progression, student support, research evaluation or other institutional outcomes. A policy requirement that “human review is required” may appear to establish human control while leaving critical governance arrangements unspecified. It may not identify who conducts the review, what information and competencies are required, whether the reviewer must independently evaluate the AI output or whether the reviewer possesses sufficient discretion to question, modify, reject or override the recommendation. It may also fail to establish who retains final accountability and whether affected individuals have meaningful opportunities to challenge or appeal an AI-influenced decision. Meaningful human authority therefore requires more than physical presence or procedural review. It requires access

to relevant information, sufficient competence to evaluate AI outputs, adequate time and institutional capacity and genuine discretion to exercise independent judgement. Crucially, the human actor must possess the authority to question, modify, reject or override AI-generated recommendations where appropriate, together with clearly defined accountability for the resulting decision.

The Authority Specification Gap thus represents a critical distinction between merely regulating AI activity and explicitly governing decision rights. HACG addresses this gap by examining whether institutional arrangements translate the principle of human oversight into clearly defined, exercisable and accountable forms of human decision authority.

Human-AI Cognitive Governance

Human-AI Cognitive Governance (HACG) refers to the institutional structuring of cognitive, evaluative and decision-making authority between human academic actors and artificial intelligence systems. It focuses on how institutions define and manage the respective roles of humans and AI when AI is used in academic, research, administrative or professional activities. The scope of HACG is limited to authority arrangements within AI-mediated cognitive and decision processes. It considers which tasks AI can perform, which responsibilities should remain with human actors and which functions can be delegated under clearly defined conditions. It also considers when AI-generated outputs need human verification, who has the authority to accept, modify, reject or override an AI recommendation and who remains accountable for the final decision. Since, AI capabilities and associated risks continue to change, these arrangements also need to be reviewed and revised over time.

HACG is not intended to measure AI adoption, technology acceptance or AI literacy by themselves. Rather, it considers whether these capabilities are accompanied by meaningful and exercisable human decision authority. In this sense, HACG connects AI governance with professional judgement, human agency, automation and institutional accountability while keeping the allocation and exercise of decision authority at the centre of analysis.

AI Assistance and AI Authority

The role of AI in higher education can vary considerably, from providing routine cognitive assistance to influencing decisions with significant academic or institutional consequences. At a relatively low level of involvement, AI may be used for brainstorming, language refinement, summarisation, generating examples or organising information. At a higher level, it may generate recommendations related to formative feedback, assessment design, literature discovery, student support or the organisation and interpretation of research evidence. More consequential uses include classifying student responses, identifying possible academic-integrity concerns, generating preliminary grades, identifying students who may be academically at risk or supporting aspects of research evaluation. The governance implications of these uses depend not only on the task itself but also on how much decision-relevant influence is given to the AI system and what consequences may follow from its output. As AI becomes more involved in evaluative or decision-relevant activities, institutions need stronger arrangements for human verification, intervention, discretion and accountability. The key issue, therefore, is not simply whether AI is being used but what role it plays in the decision process and whether human actors retain effective control over the resulting judgement.

HACG distinguishes between AI assistance and AI authority. AI assistance occurs when an AI system provides information, suggestions or cognitive support while the human actor retains substantive control over evaluation and decision-making. AI authority, as used in this framework, does not mean that an AI system independently possesses institutional authority. Rather, it refers to situations in which AI outputs acquire substantial influence over decisions due to institutional delegation, established procedures or reliance on the system's recommendations. The governance concern arises when this influence becomes greater than the human actor's practical ability to independently assess, question, modify, reject or override the output. The central concern of HACG is therefore whether institutional arrangements preserve meaningful human decision authority when AI becomes increasingly involved in consequential cognitive and evaluative activities.

Human Authority Gradient

Human oversight should not be conceptualised as a binary condition determined solely by whether a human is present in an AI-mediated process. The appropriate level of human authority should instead be calibrated to the characteristics and potential consequences of the decision. On this basis, the present study proposes the Human Authority Gradient (HAG) as a conceptual principle for determining how the strength of human decision authority should vary across AI-mediated contexts. The HAG is grounded in three interrelated characteristics of AI-influenced decisions: consequence, irreversibility and contestability. As the potential consequences of a decision become more significant, its effects become more difficult to reverse or opportunities for meaningful challenge become more limited, stronger and more explicit human authority becomes necessary. HAG thus shifts the analytical focus from the mere presence of human oversight to the substantive capacity of human actors to exercise independent judgement and control over outcomes.

Consequence

Consequence refers to the magnitude and significance of the potential effects of an AI-influenced decision on individuals, groups or institutions. Lower-consequence applications, such as brainstorming, language refinement or classroom example generation, generally carry limited implications when outputs are inaccurate. In contrast, AI applications involving grading, academic progression, admissions, disciplinary decisions, student support, employment-related decisions or research evaluation may materially affect rights, opportunities, professional standing or institutional outcomes. As the potential consequence of an AI-influenced decision increases, governance arrangements should require correspondingly stronger human authority, including clearly defined decision rights and sufficient discretion to independently question, modify, reject or override AI-generated outputs.

Irreversibility

Irreversibility refers to the extent to which an AI-influenced decision can be corrected, reversed or meaningfully remedied after implementation. Errors in low-stakes applications may often be corrected through subsequent clarification or adjustment. By contrast, decisions affecting academic progression, disciplinary status, professional standing or access to institutional opportunities may generate effects that are difficult to reverse or that persist even after formal correction. Greater irreversibility therefore warrants stronger safeguards before a decision becomes operational. These may include independent human review, explicit decision authority, adequate documentation, reconsideration procedures and accessible correction or appeal mechanisms. The less reversible the potential outcome, the stronger the justification for requiring substantive human authority at the point of decision.

Contestability

Contestability refers to the extent to which individuals affected by an AI-influenced decision can question its basis, challenge its outcome and obtain meaningful human reconsideration. It becomes particularly important where AI outputs are difficult to interpret, where affected individuals have limited access to relevant information or where institutional procedures provide weak opportunities for review.

Where contestability is limited, stronger human authority becomes essential. Without an effective mechanism for challenge and reconsideration, an AI-assisted recommendation may acquire an unwarranted appearance of objectivity or finality, even where the underlying output is uncertain, incomplete or erroneous.

The Human Authority Gradient therefore provides a risk-sensitive basis for calibrating human decision authority within AI-mediated institutional processes. By relating the required strength of human authority to consequence, irreversibility and contestability, HAG connects responsible AI governance with procedural fairness, human agency and institutional accountability. Its central proposition is that human oversight should become progressively more substantive as AI involvement moves closer to consequential, difficult-to-reverse and weakly contestable decisions. In this sense, HAG provides the conceptual bridge between the governance principles of HACG and their operationalisation through the Human-AI Cognitive Governance Index (HACGI).

Conceptual Propositions

The Human Authority Gradient provides a basis for deriving eight propositions concerning the conditions under which human decision authority should be strengthened in AI-mediated higher education. These propositions translate the framework's central assumptions into empirically testable statements and provide a foundation for future validation of HACG and HACGI.

P1. The greater the potential consequences of an AI-influenced academic decision, the stronger the level of meaningful human authority required to govern that decision.

P2. As the irreversibility of an AI-influenced decision increases, institutional requirements for human verification, documentation, review and override should become more stringent.

P3. Where opportunities for contesting an AI-influenced decision are limited, explicit human decision authority becomes increasingly important.

P4. Human review constitutes meaningful human authority only when the reviewer possesses the substantive capacity to independently evaluate, modify, reject or override AI-generated outputs.

P5. As institutional use of AI moves from cognitive assistance toward greater functional delegation, corresponding requirements for verification, intervention and accountability should increase.

P6. Verification of AI-generated outputs should be understood as a form of epistemic governance, involving the evaluation of accuracy, relevance, uncertainty and contextual appropriateness, rather than merely technical error detection.

P7. Professional accountability requires institutional arrangements that enable human actors to exercise substantive decision authority; assigning responsibility without corresponding authority constitutes insufficient governance.

P8. Adaptive policy review is a necessary component of cognitive governance as changes in AI capabilities, applications and associated risks can alter the appropriate allocation of human and AI roles over time.

These propositions establish the central logic of the Human Authority Gradient: the greater the potential for AI to shape consequential, irreversible or difficult-to-contest outcomes, the stronger and more explicitly defined human authority should be. The propositions are advanced as testable theoretical claims for future empirical research, rather than as causal relationships established by the present simulation.

Human-AI Cognitive Governance Index

Purpose and Conceptual Status

The Human-AI Cognitive Governance Index (HACGI) operationalises the Human-AI Cognitive Governance (HACG) framework as a structured documentary coding instrument for examining how institutional policies specify human authority in AI-mediated academic and professional processes. HACGI is deliberately positioned as a governance-document analysis tool rather than as a conventional psychometric or institutional-performance measure. It does not measure individual attitudes, AI adoption, implementation quality or institutional effectiveness. Nor is it intended, at this stage, to function as a validated institutional ranking instrument. Its primary purpose is to make otherwise diffuse authority arrangements systematically observable, codable and comparable across institutional policy documents.

The index comprises eight dimensions that capture the principal elements through which human authority can be specified in AI-mediated decision processes:

Dimension	Code	Core Question
Human Decision Authority	HDA	Is final decision authority explicitly retained by humans?
AI Delegation Boundary	ADB	Are AI functions and limits of delegation clearly specified?
Verification Requirement	VR	Are AI outputs subject to meaningful human verification?
Epistemic Accountability	EA	Is responsibility for accuracy, interpretation and contextual validity assigned?
High-Stakes Governance	HSG	Are stronger safeguards applied to consequential AI uses?
Disclosure and Traceability	DT	Is AI use documented, disclosed or sufficiently traceable?
Professional Accountability	PA	Does professional responsibility remain clearly assigned to human actors?
Policy Adaptability	PAD	Are mechanisms provided for reviewing and revising AI governance arrangements?

Together, these dimensions distinguish the presence of human involvement from the specification of human authority. An institution may, for example, require human oversight or disclose AI use while providing limited specification of who can challenge, modify, reject or override an AI-supported recommendation. HACGI is designed to identify precisely such differences in governance specification.

Coding Scale

Each HACGI dimension is coded on a four-point ordinal scale from 0 to 3, reflecting the degree to which the relevant governance provision is specified in the institutional document:

Score	Interpretation
0	Not specified
1	Implicit or weakly specified
2	Explicit but partial
3	Explicit and operationally specified

A score of 0 indicates that the relevant authority arrangement is absent from the document. A score of 1 indicates that the issue is addressed only indirectly or through broad principles. A score of 2 indicates an explicit provision but with incomplete specification of responsibilities, procedures or decision rights. A score of 3 indicates that the provision is explicitly defined with sufficient operational detail to identify the relevant responsibility, procedure, authority or safeguard.

With eight dimensions and a maximum score of three per dimension, the maximum possible raw score is:

[
HACGI_{\max}=8\times3=24
]

For institution (i), the descriptive HACGI score is calculated as:

[
HACGI_i=\frac{\sum_{j=1}^{8}x_{ij}}{24}\times100
]

where (x_{ij}) represents the score assigned to institution (i) on dimension (j).

The resulting percentage should be interpreted cautiously as an indicator of the formal specification of AI-related governance and human authority within the analysed documents. It should not be interpreted as direct evidence of implementation, institutional effectiveness or actual behaviour. A high score indicates

stronger documentary specification; it does not establish that the corresponding practices are consistently enacted.

HACGI Coding Rubric

The coding rubric translates the eight HACGI dimensions into observable documentary indicators. The emphasis is on what an institutional document explicitly specifies, rather than on assumptions about institutional practice.

Dimension	0 - Not specified	1 - Implicit or weak	2 - Explicit but partial	3 - Explicit and operationally specified
HDA	Human decision authority not addressed	Human involvement or responsibility implied	Human responsibility or final judgement explicitly stated	Final human authority, decision rights and/or override authority clearly specified
ADB	AI delegation boundaries absent	Broad responsible-use or cautionary language	Permitted or prohibited AI functions specified	Functional boundaries, delegation limits and Human-AI roles clearly defined
VR	No verification requirement	General caution concerning AI outputs	Human verification explicitly required	Verification procedure, responsibility and decision consequences clearly specified
EA	Accountability for AI-supported outputs not specified	General warnings concerning accuracy or reliability	Human responsibility for accuracy or judgement stated	Responsibility for validation, interpretation and contextual accuracy operationalised
HSG	No differentiated treatment of consequential uses	General caution concerning risks	Higher-risk or high-stakes uses explicitly differentiated	Enhanced safeguards, human review, intervention or appeal mechanisms specified
DT	AI use neither disclosed nor documented	General transparency expectations	Disclosure or acknowledgement of AI use required	Documentation, attribution, provenance or traceability requirements clearly specified
PA	Professional accountability not addressed	Responsibility implied	Human or professional responsibility explicitly assigned	Role-specific responsibility, decision authority and intervention or override arrangements specified
PAD	No provision for policy change	Recognition that AI practices may evolve	Periodic policy review specified	Continuous monitoring, structured review and revision mechanisms clearly established

The rubric is intended to support transparent and reproducible documentary coding. Coders should assign scores on the basis of explicit textual evidence within the analysed policy or governance document and should avoid inferring institutional practices that are not documented. Where a provision could reasonably receive two scores, the lower score should be assigned unless the document provides sufficient evidence for the higher level of specification. Dimension-level scores should always be retained alongside the

aggregate HACGI score. This distinction is important, as institutions may exhibit substantially different governance profiles despite achieving similar overall scores. For example, an institution may have strong provisions for disclosure and traceability while providing limited specification of human decision authority, verification responsibilities or override rights.

Accordingly, the aggregate HACGI score should be treated as a summary indicator rather than a substitute for the underlying governance profile. The dimension-level pattern is substantively important. The central concern of HACG is not simply how much governance a policy contains but whether the policy clearly specifies and protects human authority at the points where AI can influence consequential judgement and decision-making.

Documentary Empirical Grounding

Institutional Corpus

The documentary corpus comprises official AI-related policies, guidelines, assessment guidance, academic regulations and institutional statements from 34 universities across five national contexts: the United Kingdom, Australia, the United States, Canada and India. The institutions were purposively selected to provide variation across institutional and national contexts and to establish sufficient documentary breadth for examining how authority-related dimensions of AI governance are articulated in higher education.

The United Kingdom corpus comprises nine institutions: the University of Oxford, University of Cambridge, University College London, London School of Economics and Political Science, King's College London, University of Edinburgh, Durham University, University of Manchester and University of Warwick.

The Australian corpus comprises eight institutions: University of Melbourne, University of Sydney, University of Queensland, UNSW Sydney, Monash University, Macquarie University, Deakin University and Griffith University.

The United States corpus comprises fourteen institutions: University of Southern California, Harvard University, Massachusetts Institute of Technology, Princeton University, Cornell University, Columbia University, Stanford University, University of California, Berkeley, Duke University, University of Chicago, Yale University, University of Pennsylvania, University of California, Los Angeles and University of Michigan.

The Canadian corpus comprises two institutions: University of Toronto and University of British Columbia. The Indian corpus comprises one institution, the Indian Institute of Technology Delhi.

The corpus is purposive rather than nationally representative. Its purpose is therefore not to estimate national prevalence or support population-level comparisons across countries. Instead, the multi-jurisdictional sample provides contextual breadth for examining whether the governance dimensions proposed by HACG can be identified in actual institutional policy and guidance.

Document Selection Protocol

To establish a consistent documentary basis for analysis, institutional documents were selected using a predefined document selection protocol. The protocol was designed to identify documents that contain substantive provisions concerning the use, oversight, evaluation or governance of AI within academic and institutional processes. Documents were eligible for inclusion when they were:

1. Officially issued or hosted by the university, including central policy offices, academic governance bodies, teaching and learning units, assessment authorities, research offices or equivalent institutional entities;
2. Directly concerned with AI or AI-mediated academic activity, including generative AI, artificial intelligence in assessment, teaching, learning, research, feedback, marking, student support or academic administration;
3. Normative or operational in character, meaning that they specified permissions, restrictions, responsibilities, procedures, expectations, decision rights, verification requirements, disclosure obligations or accountability arrangements rather than merely describing AI technologies;
4. Applicable to institutional or academic practice, either institution-wide or at the level of faculties, courses, assessments or academic roles; and

5. Sufficiently substantive to permit coding against the operational meaning of the HACGI dimensions.

Priority was given to current institutional policies and guidance available during the study period. Where a university maintained several overlapping documents, documents were retained when they contributed distinct governance information; substantially duplicative documents were treated as supporting rather than independent evidence. Public-facing explanatory material was considered only when it originated from an official institutional source and contained substantive governance provisions. Documents were excluded when they were primarily news announcements, promotional material, general commentary, technical descriptions of AI tools, student-facing informational material without governance content or documents in which AI was mentioned without establishing a relevant institutional rule, responsibility, decision boundary or accountability mechanism. The selection process therefore sought to avoid treating the mere presence of the term *AI* as evidence of governance. Document inclusion was determined by substantive governance relevance rather than keyword occurrence alone. This distinction is important as a document may discuss AI extensively without specifying who retains decision authority, who must verify AI outputs or who remains accountable for consequential academic decisions.

Documentary Tracing of HACGI Dimensions

The eight HACGI dimensions were traced through the selected documents using a construct-to-document mapping procedure. The analysis did not require institutions to use the terminology of HACG or HACGI. Instead, the coding focused on whether institutional documents contained provisions that were substantively equivalent to the operational definitions of the eight dimensions. This approach recognises that universities frequently express similar governance principles using different terminology. For example, a policy may not use the term *human decision authority* but may state that an educator, assessor or academic authority must make the final judgement. Similarly, a policy may not use the term *verification requirement* but may require academic staff to critically evaluate, check or validate AI-generated content before relying upon it. The analysis therefore coded governance substance rather than terminology. The tracing procedure involved three sequential stages. First, the selected institutional documents were reviewed to identify provisions relating to AI use in teaching, learning, assessment, research, feedback, marking, student support and academic administration. Second, relevant provisions were extracted and categorised according to their substantive governance function. Third, the extracted provisions were compared with the operational definitions of the eight HACGI dimensions to establish a dimension-to-document traceability matrix. The traceability logic was applied as follows. Provisions specifying permitted, restricted or prohibited forms of AI use were examined as evidence of AI-use boundary definition. Requirements for critical checking, fact-checking, validation or human review of AI outputs were examined as evidence of verification requirements. Provisions assigning final academic judgement or responsibility to educators, assessors, researchers or designated institutional authorities were examined as evidence of human decision authority and accountability. Provisions allowing academic actors to modify, reject, challenge or override AI-generated recommendations were examined as evidence of human intervention or override authority. Requirements concerning disclosure, documentation, attribution, communication or explanation of AI involvement were examined as evidence of the relevant transparency and disclosure dimension. The remaining HACGI dimensions were traced through the same principle by identifying institutional provisions that substantively corresponded to their respective operational definitions.

A HACGI dimension was considered documentarily observable when at least one substantive institutional provision could be mapped to its operational definition. Documentary observability does not imply that the dimension is comprehensively implemented within the institution. It indicates only that the underlying governance construct has a recognisable institutional manifestation in formal policy or guidance. This distinction is methodologically important. The purpose of documentary tracing is not to convert every policy statement into a numerical institutional score. Rather, it establishes a defensible empirical link between the conceptual framework and real institutional governance arrangements. The traceability matrix consequently functions as a bridge between construct conceptualisation and empirical operationalisation while avoiding the stronger and unwarranted claim that documentary presence constitutes validated institutional performance.

Documentary Evidence of Authority-Oriented Governance

The documentary analysis provides substantive grounding for the eight dimensions of the Human-AI Cognitive Governance Index (HACGI) by demonstrating that authority-related governance provisions are identifiable in contemporary university policies and guidance. The evidence does not imply that institutions have explicitly adopted HACG or HACGI; rather, it shows that policy provisions corresponding to the framework's underlying constructs are already present in institutional practice.

For example, the University of Oxford requires assessment setters to clarify whether and how AI may be used in particular assessments and provides guidance concerning disclosure and assessment design. These provisions illustrate the specification of task-level AI permissions and transparency requirements.

King's College London provides particularly explicit guidance on AI-assisted marking and feedback. Its policy distinguishes AI assistance from the replacement of human expertise and academic judgement and does not permit AI to independently determine marks or academic judgements. This provides a clear documentary basis for dimensions concerned with human decision authority and professional accountability.

University College London distinguishes categories of AI use in assessment and requires assessment-specific communication regarding permissible AI involvement. This illustrates the importance of clearly defined boundaries around AI use and the communication of those boundaries to academic actors and students.

Similarly, UNSW establishes assessment-specific conditions for AI use and emphasises critical review and due diligence in relation to AI-generated outputs. Such provisions provide observable evidence for verification and evaluative accountability within AI-mediated academic processes.

Monash University uses assessment-specific AI statements and assigns responsibility to relevant academic authorities for determining how AI may be used. This illustrates the importance of explicit role allocation and governance at the level where academic decisions are actually made.

The University of Queensland provides a further example of increasingly differentiated AI governance through its assessment conditions, including an “AI required” category involving tool selection, prompting, verification and integration. Such provisions suggest a movement beyond simple permission or prohibition towards more explicit specification of how AI is expected to participate in academic work.

Comparable patterns are visible in the North American corpus. Princeton places substantial responsibility for AI-use decisions at the instructor or course level and addresses disclosure where AI use is permitted. Harvard requires faculty to communicate expectations concerning AI use and emphasises responsibility for evaluating AI-generated outputs. Cornell similarly addresses policy clarification, documentation, attribution and verification. Together, these examples demonstrate that several elements of authority-oriented governance are already observable across institutional contexts.

The documentary evidence therefore serves two purposes. First, it establishes that the conceptual dimensions underlying HACGI are empirically traceable in real institutional governance documents rather than being purely abstract constructs. Second, it provides contextual evidence for examining the extent to which existing policies specify human responsibilities, verification requirements, decision boundaries and accountability when AI is incorporated into academic activity.

Documentary Grounding Versus Institutional Measurement

The documentary evidence should not be interpreted as equivalent to a validated institutional HACGI score. The presence of a relevant policy provision does not, by itself, establish how consistently the provision is implemented, understood, enforced or experienced in practice. A university may formally assign responsibility to human actors while implementation remains uneven or may specify verification requirements without providing sufficient institutional capacity to support them. Accordingly, the study maintains a clear distinction between documentary grounding, construct operationalisation and quantitative validation. Documentary analysis establishes that the HACGI dimensions have identifiable institutional manifestations and provides a transparent basis for their operational definitions. It does not establish institutional performance levels or demonstrate that the corresponding governance arrangements operate effectively in practice.

The subsequent quantitative analysis therefore does not treat the documentary corpus as a source of empirical HACGI scores. Instead, synthetic data are used separately to examine the behaviour, robustness and sensitivity of the proposed MCDM architecture under alternative weighting and performance conditions. The use of synthetic observations allows the methodological properties of the HACGI framework to be examined without incorrectly attributing simulated scores to the 34 universities. This separation is essential to the empirical logic of the study: institutional documents ground the constructs; the traceability procedure establishes their observable policy manifestations; and synthetic data test the quantitative architecture. The documentary corpus consequently provides empirical plausibility and construct traceability while the MCDM and Monte Carlo procedures provide methodological evaluation rather than claims of validated institutional performance.

Research Design

The study adopts a two-component research design combining documentary empirical grounding with MCDM simulation.

Component 1: Documentary Empirical Grounding

Official institutional policies, guidelines and assessment documents are examined to determine whether governance provisions corresponding to the eight HACGI dimensions are observable in contemporary higher education. This establishes the institutional traceability and empirical plausibility of the proposed constructs, without assuming that the sampled universities have adopted HACG or HACGI.

Component 2: MCDM Simulation

Synthetic governance profiles are used to examine the computational behaviour of HACGI within a multi-criteria decision-making framework, including its sensitivity to alternative weights and governance conditions. The simulation evaluates the methodological coherence of the index rather than actual institutional performance.

The two components therefore serve distinct but complementary purposes: the documentary analysis establishes whether the constructs are observable in real institutional governance while the simulation examines whether those constructs operate coherently as a quantitative index. Neither component alone establishes external validity. The design consequently avoids conflating documentary evidence or synthetic results with validated institutional rankings.

Documentary Coding Protocol

The empirical application will rely primarily on official institutional sources to ensure that coding reflects formally recognised governance arrangements. Eligible documents include university-wide AI/GenAI policies, assessment regulations and guidance, academic-integrity policies, student and staff guidance, research-AI guidance, institutional AI governance frameworks and formal procedures or protocols. Third-party commentary, media reports, student-generated summaries and commercial AI guidance will not be treated as primary coding evidence.

For each coded provision, the researcher will retain the source document, version or publication date, relevant page or section and the substantive policy meaning. This ensures that each HACGI coding decision remains transparent, traceable and reproducible.

Unit of Analysis

The principal unit of analysis is the policy meaning unit, the smallest textual element that conveys a distinct governance provision. Depending on the document, this may be a sentence, clause, paragraph, definition, table entry, procedural instruction or responsibility statement.

Meaning-unit coding is preferred to treating an entire policy as a single observation given that governance provisions are often distributed unevenly within the same document. For example, a policy may specify disclosure requirements without defining who has authority to reject or override an AI recommendation. Document-level coding could conceal this distinction, whereas meaning-unit analysis allows each governance mechanism to be identified and evaluated independently.

Coding and Inter-Rater Reliability

A full institutional application of HACGI should employ at least two independent coders to enhance coding consistency and reduce subjective interpretation. Across 34 institutions and eight dimensions, this would generate 272 institution-by-criterion observations. Each coded observation should be traceable to its source document, version or date, page or section, relevant policy provision and coding rationale, thereby ensuring documentary transparency and auditability. Since, HACGI uses an ordinal coding scale, weighted Cohen's kappa (κ) is appropriate for assessing inter-rater reliability, as it distinguishes minor disagreements between adjacent categories from more substantial disagreements between distant categories. A complete empirical application should report weighted κ for each dimension, percentage agreement, identified disagreements, coder reconciliation procedures and final consensus scores. Cronbach's alpha is not required as a primary validation statistic. HACGI is conceptualised as a formative documentary index, in which its eight dimensions capture distinct and potentially non-interchangeable governance mechanisms. They are therefore not expected to function as reflective indicators of a single latent construct, making internal-consistency statistics such as alpha conceptually inappropriate as a prerequisite for index validation.

MCDM Methodology

Rationale

Institutional AI governance is inherently multi-dimensional. An institution may combine strong disclosure and professional accountability with moderate verification, weak delegation boundaries and limited policy adaptability. A single aggregate judgement can therefore obscure meaningful differences across governance dimensions. A multi-criteria decision-making (MCDM) approach is appropriate as it preserves these dimensions while providing a transparent framework for their systematic aggregation and comparison (Hwang & Yoon, 1981).

Selection of TOPSIS

The Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) ranks alternatives according to their relative proximity to a positive ideal and distance from a negative ideal solution (Hwang & Yoon, 1981). TOPSIS is particularly suitable for the present methodological demonstration for three reasons. First, the eight HACGI dimensions can initially be treated as beneficial governance criteria, where higher values represent stronger or more explicit governance specification. Second, its computational logic is transparent and readily interpretable. Third, its weighting structure allows criterion importance to be varied, making it well suited to subsequent sensitivity and Monte Carlo analysis.

Accordingly, TOPSIS is used not to claim that one university is objectively superior to another but to demonstrate how an empirically populated HACGI dataset could be translated into a transparent and reproducible MCDM framework. The resulting simulation therefore illustrates the analytical behaviour of the proposed index rather than constituting an institutional ranking.

Weighting Strategy

Criterion weighting is a critical methodological issue in MCDM. Although AHP can derive weights from expert pairwise comparisons (Saaty, 1980), such judgements must be based on genuine elicited preferences and cannot be retrospectively reconstructed or fabricated. As no expert comparison dataset was available, this study does not impose artificial AHP judgements.

Instead, equal weighting is adopted as a transparent baseline:

[
w_j=\frac{1}{8}=0.125
]

This assumption does not imply that the eight HACGI dimensions are substantively or normatively equivalent. It provides a neutral reference point in the absence of empirically elicited preferences. Future studies can replace these baseline weights with expert-derived weights obtained through systematic pairwise comparison.

TOPSIS Procedure

Let (X=[x_{ij}]) denote the decision matrix, where (i) represents a governance profile and (j) represents one of the eight HACGI dimensions. TOPSIS is implemented in six steps.

First, the decision matrix is vector-normalised:

[
r_{ij}=\frac{x_{ij}}{\sqrt{\sum_{i=1}^{n}x_{ij}^{2}}}
]

The normalised values are then weighted:

[
v_{ij}=w_jr_{ij}
]

The positive and negative ideal solutions are defined respectively as:

[
A^+={\max_i(v_{ij})}, \qquad
A^-={\min_i(v_{ij})}
]

The Euclidean separation of each profile from the two ideal solutions is calculated as:

[
D_i^+=\sqrt{\sum_j(v_{ij}-A_j^+)^2}, \qquad
D_i^-=\sqrt{\sum_j(v_{ij}-A_j^-)^2}
]

Finally, the relative closeness coefficient is:

[
C_i=\frac{D_i^-}{D_i^++D_i^-}
]

Higher (C_i) values indicate greater proximity to the positive ideal governance profile.

Monte Carlo Sensitivity Analysis

As MCDM rankings can be sensitive to criterion weights, the robustness of the baseline results is assessed through 10,000 Monte Carlo simulations. Each simulation draws an alternative weight vector from a symmetric Dirichlet distribution:

[
w^{(s)}\sim Dirichlet(20,\ldots,20)
]

The distribution constrains the eight weights to sum to one while allowing moderate departures from the equal-weight baseline.

TOPSIS is recalculated for every simulated weight vector. The analysis records mean rank, rank variability, top-rank frequency and Kendall's (\tau) relative to the baseline ranking. This procedure evaluates whether the resulting governance ordering remains stable under plausible changes in criterion importance, thereby providing a more rigorous robustness assessment than reliance on a single predetermined weighting scheme.

Simulation Design

As a fully independently coded 34 × 8 institutional governance matrix was not available, the MCDM experiment uses 34 synthetic governance profiles, each defined across the eight HACGI dimensions with ordinal scores ranging from 0 to 3. The profiles are intentionally anonymous and non-institutional. They do not represent Oxford, Harvard, Melbourne, IIT Delhi or any other specific university and must not be interpreted as empirical observations of those institutions. A fixed random seed is used to ensure reproducibility.

The simulation addresses a methodological rather than empirical question: if comparable HACGI scores were available for 34 institutions, would the proposed HACGI–TOPSIS architecture produce a coherent ordering and remain reasonably stable under alternative criterion weights? It does not determine which

real university has the strongest AI governance. That question requires systematic documentary coding and empirical institutional data.

Statistical Results

Association between HACGI and TOPSIS

Under the baseline equal-weight specification, the 34 synthetic governance profiles produced a very strong association between the additive HACGI score and the TOPSIS closeness coefficient.

Statistic	Result
Synthetic governance profiles	34
HACGI dimensions	8
Baseline weighting	Equal
Monte Carlo replications	10,000
Pearson correlation	$r = .998$
Spearman correlation	$\rho = .996$
Pearson p-value	$< .001$
Spearman p-value	$< .001$
Mean Kendall rank association	$\tau \approx .932$

The Pearson correlation was .998 ($p < .001$) while the Spearman correlation was .996 ($p < .001$), indicating that TOPSIS closely preserves the ordering generated by the additive HACGI score within the simulated setting. This result should, however, be interpreted as evidence of internal computational coherence, not external validity. Both measures are constructed from the same eight governance dimensions, making a strong association structurally expected. The finding therefore supports the consistency of the proposed HACGI–TOPSIS architecture but does not establish that HACGI accurately measures real-world institutional AI governance.

The appropriate conclusion is consequently modest but important. The simulation demonstrates that the proposed index and decision architecture behave coherently under the specified assumptions; it does not demonstrate empirical validity for actual universities.

Ranking Robustness

Across the 10,000 alternative-weight simulations, the mean Kendall rank association with the baseline ordering was approximately $\tau = .932$, with a median of $\tau = .936$ and a central simulation interval of approximately [.882, .968]. These results indicate substantial stability under moderate changes in criterion weights.

However, closely matched synthetic profiles were more sensitive to weighting changes. This is not a weakness of the index; rather, it reflects an important property of multi-criteria governance assessment. A robust index should preserve broad differentiation while allowing legitimate uncertainty where alternatives have similar profiles. The simulation therefore suggests that HACGI can provide stable broad differentiation while appropriately retaining uncertainty among closely matched governance profiles.

Interpretation of the MCDM Simulation

The simulation supports four methodological conclusions. First, the eight HACGI dimensions can be structured as a multi-criteria decision matrix. Second, TOPSIS can integrate these dimensions without requiring fabricated expert preferences. Third, equal weighting provides a transparent baseline when empirical preference data are unavailable. Fourth, Monte Carlo analysis makes the sensitivity of rankings to alternative weighting assumptions explicit.

The simulation does not establish that the 34 universities have particular HACGI scores, that one institution is better governed than another, that HACGI predicts institutional outcomes or that the proposed dimensions and their weights possess external or empirically optimal validity.

The distinction between methodological validation and substantive institutional validation is therefore fundamental: the simulation evaluates how the proposed architecture behaves, not how actual universities perform.

Documentary Findings

Permission Is More Visible Than Authority

The documentary evidence indicates that contemporary university policies increasingly specify whether AI is permitted, where and when it may be used, which assessment activities allow it, what users must disclose and which uses are prohibited. Evidence from Oxford, UCL, UNSW, Monash, Queensland, Princeton, Harvard, Cornell and other institutions illustrates this growing emphasis on defining permissible AI use. Yet permission alone does not establish who decides whether an AI recommendation should be accepted, modified, rejected or overridden. This distinction captures the central empirical relevance of the Authority Specification Gap: institutions may regulate AI use without fully specifying the distribution of decision authority surrounding that use.

Human Oversight Becomes Meaningful When It Includes Decision Rights

Some institutional guidance moves beyond generic references to "human oversight." King's College London's guidance on AI-assisted marking and feedback provides a particularly clear example by distinguishing AI support from the replacement of human expertise and academic judgement. The important feature is not simply the presence of a human in the workflow but the preservation of human decision authority and professional responsibility.

This distinction is central to HACG: meaningful human oversight depends on the capacity to evaluate, modify, reject or override AI outputs-not merely on human participation in the process.

Verification Is Epistemic Governance

UNSW, Queensland, Cornell and other institutions increasingly require users to review or verify AI-generated outputs. Verification, however, extends beyond technical error checking. It concerns a deeper epistemic question: who determines whether an AI-generated claim is sufficiently accurate, reliable, relevant and contextually appropriate to inform an academic judgement?

Verification therefore represents a form of epistemic governance, directly connecting institutional procedures with the HACGI dimensions of Verification Requirement and Epistemic Accountability.

Disclosure Is Not Accountability

Disclosure establishes that AI was used; traceability clarifies how it was used; accountability identifies who remains responsible for the resulting judgement. These are related but distinct governance functions. A sophisticated disclosure regime can therefore coexist with weak specifications of decision rights and responsibility. HACGI deliberately separates these functions by asking three distinct questions: Was AI used? How did AI participate? Who remains accountable for the resulting decision?

High-Stakes Activities Require Differentiated Governance

AI applications do not carry equivalent consequences. Generating classroom examples is fundamentally different from using AI to influence grading, academic progression, academic-integrity investigations, admissions, disciplinary decisions or research evaluation. This difference is captured by the Human Authority Gradient: as the consequences and potential irreversibility of AI influence increase, governance should place stronger decision authority, verification capacity and accountability with human academic actors.

Recent research examining more than 750 essays across three UK universities reported that frontier AI systems matched human degree classifications only 35–65% of the time. Although this finding does not establish that AI should be excluded from assessment, it illustrates the uncertainty surrounding AI-supported high-stakes evaluation and reinforces the need for meaningful human verification, judgement and accountability.

Discussion

From Human-in-the-Loop to Human-in-Authority

The central conceptual contribution of HACG is its distinction between human-in-the-loop and human-in-authority. Human-in-the-loop merely indicates that a person participates somewhere in an AI-mediated process. Human-in-authority requires something stronger: the human must retain meaningful decision rights, interpretive authority, verification responsibility, capacity to modify or reject AI outputs, override authority and ultimate accountability.

This distinction is particularly important in higher education, where AI-supported decisions can affect assessment, academic progression, student rights, professional judgement and institutional reputation. The governance objective should therefore not be simply to keep a human present but to ensure that the human retains the authority and practical capacity to exercise independent judgement.

HACG as an Extension of Responsible AI Governance

Existing responsible-AI frameworks emphasise principles such as fairness, transparency, privacy, safety, accountability and human oversight. HACG extends this agenda by making authority allocation an explicit object of governance.

Responsible-AI principle	HACG question
Human oversight	Who retains final decision authority?
Transparency	What must be disclosed and documented?
Accountability	Who remains responsible for the decision?
Fairness	Who evaluates potentially harmful or biased outputs?
Safety	Which decisions require limits on AI delegation?
Explainability	Who determines whether an explanation is adequate?

HACG thus moves responsible AI governance from broad principles toward the institutional allocation and exercise of cognitive authority.

Verification as Epistemic Governance

Within HACG, verification is more than factual checking. It involves assessing accuracy, evidence quality, disciplinary context, uncertainty, missing information, bias, hallucination, methodological limitations and the consequences of error. The required depth of verification should increase with the significance of the decision.

This extends established work on appropriate reliance and human interaction with automation (Lee & See, 2004; Parasuraman et al., 2000) into institutional AI governance. Given that generative AI can produce plausible but inaccurate or contextually inappropriate outputs, verification becomes an epistemic responsibility of the human decision-maker, not merely a technical quality-control step.

High-Stakes Governance

AI-supported decisions should not be governed uniformly. Governance intensity should increase with the consequence, contestability and potential irreversibility of AI influence. For low-consequence applications, routine human review may be sufficient. Moderate-consequence applications warrant explicit verification and documentation. High-consequence decisions require stronger safeguards, including clearly assigned human authority, independent verification, documentation, explanation, appeal mechanisms, role-specific accountability and meaningful limits on automation.

This differentiated approach is consistent with broader human-centred AI governance in education (Miao & Holmes, 2023; Crompton et al., 2026; OECD, 2026) and reinforces the central HACG proposition: the greater the consequences of AI-mediated judgement, the stronger the human authority required to govern it.

Disclosure, Traceability and Accountability

Disclosure, traceability and accountability are related but distinct governance mechanisms. Disclosure establishes whether AI was used; traceability establishes how it participated; and accountability identifies who remains responsible for the resulting decision.

A governance system that answers only whether AI was used is therefore incomplete. HACGI separates Disclosure and Traceability (DT) from Professional Accountability (PA) and Human Decision Authority (HDA), recognising that transparency does not necessarily establish who holds final decision rights. An institution may have strong disclosure requirements while leaving responsibility for the resulting judgement insufficiently specified.

Professional Accountability

AI governance ultimately remains embedded in professional responsibility. Faculty, researchers, academic administrators and other institutional actors operate within disciplinary, ethical and professional norms. AI may support their work but consequential judgements should remain attributable to an authorised human actor or institutional role.

Meaningful professional accountability requires more than formal responsibility. It depends on adequate AI literacy, time and capacity for verification, institutional support, clear delegation boundaries, reliable tools, protection of legitimate professional judgement and the authority to reject or override AI recommendations. HACG therefore connects authority specification with the broader literature on AI literacy, professional judgement and responsible AI readiness (Long & Magerko, 2020; Kaur & Pundir, 2026a, 2026b).

Policy Adaptability

AI governance cannot be treated as a one-time policy exercise. As AI capabilities, deployment models, interfaces and institutional applications evolve, static policies can rapidly become misaligned with practice. Policy adaptability is therefore itself a governance capability.

A mature governance system should specify who monitors technological developments, who reviews policy, when formal review occurs, what events trigger interim review, how changes are communicated and how incidents feed back into policy revision. This becomes increasingly important as AI systems move beyond isolated recommendations towards more agentic, multi-step forms of participation.

PEVJA-R: A Practice-Oriented Governance Protocol

To translate HACG from institutional governance into everyday professional practice, this study proposes PEVJA-R: Purpose, Engage, Verify, Judge, Account, Reflect.

Purpose asks why AI is being used and whether its use is justified for the educational, research or administrative objective.

Engage clarifies the role assigned to AI-whether it is assisting, recommending, evaluating or performing an action.

Verify requires independent evaluation of the output, including its accuracy, evidence, context, limitations, bias, uncertainty and potential consequences.

Judge places the final academic or professional judgement with the authorised human actor, who applies disciplinary knowledge and professional standards.

Account identifies who owns the resulting decision and ensures that responsibility remains attributable to an identifiable human actor or authorised institutional role.

Reflect examines whether AI genuinely supported human judgement or inadvertently displaced it, including whether delegation remained within defined boundaries, verification was adequate, unexpected risks emerged or policy revision is required.

PEVJA-R therefore complements HACGI at a different level of analysis. HACGI operates primarily at the institutional governance level while PEVJA-R translates those governance principles into professional practice.

HACG and AI Readiness

HACG should be distinguished from AI readiness. AI readiness concerns whether institutions and professionals possess the capabilities and conditions required for responsible AI adoption. HACG concerns how authority is structured and exercised once AI becomes part of cognitive work. An institution may be highly AI-ready yet remain cognitively under-governed if it has substantial AI capability but unclear decision rights. Conversely, clearly defined authority structures may be ineffective when professionals lack the knowledge, time or institutional capacity to exercise independent judgement. AI readiness and HACG are therefore complementary rather than competing constructs.

This distinction also provides a basis for future empirical research linking institutional AI-readiness measures with HACGI. Such work could examine whether greater organisational readiness is associated with more explicit specification of human decision authority, verification responsibility and accountability.

Implications for Higher-Education Policy

Institutional AI policies should distinguish four governance levels: permission, delegation, verification and authority. Permission asks whether AI may be used; delegation specifies what AI is allowed to do; verification identifies who must evaluate its outputs; and authority establishes who makes the final decision.

Many existing policies appear strongest at the level of permission. HACG argues that governance maturity requires movement beyond permission towards explicit specification of functional boundaries, verification duties, responsible roles, override rights, accountability and review or appeal mechanisms.

Implications for Assessment

Assessment governance requires task-specific rather than generic AI rules. Policies should specify permitted and prohibited functions, documentation requirements, verification expectations, evidence of independent judgement, responsibility for errors, the extent to which AI may influence evaluation and the human authority over the final grade.

Where assessment consequences are substantial, statements such as “AI may be used responsibly” are insufficient. Effective governance must specify how AI may participate in the particular task and where human judgement remains determinative.

Implications for Faculty Development

Faculty development should extend beyond technical AI literacy towards AI governance literacy. Academic professionals need the capacity to recognise inappropriate delegation, assess AI reliability, verify outputs, calibrate reliance, identify high-stakes uses, preserve professional judgement, exercise override rights and document AI involvement.

Training should therefore address not only how to use AI but also when not to delegate, when verification must intensify and when human judgement must remain decisive.

Implications for Indian Higher Education

HACG has particular relevance to Indian higher education, where institutions vary considerably in digital infrastructure, faculty capability, institutional autonomy, technological maturity and governance capacity. A uniform model of AI adoption may therefore be neither feasible nor desirable.

An authority-centred approach allows institutions to determine where AI provides useful assistance, where delegation should be restricted, which decisions require stronger safeguards, who retains final authority and what level of verification is appropriate to institutional capacity. In this respect, HACG complements the responsible AI-readiness perspective of Kaur and Pundir (2026a, 2026b) by addressing authority allocation rather than readiness alone.

Proposed Governance Maturity Bands

If future institutional coding provides sufficiently validated data, HACGI percentages may be interpreted descriptively through provisional maturity bands:

HACGI (%)	Descriptive interpretation
0–24	Minimal specification
25–49	Emerging governance
50–69	Developing governance
70–84	Advanced governance
85–100	Highly specified governance

These thresholds are provisional interpretive categories, not validated performance standards. They should not be used to classify institutions as "good" or "poor" until external evidence demonstrates that the bands correspond to meaningful differences in governance practice or outcomes.

Methodological Contribution

The study makes six principal methodological contributions. It translates the abstract concept of human authority into observable documentary dimensions; distinguishes permission, delegation, verification and authority; conceptualises HACGI as a formative governance index rather than a reflective psychometric scale; integrates HACGI with TOPSIS without fabricating expert preferences; uses Monte Carlo analysis to examine ranking sensitivity; and maintains a clear separation between empirical documentary evidence and synthetic MCDM simulation. This final distinction is particularly important. It strengthens the methodological contribution of the study without converting simulated results into unsupported claims about actual institutional performance.

Robustness and Statistical Interpretation

The simulation results provide evidence at three complementary levels. First, the strong association between HACGI and TOPSIS indicates aggregation coherence: under the simulated conditions, TOPSIS largely preserves the broad ordering generated by the additive index. Second, the high mean Kendall's τ demonstrates substantial weight robustness under moderate perturbations. Third, the greater sensitivity of closely matched profiles highlights ranking uncertainty rather than undermining the framework.

A responsible governance index should therefore not conceal uncertainty behind a single deterministic ranking. Its value lies in providing stable broad differentiation while making sensitivity and uncertainty visible where the evidence warrants caution.

Limitations

This study has several limitations. First, the MCDM analysis is based on synthetic institutional profiles and therefore demonstrates the computational behaviour and robustness of the HACGI–TOPSIS framework rather than actual university performance. Actual institutional scores require systematic coding of the full 34×8 documentary matrix. Second, the purposive and unevenly distributed institutional corpus provides contextual breadth rather than national representativeness. Third, the absence of expert pairwise-comparison data prevented the use of AHP; the equal-weight baseline should therefore be regarded as a methodological reference rather than an empirically established weighting structure.

A further limitation concerns the distinction between formal and enacted governance. Documentary analysis identifies what institutional policies specify but cannot establish whether those provisions are understood, followed, enforced or reflected in practice. Similarly, the strong HACGI–TOPSIS association demonstrates computational coherence rather than construct or external validity, since both are derived from the same governance dimensions. Future research should therefore validate the framework through independent coding, inter-rater reliability, expert assessment, empirical weighting and outcome-based evidence. Validation should also examine whether the current coding categories and equal-weight structure remain appropriate across institutional and high-stakes contexts.

Future Research and Empirical Validation

Future research should advance HACGI from methodological demonstration to empirical institutional assessment. This requires systematic collection and archival of official documents, independent coding of all eight dimensions by trained coders, inter-rater reliability testing and expert evaluation of the framework's conceptual relevance and completeness. Once genuine expert judgements are available, AHP or another empirically justified weighting method can be introduced, followed by TOPSIS and sensitivity analysis. Validation should also extend beyond institutional ranking. Future studies should examine whether stronger human-authority provisions are associated with meaningful outcomes such as policy compliance, clarity of decision responsibility, procedural fairness, professional trust, AI-related incidents and stakeholder confidence. This would allow assessment of criterion validity and clarify whether formal governance specification translates into enacted and perceived governance.

Accordingly, future research should extend HACGI from policy specification to enacted governance, perceived governance, and governance outcomes, providing stronger empirical evidence of its theoretical and practical value.

Conclusion

The central governance challenge posed by generative AI in higher education is no longer simply whether AI should be permitted but who retains authority when AI advises. This study develops Human-AI Cognitive Governance (HACG) as an authority-centred framework for addressing this question. Its central theoretical contribution is the Authority Specification Gap: policies may regulate AI use without clearly specifying who may accept, question, modify, reject or override AI-generated outputs. The Human Authority Gradient further proposes that human authority should strengthen as the consequences and irreversibility of AI-influenced decisions increase and contestability decreases. The Human-AI Cognitive Governance Index (HACGI) operationalises this perspective through eight dimensions covering decision authority, delegation, verification, epistemic accountability, high-stakes governance, disclosure and traceability, professional accountability and policy adaptability. Documentary evidence shows that these governance concerns are already observable in higher-education policies while also demonstrating that AI permission, delegation, oversight and human authority represent distinct governance conditions.

The MCDM analysis further demonstrates the computational coherence and robustness of the proposed index. Pearson's $r = .998$ establishes near-perfect linear correspondence between HACGI and TOPSIS scores while Spearman's $\rho = .996$ shows that the resulting rank order is almost completely preserved. Monte Carlo analysis across 10,000 alternative weighting configurations produced a mean Kendall's $\tau \approx .932$, indicating substantial ranking stability under moderate weighting uncertainty. Together, these findings show that the framework is not materially dependent on a single equal-weight specification while the less-than-perfect rank stability appropriately signals greater uncertainty among closely matched profiles. These statistics establish methodological robustness, not external validity. They demonstrate that HACGI behaves coherently as an MCDM architecture and remains broadly stable under weighting variation; they do not establish that particular universities perform better or that stronger HACG produces better outcomes.

The study therefore contributes theoretically, conceptually, methodologically and empirically by shifting responsible AI governance from permission to explicit allocation of authority. Meaningful human oversight requires more than human presence in an AI-enabled workflow; it requires the institutional capacity and authority to understand, verify, question, modify, reject, override and ultimately take responsibility for consequential decisions. As AI systems become increasingly capable and potentially agentic, institutions will need to specify who may delegate, who must verify, who may override, who decides and who remains accountable. Future research should replace synthetic profiles with independently coded institutional evidence, establish expert and inter-rater validation, introduce empirically grounded weighting and test whether stronger authority-centred governance is associated with better institutional and educational outcomes. AI participation should not silently become AI authority.

References

- Alfiras, M. I. I., Emran, A. Q. & Mohamed, A. M. (2026). Ethics and governance of generative AI in education: A systematic review on responsible adoption. *Discover Education, 5*, 37. https://doi.org/10.1007/s44217-025-01051-y
- Bacalso, F. D., Villamor, F. E., Paclipan, C. P. orquia, C. J. B., Rendon, I. P., Albiso, A. T. & Himang, C. M. (2026). Institutional approaches to generative AI management in higher education: A systematic review. *Frontiers in Education, 11*, 1814426. https://doi.org/10.3389/feduc.2026.1814426
- Baggia, A., Tratnik, A. & Brezavšček, A. (2026). University teachers' perceptions, attitudes and practices with generative AI: A systematic literature review with meta-analytic component. *Quality & Quantity, 60*, 14107–14139. https://doi.org/10.1007/s11135-026-02717-x
- Barus, O. P., Hidayanto, A. N., Eitiveni, I., Phusavat, K. & Sudibjo, N. (2026). Identifying key components and stakeholders for generative AI governance in higher education: A systematic literature review. *Interactive Learning Environments, 34*(6), 4316–4339. https://doi.org/10.1080/10494820.2025.2596058
- Bittle, K. & El-Gayar, O. (2025). Generative AI and academic integrity in higher education: A systematic review and research agenda. *Information, 16*(4), 296. https://doi.org/10.3390/info16040296
- Crompton, H., Burke, D., Nickel, C., Bozkurt, A., Miao, F., Sharples, M., Greene, J. A., Parsons, D., Gill-Simmen, L., Edmett, A., Pegrum, M., de Waard, I., Bonk, C. J., Garcia, M. B., Curry, J. H., Lindsey, L. A., Yang, M., Marshall, S., Bali, M., Deutsch, N., le Roux, S., Benali, M., Samsudin, M. A. B., Tinmaz, H., Bernacki, M. L., van Wyk, M., Singh, L., Chigona, A., Eaton, L., Xiao, J., Velander, J., Kim, J., Bellas, F., Rajalakshmi, R., Machado, A. de B., Palalas, A. & Yu, S. (2026). Governing generative AI in higher education: A global Delphi study on policy and practice. *International Journal of Educational Technology in Higher Education, 23*, Article 21. https://doi.org/10.1186/s41239-026-00602-z
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use and user acceptance of information technology. *MIS Quarterly, 13*(3), 319–340.
- Hwang, C. L. & Yoon, K. (1981). *Multiple attribute decision making: Methods and applications: A state-of-the-art survey*. Springer.
- Jiang, X. & Abdullah, Z. (2026). AI-enabled governance in higher education: A systematic review of applications, outcomes and emerging implications. *Frontiers in Education, 11*, 1856440. https://doi.org/10.3389/feduc.2026.1856440
- Kaur, K. & Pundir, P. S. (2026a). *Measuring AI-enabled educational readiness in higher education: Scale development, higher-order structural model and institutional indexing*. https://doi.org/10.5281/zenodo.22125460
- Kaur, K. & Pundir, P. S. (2026b). Responsible AI readiness in higher education: A multilevel framework integrating faculty capability, professional judgement and institutional conditions. *International Journal for Multidisciplinary Research, 8*(5).
- Lee, J. D. & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392
- Long, D. & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–16). Association for Computing Machinery. https://doi.org/10.1145/3313831.3376727
- Miao, F. & Holmes, W. (2023). *Guidance for generative AI in education and research*. UNESCO. https://doi.org/10.54675/EWZM9535
- OECD. (2026). *Policies supporting responsible and systematic generative AI adoption in higher education*. OECD Education Spotlights No. 23. OECD Publishing. https://doi.org/10.1787/c4e5621f-en
- Parasuraman, R., Sheridan, T. B. & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans, 30*(3), 286–297. https://doi.org/10.1109/3468.844354

- Qian, Y. (2026). Governing generative AI in higher education: Emerging policy approaches and support ecosystems at innovative U.S. universities. *International Journal for Educational Integrity, 22*, Article 25. https://doi.org/10.1007/s40979-026-00233-x
- Saaty, T. L. (1980). *The analytic hierarchy process: Planning, priority setting, resource allocation*. McGraw-Hill.
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–478.