World Journal of Educational Studies
Print ISSN: 2959-9989
Online ISSN: 2959-9997
DOI: https://doi.org/10.61784/wjes3183

ARTIFICIAL INTELLIGENCE-ENABLED DIGITAL GOVERNANCE IN HIGHER EDUCATION: OPERATING MECHANISMS, RISK BOUNDARIES, AND AN ADAPTIVE GOVERNANCE FRAMEWORK

WeiJia Zou, ZeYu Wang*
School of Public Administration, Guangzhou University, Guangzhou 510006, Guangdong, China.
**Corresponding Author: ZeYu Wang*

Abstract: Artificial intelligence is increasingly becoming a part of the governance of universities. While there has been a significant amount of research discussing the applications of artificial intelligence in higher education and the ethics of artificial intelligence, there has been less discussion of how artificial intelligence impacts the governance of universities. This article reviews 36 sources relating to artificial intelligence and higher education to develop a framework for understanding artificial intelligence-enabled digital governance in higher education. The article identifies four mechanisms through which artificial intelligence functions within higher education governance: data-driven sensing, cognitive augmentation in decision-making, coordinated execution, and recursive institutional learning. These mechanisms create risks for higher education institutions, including data extraction and purpose drift, algorithmic bias and opacity, automation dependence and responsibility diffusion, and misalignment with educational values. To manage these risks, the article develops an adaptive governance framework that includes risk-tiered authorization, impact assessment, human oversight, decision logging, audit, contestability, and institutional updating. The framework is applied to major university functions. The article argues against the use of artificial intelligence as an autonomous decision maker in higher education and against presenting artificial intelligence as a neutral productivity device. The article suggests that the appropriate role for artificial intelligence in higher education is to strengthen institutional judgment within a dual closed loop of feedback. This article contributes to the field of digital governance and provides a framework through which universities may use artificial intelligence while protecting student rights, academic values, and public trust in higher education.
Keywords: Artificial intelligence; Digital governance; Higher education; Adaptive governance; Algorithmic accountability

1 INTRODUCTION

Artificial intelligence has made its way into universities in a variety of ways. From tutoring and grading systems to admissions and application processing, universities make use of AI in a variety of ways. The individual systems are often adopted by different offices within the university for different reasons. Yet, the combined impact of these systems is felt at the institutional level. Once AI is used to classify students, recommend interventions, rank risks, distribute attention to individuals or programs within the university, or generate content for the institution, AI is involved in the creation of university knowledge and the allocation of discretion within the institution [1]. Therefore, the question of whether or not AI will be used by universities is no longer relevant in many instances. Instead, the question of what kind of governance is created through the use of AI within the university is the new question to be asked of higher education institutions.

Digitalization and digital governance are terms that are often used to discuss the changes that are occurring within universities as a result of AI. Digitalization refers to the movement of information into a digital format. Digital governance, on the other hand, relates to the way in which public institutions and the public value that is created by those institutions are reshaped through the use of digital technologies [2]. Many of the discussions regarding AI within the public sector relate to these two ideas. AI can improve the capacity and responsiveness of the public sector, but it can also introduce issues into these organizations. Higher education is facing similar issues [3-5]. AI can identify learning challenges for students, reduce the workload of administrative staff, and improve the services that are offered by universities. However, the outputs of AI systems are often converted into decisions that have an impact upon students and staff of the institution.

The current literature has advanced in three directions. Systematic reviews map the growth of AI applications in higher education, especially prediction, assessment, personalization, and support systems [6,7]. Work on generative AI discusses both its promise and its risks for universities [8]. A further body of ethical research sets out principles and safeguards for responsible AI use in education, including ethics, bias, privacy, and explainability [9]. These conversations, however, still do not meet often enough [10]. Application studies may describe what AI can do without examining how ethical principles should constrain it. Ethics studies may state values without showing how they become institutional procedures. Policy and governance studies also point to the need for university-level frameworks and more specific guidance rather than scattered teaching-use rules [11].

This article begins to fill that gap in the literature. Instead of considering AI-enabled digital governance as a loose collection of tools, this article considers such an arrangement from a socio-technical perspective. The article aims to answer three main questions: through what mechanisms does AI change the governance capacity of universities? What risk boundaries emerge with the inclusion of AI in decision-making and service delivery? And, what kind of governance architecture for AI-enabled digital governance in universities can enable innovation yet preserve the values of education, legal responsibility, and public trust in university education? The article creates a framework based on socio-technical systems theory, public value theory, and adaptive governance theory. This framework is then applied to the five main domains of AI-enabled digital governance in universities: student services, teaching and assessment, quality assurance, research administration, and campus risk management.

AI articles usually have a few main contributions. The first is usually theoretical. The second is usually analytical. The third is usually practical. In this article, the authors have stated that their contributions include the theoretical, analytical, and practical aspects of AI. The theoretical aspect of their contribution is the connection between AI capability and institutional discretion and accountability. The analytical aspect of their contribution is the proposal of a model that outlines the mechanisms by which one AI capability can create both public value and harm to the institution that created that AI. Finally, their practical contribution is the proposal of a governance framework for the institution to manage its AI projects. The goal of the authors is not to hold back AI adoption in institutions, but to ensure that the adoption of AI in institutions is educationally purposeful, institutionally controllable, and publicly defensible.

2 THEORETICAL BASIS AND CONCEPTUAL DEFINITION

2.1 Core Concepts

AI articles usually have a few main contributions. The first is usually theoretical. The second is usually analytical. The third is usually practical. In this article, the authors have stated that their contributions include the theoretical, analytical, and practical aspects of AI. The theoretical aspect of their contribution is the connection between AI capability and institutional discretion and accountability. The analytical aspect of their contribution is the proposal of a model that outlines the mechanisms by which one AI capability can create both public value and harm to the institution that created that AI. Finally, their practical contribution is the proposal of a governance framework for the institution to manage its AI projects. The goal of the authors is not to hold back AI adoption in institutions, but to ensure that the adoption of AI in institutions is educationally purposeful, institutionally controllable, and publicly defensible.

In higher education, digital governance refers to the institutional arrangement through which data, algorithms, and humans make educational decisions. Four points follow from this definition. First, governance is about authority, not technical aspects of digital systems. Second, governance includes formal and informal decisions within higher education. Third, governance covers the entire lifecycle of digital systems in higher education. Fourth, the legitimacy of digital governance in higher education depends upon educational and public values in addition to technical metrics like accuracy and cost.

AI-enabled digital governance means that artificial intelligence extends the sensing, analysis, coordination, and learning of a digital institution under its human authority. The word enabled is important here as well. AI can extend the capabilities of a university to notice and process information, but it should never replace the institution's legal and ethical responsibilities [12]. The object of digital governance of an institution is not the algorithm alone but the whole arrangement of datasets, models, interfaces, routines, judgments, vendors, and stakeholders.

The standard by which the arrangement should be assessed is public value. Value in higher education goes beyond considerations of speed and efficiency. Value includes considerations of educational quality, equal access and treatment, academic freedom, student development, privacy, fairness in procedures, and trust in the institution. Even if an AI system performs technically well, it may be unacceptable if it damages these values. The value created by a governance design that places caution on the use of AI may create value by making the AI use explainable, contestable, and trustworthy despite the added cost in procedures [13,14].

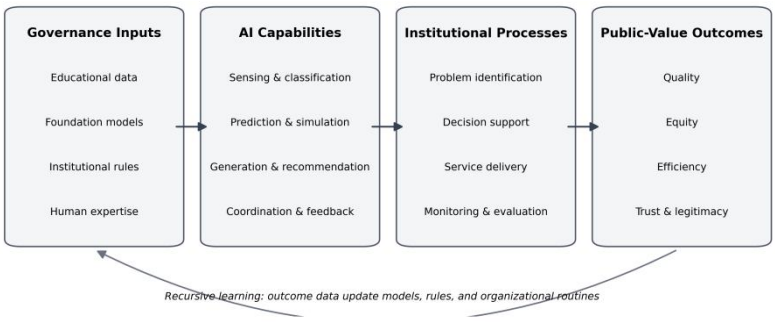

Figure 1 Analytical Framework of AI-Enabled Digital Governance in Higher Education

2.2 Theoretical Support

Socio-technical systems theory challenges the idea that technology produces stable effects on its own. The operational

system produces performance through its combination of technical elements and human activities and institutional guidelines and routine operational methods. The development of AI governance depends on this essential element because organizations need more than accurate models to establish their operational quality. A model which shows strong validation results will fail when used in practice because of various factors which include unstandardized data definitions and staff misinterpretation of confidence scores and user intentional behavior modifications and manager implementation of recommendations as mandatory commands. The socio-technical perspective lets us study how people interact with technology systems and the institutional rules which control their behavior in their operational environment.

The theoretical framework of public value includes its normative elements which support the argument. Universities operate as mission-oriented organizations which perform knowledge generation and personal development and learning certification and community service. AI systems need to achieve their intended goals through implementation systems which establish equitable advantage and disadvantage distribution. The perspective shows that decision-making based on efficiency and individual needs does not meet all necessary requirements. A system which operates with high efficiency but produces negative effects on students who lack complete data and restricts academic evaluation produces more damage than benefits. The evaluation process requires assessment of rights together with equity and legitimacy and mission alignment.

Adaptive governance introduces a dynamic element to the governance of AI systems. Because these systems and their environments are constantly changing, adaptive governance is essential. Adaptive governance permits experimentation within limits, continuous monitoring and adaptation of the AI system. For AI systems, adaptive governance requires the establishment of thresholds and owners of those thresholds, as well as the ability to review the system and to suspend or retire it, if necessary [15].

Together, these theories lead to a central proposition regarding the value of AI in higher education: the value of AI depends upon the institutional quality of the human-machine arrangement. Socio-technical systems theory explains how the capability of the human-machine arrangement is produced; public value theory clarifies what the capability of that arrangement should serve; and adaptive governance theory explains how the arrangement can be corrected to provide a better value to the public. Figure 2 presents this theoretical foundation for the value of AI in higher education.

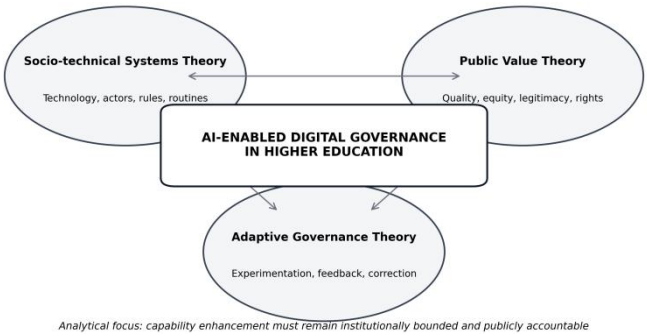

Figure 2 Theoretical Support for AI-Enabled Digital Governance

3 RESEARCH DESIGN AND METHODS

3.1 Overall Research Framework

The research implements a theory-building approach which unites integrative review with comparative policy analysis and scenario-based institutional deduction methods. The analysis does not focus on calculating one specific causal effect. The framework development process aims to create an explanatory system which links technical operations with organizational systems and risk management parameters and governance frameworks. Because the problem of study spans public administration and digital governance and AI ethics and educational technology and learning analytics and higher-education policy domains, an integrative review is appropriate. Comparative policy analysis enables the researchers to detect the fundamental principles of regulation which are present in the different frameworks studied. The scenario deduction process enables the researchers to convert those principles into operational organizational structures for a university.

Five phases of work were completed in the research. The first phase involved compiling a list of 36 essential sources to include in the study. These sources included 15 studies on AI digital governance, 15 studies on AI in higher education and educational ethics, and 6 documents related to institutional and regulatory frameworks. The research process started with open coding of the sources to identify AI capabilities and institutional effects and risk types and governance principles and implementation instruments. The third phase involved axial synthesis which linked recurring capabilities to the organizational processes they change. The research team employed scenario deduction to determine if university application categories would identify governance needs during their evaluation of typical university applications. The categories found their place within an adaptive governance framework which structured them according to system risk levels and their developmental stages.

The study uses analytical generalization to support its findings instead of statistical generalization methods. The

research framework achieves its validity through the alignment of conceptual elements with multiple data sources which produce distinct results. A functional framework needs to show how AI produces varying impacts in multiple environments and it should detect clear governance breakdowns and produce separate control needs for systems with different severity levels. The method requires researchers to follow a clear path which connects their literature review and policy framework studies to their institutional design development.

3.2 Source Selection and Coding

The research selected governance literature to study digital transformation, public-sector AI, administrative discretion, algorithmic regulation and AI governance, AI ethics principles, algorithmic accountability and auditing, fairness in sociotechnical systems, transparency in public decision-making, and adaptive governance. The educational research field studies AI applications in higher education, historical directions in AI in education, ethics and bias, privacy in learning analytics, explainable AI in education, generative AI and ChatGPT in education, AI policy and responsible adoption in universities, academic integrity, and quality assurance. The policy corpus includes UNESCO ethics guidance, UNESCO guidance for generative AI in education and research, the NIST AI Risk Management Framework, OECD guidance on digital education, the European Union AI Act, and China's Interim Measures for the Management of Generative AI Services.

Coding took place at three different levels of analysis. At the functional level, codes were developed to understand the functions of sensing, prediction, generation, recommendation, automation, and feedback. At the institutional level, the codes related to changes in information asymmetry, discretion, coordination, resource allocation, and responsibility. At the normative level, the codes related to the concepts of privacy, fairness, transparency, human agency, safety, contestability, and public value. By using a multi-level approach for coding the data, it avoids the issue of treating a technical feature as an outcome of governance. For instance, sensing is a technical feature of the AI system, while risk prioritization is an institutional use of the AI system, and the issue of unequal false-positive rates is a normative problem of the AI system.

Two analytical chains emerged from the coding of the responses. The first analytical chain follows the creation of value from the use of data and AI, while the second analytical chain follows the creation of risks from the use of data and AI. These two chains are intentionally symmetrical, as many of the benefits of data and AI also lead to associated risks. For instance, the ability to strongly predict an individual's needs may allow for earlier support to that individual, yet it may also lead to the profiling of that individual. Similarly, the ability to automate the delivery of a service may lead to improved service delivery, yet it may also lead to the weakening of the responsibility of the organization that is providing that service.

3.3 Analytical Structure

The final analytical structure contains four operating mechanisms and four matching risk boundaries. The mechanisms consist of data-driven sensing and cognitive augmentation and coordinated execution and recursive institutional learning. The risk boundaries consist of four main areas which include data extraction and purpose drift and algorithmic bias and opacity and automation dependence and responsibility diffusion and educational value misalignment. The categories exist beyond single boxes because they extend across multiple fields. The process follows a sequence where data practices determine model outputs which then affect human decision-making to produce organizational actions that generate outcomes which will affect upcoming data and rules.

Five institutional scenarios were used to evaluate the framework. Each scenario was examined to see if it could be assigned a consequence level, a responsible owner, a minimum human-review requirement, an audit trail, and a redress mechanism. Those that could not meet these conditions were considered to be institutionally immature. Figure 3 presents the research process, and Table 1 summarizes the corpus and coding dimensions.

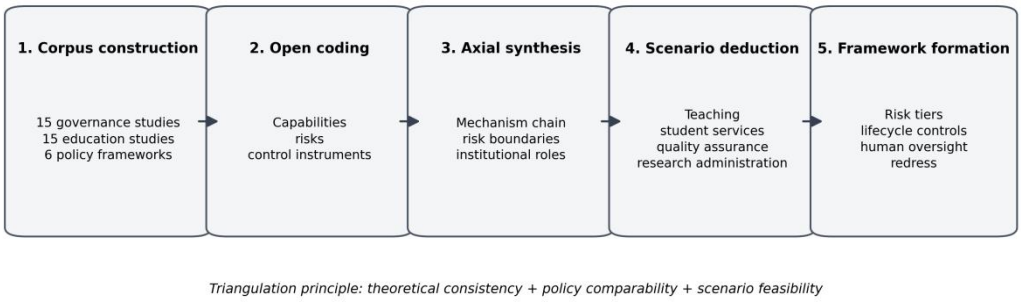

Figure 3 Research Design and Analytical Procedure

Table 1 Corpus Structure and Coding Dimensions

Corpus Category	Number	Main Topics	Coding Output
AI and digital governance	15	Digital transformation; public-sector AI; discretion; transparency; fairness; audit	Institutional mechanisms; accountability failures; control instruments
AI in higher education	15	Applications; learning analytics; generative AI; policy; integrity; quality assurance	Educational scenarios; stakeholder risks; mission constraints
Policy and regulatory frameworks	6	UNESCO; NIST; OECD; EU AI Act; Chinese regulation	Risk principles; lifecycle duties; human oversight; redress
Total	36	Cross-disciplinary integrative corpus	Four mechanisms; four risk boundaries; adaptive framework

4 OPERATING MECHANISMS OF AI-ENABLED DIGITAL GOVERNANCE

4.1 Data-driven Sensing and Problem Identification

The first mechanism is expanded institutional sensing. Universities have identified problems through various methods for decades: surveys, reports, complaints, and observation of lagging indicators. AI can analyze massive amounts of data and recognize patterns that would be difficult for humans to detect. For instance, university systems like learning-management logs, service records, text analytics, and anomaly detection can all provide information about problems within the institution. The main benefit to institutional governance is in gaining awareness of problems as they occur, rather than after the fact.

The act of sensing is never neutral. Every sensing system makes decisions about what is a signal, what population to make visible, and what variation within that sensing system to focus upon. A student that avoids a learning platform may appear to be disengaged from the learning, a classifier may interpret different types of language as negative sentiment, and data from administrative processes may reflect existing inequalities within an institution. Thus, there must be a separation between sensing and diagnosing an issue. While AI can sense that a pattern exists within an institution, the individuals within that institution must still investigate the context of that pattern before assigning any meaning or consequences to it.

Effective sensing within an educational institution depends on a layered data architecture [16]. First, there must be a layer that establishes lawful and purpose-specific access to the data. Second, there must be a layer that records the data's provenance, quality, missingness, and representativeness. Third, there must be a layer that translates those indicators into educational terms. Fourth, there must be a layer that adds a human interpretation and verification of the data's context. Without these layers to manage the data, universities run the risk of building high-resolution sensing systems that measure convenience rather than the educational reality of the institution. Additionally, data minimization is essential: just because an institution has the capacity to collect more data does not create a reason for that data to be collected.

AI-enabled sensing systems allow organizations to achieve their intended objectives. The university stores its data in various locations which include academic affairs and student services and finance and research management and information technology departments. The proper management of linkages between departments enables organizations to detect hidden common causes which individual departments would not discover on their own. Organizations face an increased risk of purpose drift because they collect data for specific functions which can be used for different purposes. The system permits sensing operations only when it shows clear boundaries and access restrictions and maintains distinct lines between support services and disciplinary actions.

4.2 Cognitive Augmentation of Institutional Decision-making

Another of the ways in which artificial intelligence can enhance higher education is through the augmentation of human cognition. Many tasks that AI systems can perform for humans include summarizing information, simulating scenarios, estimating probabilities, comparing options, and drafting recommendations. Each of these functions can significantly reduce the amount of information that a human must process for a given task and increase the amount of evidence that is made available to those humans. For instance, AI can be implemented to perform quality assurance functions for higher education programs, to handle research administration tasks, and to respond to student inquiries in student services. In each case, the use of AI increases the amount of attention and analysis that can be given to a task by the professionals involved in that task.

Cognitive augmentation by AI involves changes in the discretion that humans have over a task: enrichment and narrowing. Enrichment of discretion means that AI systems increase the amount of evidence and alternative perspectives that humans have at their disposal. Narrowing of discretion means that AI systems present humans with recommendations that frame tasks in particular ways, which causes humans to adopt those recommendations. For instance, if an AI interface presents humans with a single risk score or a ranked list of recommendations, humans may follow those recommendations due to the automation bias of the interface. Such automation bias reflects errors in the design of the interface, rather than in the cognition of the humans using it.

Meaningful human oversight of AI systems requires more than simply placing a human at the end of an automated process. The human that reviews the process must have the ability to disagree with the automated process, have access to all relevant evidence, have time to review the automated process, and be responsible for the final decision in the process. The interface between the automated process and the human reviewer should display the uncertainty, alternative explanations, and limitations of the automated process. For high-consequence decisions, there should not be a reliance on a single automated process or model. Instead, multiple sources of evidence should be used, and the reason for any automated recommendation that materially affects the outcome should be recorded.

Explainability must fit both the user and the decision. For instance, a technical explanation is appropriate for the model developers, but not for the students or academic committees. Operational explanation involves explaining the main factors behind the recommendation, the data used, the confidence involved, and the route for correction. Procedural explanation involves explaining who authorized the system, who owns the decision, and how it can be challenged. Research on explainable AI in education involves research into these aspects of explanation as a social and pedagogical process, rather than a technical feature [17].

4.3 Coordinated Execution through Human-machine Collaboration

The third mechanism is coordinated execution. AI systems connect detection with decision support and organizational unit communication and follow-up operations. The student-service system detects incoming requests before it obtains policy data which it uses to produce responses and schedule meetings and keep track of current cases. The quality-assurance system collects evidence while it performs indicator comparisons to create review summaries which it uses to assign and track corrective actions. The process of coordination reduces waiting times for handoffs while it establishes standard service delivery pathways. AI systems generate the most substantial coordination improvements because they handle repetitive system translation work yet human operators manage all ambiguous and sensitive and value-based cases. Organizations should choose specific tasks for automation according to the fundamental principle of selective automation. The automation process can handle tasks which follow standard procedures and have defined instructions and produce minimal negative outcomes and allow straightforward recovery from mistakes. The management of rights and discipline and assessment and employment and access to vital opportunities needs human intervention at its highest level. The tool which works at one development stage might fail to work at any other stage. A language model can create disciplinary notices but it should not make decisions about the disciplinary actions which appear in those notices.

Human-machine collaboration requires persons to have explicit duties which they must complete. Every AI-supported process should name a business owner, a data steward, a technical owner, a frontline decision-maker, and an independent oversight function. The business owner establishes the educational objectives while taking complete accountability for all resulting results. The data steward manages all aspects of data quality and ensures proper legal compliance for all data operations. The technical owner maintains model performance through their management of security systems. The frontline decision-maker makes sense of system outputs by using their knowledge of operational situations. The oversight function performs audits to check for compliance and it handles all received complaints. The process of separating these roles prevents duties from disappearing into vendor agreements or IT system management operations.

Procurement is part of execution governance, not a separate administrative detail. AI is often acquired by universities through cloud platforms and embedded software features. The contracts for these features should specify the data use, data retention, model changes, subcontractors, security, performance reporting, audit access, and termination support. Vendor assurances cannot replace the accountability of the university. The university is still responsible for the decisions made by the external AI system that it uses to make or shape those decisions.

4.4 Recursive Institutional Learning

The fourth mechanism is recursive institutional learning. Because digital systems produce detailed records of how the institution performs, those records can be analyzed to enable the institution to learn how it performs, where it fails to meet its goals, and how different groups experience the same process. AI can assist in this process by detecting drift, summarizing feedback, and comparing outcomes with intended outcomes. In this way, the institution can move from a focus on creating digital projects to creating routines that can be improved over time.

Recursive learning produces value through rule updates which happen because of feedback. Organizations keep track of operational uptime and total system accuracy yet they tend to disregard how their systems modify performance and produce unequal error distribution and unexpected workload challenges. A complete learning loop should include performance indicators, equity indicators, user experience, complaints, override patterns, and qualitative review. The system needs human intervention because users need to verify the model results since the model generates predictions which they do not trust. The system achieves its goals through low override rates yet it might need human support to operate its automated processes. The process of interpretation needs multiple types of evidence to function properly.

In institutional learning, there must be accountability feedback as well as performance feedback. Performance feedback considers if the system achieved the stated objective. Accountability feedback considers if the objective is still legitimate, the data used, the process, and the effects. An AI system may become more accurate at predicting student withdrawal, but the institution may no longer accept the system. The institution must be able to revise or stop the

system despite its improved performance.

The mechanism chain displayed in Figure 1 shows the ability to interpret the system from either forward or backward perspectives. The development of AI capabilities emerges through the combination of data with models and rules and human knowledge which then transforms institutional operations and results in public-value creation and model and rule and routine updates from outcome feedback. The reverse loop matters because error or value conflict can enter at every stage. The quality of governance becomes dependent on how well the feedback system operates through clear procedures which allow stakeholders to challenge the process and require institutional involvement for execution.

5 RISK BOUNDARIES OF AI-ENABLED DIGITAL GOVERNANCE

5.1 Data Extraction, Privacy, and Purpose Drift

Data extraction represents the initial risk boundary which needs to be addressed.AI systems motivate universities to gather and connect their data because extra data points will enhance their ability to predict and create personalized solutions.Young adults who attend college make up the primary group which higher education institutions base their data collection on because they depend on these institutions.Attendance records together with learning behaviors and financial information and disability-related data and communication records and location data reveal personal details about individuals.Basic data points become capable of generating intrusive profiles when they get connected to one another.

The central governance problem with learning analytics is the issue of purpose drift. Data gathered for one purpose can be used for others. Consent is not enough to protect students' privacy. Legitimate data use requires clear purpose limitation, ethical handling of learning data, and attention to privacy principles in educational analytics [18,19].

Data minimization should be a design requirement for universities. The owners of the systems should be required to indicate why each variable should be added to the system. Even if sensitive attributes are required for fairness testing, they should not be used for prediction, and the two uses should be kept separate. The data used to develop models should be examined for provenance, representativeness, and bias. Though data can be de-identified, it does not eliminate the risk of re-identification or group-level harm.

Data collection crosses a boundary when it becomes disproportionate to the educational purpose for which the data were to be collected in the first place, when data collected for support are used for punitive action without authorization, or when individuals cannot reasonably understand how the data being collected about them will impact the decisions made by the administrators of the school or education facility. At such a point, collecting more data and sensing more about the students and their environments does not lead to better governance of the education facility; rather, it moves

5.2 Algorithmic Bias, Opacity, and Unequal Error

There are also concerns with the bias and opacity of AI systems. AI systems learn from the data generated by the existing institutions. If the processes used to generate that data contain bias, the AI models will learn to reproduce that bias. Additionally, bias may also arise from the way in which the algorithms are designed and deployed in educational settings, where many of the constructs used in education are contested rather than factual [18,20].

Aggregate accuracy can often hide the presence of unequal error. A model that performs well overall may produce more false positives for a subgroup of individuals with different language patterns or digital traces. In student-support settings, a false positive can create unnecessary intervention for the student, while a false negative can withhold necessary support. In admissions, assessment, discipline, or employment decisions, these errors can affect a student’s rights and their chances in life. Therefore, evaluating the fairness of a model requires looking at the model’s performance on subgroups of individuals, the data that they contribute to the model, and the consequences of each type of error.

Opacity in AI can take two forms: technical and institutional. Technical opacity occurs due to the complexity of AI models and the limited interpretability of their decisions. Institutional opacity occurs when users of AI do not know that it is involved in the decision-making process, when procurement contracts restrict access to AI, or when no one can explain how an AI output entered a decision. While transparency does not require disclosing the code of an AI system, it does require providing enough information to allow for oversight and challenge. Research into public decision-making shows that providing explanations of decisions supports the legitimacy of those decisions only when the explanation is relevant to the person affected by the decision and when the decision follows an accountable procedure [21].

An institution reaches its boundary when it fails to detect both data sources and model versions and decision authority and essential factors which lead to significant results. Organizations need to stay away from these systems for decision-making purposes because their accuracy claims become irrelevant during such circumstances. Unknown limits and origins of probability scores make them unsuitable to serve as replacements for institutional decisions which require logical reasoning.

5.3 Automation Dependence and Responsibility Diffusion

The third boundary is automation dependence. With repeated use of the automated system and reliance on its recommendations, staff may begin to lose their skills in the domain, stop seeking out contradictory evidence to the automation’s recommendations, and begin to consider it risky for them to depart from the model’s recommendations.

Such dependence upon automation is especially dangerous in rare or changing conditions.
Often, with responsibility diffusion, each of those involved can point to the other as being responsible for the outcome. Developers supply the tool; administrators are responsible for the model; managers rely on vendor certification; committees assume the responsibility of each individual is diffused through collective review. As a result, there is often an accountability gap in which no actor is able to justify the outcome of the algorithm. Research into algorithmic accountability therefore often stresses an end-to-end audit of the algorithms and the explicit assignment of responsibility for each actor involved in the process [22,23].
Introducing a formal human-in-the-loop rule does not solve the problem if the role of the human is only symbolic. Institutions should monitor the rate at which humans override the AI system, review a sample of the accepted human recommendations, and create a safe environment for humans who question the outputs of the AI system. High-consequence AI systems also require fallback procedures to ensure the continuation of essential services in the event that an AI system is unavailable or suspended.
The boundary is breached when the employees cannot make a defensible decision without the model, when disagreement is punished or made difficult, and when no one accepts responsibility for the adverse outcome. At that point, the AI is displacing the governance capacity of the institution.

5.4 Educational Value Misalignment

The fourth boundary is value misalignment. AI systems are developed to optimize measurable objectives. Many of the goals of education, however, are difficult to measure. For example, intellectual independence, creativity, care for others, dialogue, academic freedom, and the formation of judgment. An AI system optimized for retention may push the teaching staff away from academically demanding experiences. An AI system optimized for rapid feedback may encourage students to focus on superficial learning rather than deep comprehension. An AI system optimized for standardized quality indicators may reduce the diversity of curricula offered by educational institutions.
Generative AI establishes a clear separation because it generates natural language responses which seem finished yet it conceals its lack of knowledge and poor reasoning abilities. In teaching and assessment, the problem is not only misconduct. The process requires educators to establish new definitions for student learning achievements and their ability to produce original work and demonstrate authentic academic progress. The governance system needs to establish rules which control assessment methods, acceptable student help, disclosure, and academic-integrity standards in the era of generative AI [24].
Value misalignment may also result from unequal access. Those students or departments with more resources and knowledge of AI will gain the most from its use. However, if a university adopts AI without considering accessibility for all students, there is a risk that the average efficiency of the university will increase but the equity of its value to all students will decrease. Public value must consider distribution of value to all individuals as well as the total value created by the AI system.
Such a situation may be recognized as a boundary being crossed when the optimization of an objective replaces the purpose of education, when the development of humans is treated as a problem of data management, or when the efficiency of the system depends upon the exclusion of individuals who cannot effectively use the AI system. In such cases, redesigning the objective of the AI system, introducing qualitative judgment into the system, or discontinuing the use of AI in the application may be appropriate responses to these issues.

Table 2 Risk Classification and Minimum Governance Requirements

Risk Level	Typical Uses	Decision Effect	Minimum Controls	Human Role
Low	Drafting; translation; summarization	No direct effect; easily reversible	Security review; user verification; disclosure where relevant	User checks output
Moderate	Service routing; recommendation; internal analytics	Influences attention or service pathway	Named owner; notice; validation; periodic monitoring	Staff reviews exceptions and samples
High	Admission; grading; discipline; employment; financial support; intensive profiling	Material effect on rights, qualification, or opportunity	Impact assessment; subgroup testing; independent audit; logs; appeal	Qualified human decides and records reasons
Prohibited	Manipulative use; unjustified biometric/emotion inference; fully automated consequential judgment	Unacceptable threat to rights or educational values	Do not deploy; remove or suspend	No delegation permitted

6 AN ADAPTIVE GOVERNANCE FRAMEWORK

6.1 Risk-tiered Authorization

The framework begins with risk-tiered authorization. AI uses do not all call for the same controls. Universities need to establish application categories which depend on their impact level and ability to reverse damage and their operational range and their data protection requirements and their automated processing level and their influence on individual rights and available resources. The system enables three basic functions which include grammar support and meeting notes and draft verification before final application. The system handles three levels of risk which include service routing and non-binding recommendations and internal analytics operations. The system operates at high risk because it makes choices about student admission and assessment and discipline and employment and financial support and service availability and complete population profiling. The system should ban any operations which use defenseless people for manipulation while performing unjustified biometric or emotional assessments and making important choices without human intervention [25]. Risk classification should guide the authorization pathway. Tools with minimal risk may go through ordinary information-security and data-protection checks. For moderate-risk uses, the university should have a written purpose, a named owner, basic performance testing, user notice, and regular monitoring. High-risk systems should be subject to algorithmic impact assessment, legal and ethics review, pilot testing, subgroup evaluation, independent audit, executive approval, and a formal appeal channel. Classification should also be revisited when the actual use of a tool changes. A chatbot that only answers general questions may be low risk, but it becomes much more sensitive if it starts to suggest disciplinary action or mental-health intervention.

A central AI register should record each system's purpose, owner, vendor, data sources, model version, risk tier, authorization date, review schedule, and retirement status. Such a register gives the university a clear view of where AI is being used and helps prevent hidden adoption through personal subscriptions or embedded software functions. For systems that may substantially affect students or staff, a public-facing summary should also be made available.

6.2 Lifecycle Impact Assessment and Control

Governance needs to run through the whole lifecycle of an AI system. Before a system is purchased or developed, the university should first clarify the educational problem it is trying to address, whether non-AI options are available, which groups may be affected, and how both success and harm will be judged. At the design stage, attention should be given to data quality, security, fairness, accessibility, and explainability. Before the system is put into use, testing and user training are necessary. Once it is operating, the university should continue to watch for performance problems, model drift, errors, complaints, human overrides, and uses that were not originally expected. When the system is retired, arrangements should also be made for data deletion or archiving, vendor exit, and service continuity.

An algorithmic impact assessment can serve as the main record for this process. It should set out the system's purpose and legal basis, the decisions or services it may affect, the data and models involved, expected benefits, possible harms, affected groups, human oversight arrangements, notice and explanation mechanisms, security measures, audit plans, and routes for redress when negative effects occur. The assessment should be proportionate to the risk of the system, but should be concrete enough to allow for later evaluation of the system.

The system should be monitored using both quantitative and qualitative evidence. Quantitative evidence includes indicators of the system's accuracy, calibration, latency, drift, and security incidents. Qualitative evidence includes indicators of subgroup error, overrides, complaints, accessibility issues, workload, and alignment with the mission of the institution. Qualitative evidence is required because many of the harms that can result from AI systems can first be seen in the behavior of those affected by the system.

Any material changes to the AI system should trigger a reauthorization to use the system. Material changes include a new model provider, major version updates, expansion of the system to new populations, inclusion of sensitive data, change in the use from advisory to automated, and evidence of harm. Material changes in the cloud services for the system make the control more important, not less. The institution should require the cloud providers to notify them of any changes to their systems.

6.3 Meaningful Human Oversight and Traceable Responsibility

The framework sets out five roles to assign responsibility: institutional sponsor, scenario owner, data steward, model or technical owner, and independent oversight body. The institutional sponsor authorizes the use of the model in high-risk scenarios and provides the resources to make this happen. The scenario owner is responsible for the scenario in which the model is used and the outcomes of that scenario. The data steward is responsible for the data used in the model. The technical owner is responsible for the model itself and its configuration, validation, security, and change control. The independent oversight body is responsible for reviewing the use of the model and for handling complaints about its use.

Human oversight should be placed in areas of the system where human judgment is most needed. In moderate-risk systems, staff may be required to review the system's exceptions and outputs in sample scenarios. In high-risk scenarios, a qualified individual should review the model's recommendations before they are enacted in the scenario. This individual should be appropriately trained in the scenario and in the capabilities and limits of the model. The performance of these individuals should not be judged based on the speed of their reviews or their conformity to the model's recommendations.

Traceability of the model's outputs requires logs that show the inputs and outputs of the model, the version of the model used, the actions of the user, and the final decision that was made. These logs must be protected from unauthorized access and retained only for as long as is necessary to support audits, investigations of incidents, and appeals of decisions. Additionally, these logs must not be used to surveil the university's employees without appropriate controls over access to those logs.

Responsibility for the purchasing of models also lies with the university. Contracts with vendors of models should include provisions that ensure the university will have access to the documentation, audit evidence, incident notifications, data-location information, security commitments, and termination of the model from the vendor. For high-risk scenarios, additional evidence from the vendors should be requested to demonstrate their compliance with the university's requirements. A disclaimer in contracts cannot transfer the university's responsibilities to the vendor of the model.

6.4 Contestability, Redress, and the Dual Closed Loop

Contestability in AI means that persons affected by the AI system can question the use of the AI system. The notice must state when AI has contributed to the decision or service, what role the AI had in the decision or service, what information was used by the AI, and how to request a human review of the AI decision or service. The explanation of the AI must be understandable to those affected by the AI system. Simply stating that an algorithm was used in an AI system does not allow for anyone to challenge the AI system.

There must be procedures for redress of the AI system. These procedures should allow for the correction of data that is used by the AI system that is inaccurate, for the AI system decision or service to be reconsidered by a qualified human, for the decision or service to be escalated to another independent body, and for a reasoned response. There should also be time limits for redress of the decision or service by the AI system. The redress procedure for an appeals process should also be analyzed for evidence of systemic problems in the AI system. Repeated successful appeals of decisions by the AI system should trigger a review of the AI model or process.

The dual closed loop system links performance feedback and accountability feedback into the AI system. Performance feedback examines if the AI system achieves the objective that is set for the AI system. Accountability feedback examines if the objective set for and process of the AI system remain legitimate. Both feedback loops feed into decisions regarding the AI system such as authorization, adjustment, revision, training, and termination of the AI system. Figure 4 shows this dual closed loop system for the governance of an AI system.

This dual closed loop system follows the logic of adaptive governance of an AI system. It allows for experimentation with AI systems, but within defined boundaries. It allows for learning from the AI system, but with evidence and feedback. It encourages innovation in AI systems, but also allows for the refusal of an AI system to perform a function. Most importantly, it recognizes that an AI system is only governable if the institution can identify it, understand it, intervene in it, explain it, and stop it.

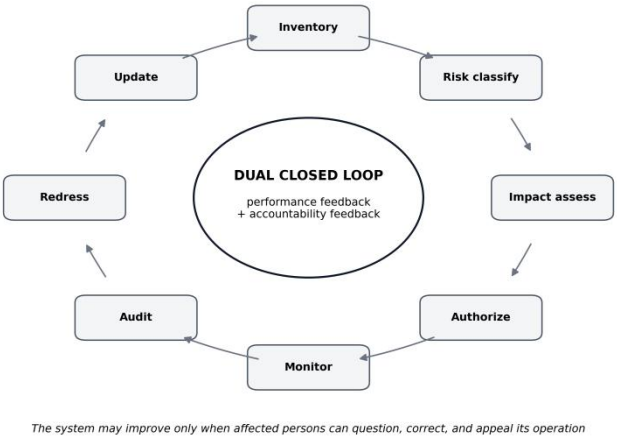

Figure 4 Adaptive Governance Cycle and Dual Closed Loop

7 SCENARIO-BASED IMPLEMENTATION IN HIGHER EDUCATION

7.1 Student Services

Student services are good places to begin adopting AI in a controlled manner. Chatbots can be implemented to answer routine questions from students, translate information into different languages, and route students to the appropriate services [26]. Recommendation systems can be used to help students find courses that meet their needs, scholarships they may be eligible for, and other support services. These uses of AI can create value for student services by reducing the barriers students face in accessing services and freeing up staff to handle the more complex student cases. These AI services should always be clearly identified as being assisted by AI, allow students to easily escalate to a person, and avoid presenting students with information that may be uncertain but is presented as authoritative.

AI systems that infer information about students' vulnerabilities, mental health, misconduct, or likelihood to withdraw pose a higher risk to those students. While such predictions may allow student services to provide assistance to those students earlier than they otherwise would, those predictions may also lead to stigmatization of those students and the provision of potentially intrusive interventions. Student analytics of this type should be performed using minimal data about those students, by creating a separation between providing assistance and punishment, by testing for unequal errors in the analytics, and by requiring a review of the context of each analytics.

7.2 Teaching, Learning, and Assessment

In the teaching context, AI can assist with lesson preparation, providing feedback for students, making lessons accessible for all students, providing language assistance, and providing personalized practice for students. The governance for AI in education should distinguish between using AI for learning support and using AI to replace the intellectual work of the students. At the course level, rules should be established regarding the permitted uses of AI and ensuring that these uses align with the learning outcomes for the course [27]. Students should be provided with opportunities to develop AI literacy, learn how to verify AI-generated content, disclose.

Assessment functions as a critical evaluation because it determines if students have reached their learning objectives. AI systems can help instructors with rubric application and feedback creation and moderation tasks but teachers must keep their authority to assign grades which determine student advancement and final certification. Institutions need to transform their assessment methods by focusing on student work processes and presentations and real-world tasks and self-assessment instead of depending on detection tools. The established error patterns of Generative-AI detectors remain unknown because these tools fail to establish reliable misconduct indicators.

Academic-integrity governance needs to extend its operations beyond its current practice of prohibition. The document needs to establish rules which specify how to identify authors and determine acceptable levels of help and handle disclosures and evidence presentation while maintaining fair treatment for all parties involved. Staff members need instructions to operate AI systems which help them create educational resources and assess student learning outcomes. The institution must follow the same rules for transparency and verification which students need to follow.

7.3 Quality Assurance and Institutional Planning

Quality assurance is a promising domain for AI integration. AI can analyze program data and quality assurance evidence to identify anomalies and perform comparative analyses. Recent work proposes models in which AI can support quality assurance processes while the governance structures of institutions maintain human accountability for quality assurance decisions [28]. The main risk of using AI in quality assurance is metric fixation, where focusing on easily measured quality assurance indicators ignores more qualitative aspects of education quality.

To mitigate the risk of metric fixation, quality assurance processes that use AI should rely on mixed evidence and academic interpretations of quality assurance data. While AI models can analyze quality assurance data and identify patterns, quality assurance review panels should examine the data themselves, and programs should be allowed to provide explanations for any differences in their quality assurance data. Any quality assurance planning tools that use AI should disclose the assumptions made by the AI, as well as any uncertainties in these assumptions. This is especially important for quality assurance planning related to enrollment, finance, and staffing of educational institutions. Finally, any outputs from quality assurance scenarios should be treated as conditional evidence for quality assurance planning rather than as predictions of the future of educational institutions.

7.4 Research Administration and Knowledge Governance

Artificial intelligence can assist with research administration tasks such as matching funding calls, checking administrative completeness, summarizing research portfolios, and supporting compliance reviews. AI can also assist researchers with tasks such as searching for literature, coding data, or creating draft documents. Research governance should protect confidential research proposals, intellectual property, unpublished research data, and any research data that is subject to export control or other sensitivity considerations. Public consumer tools for artificial intelligence should not be provided with any restricted research data unless there are institutional agreements in place to provide appropriate safeguards for that data.

Decisions about funding, hiring, promotion, and research integrity are high risk. AI can help organize evidence for these decisions, but it should not autonomously make rankings or inferences about research misconduct. Indicators used in bibliometrics and text analyses can include biases related to research fields, languages, and the career stages of researchers. Committees should retain the responsibility for the standards and reasons for their decisions. Researchers should disclose any material assistance from AI in a manner that complies with the requirements of publishers and research disciplines.

7.5 Campus Operations and Risk Management

Artificial intelligence can contribute to the improvement of campus operations in a variety of areas. Many of these applications have the potential to create significant benefits for the campus community. Some of the systems that pose a high risk to the rights of individuals on the campus should be used only in exceptional circumstances and in accordance

with the law and campus policies.

Systems related to cybersecurity are one of the areas in which AI can be used to improve campus operations. However, there are a variety of considerations that must be made to effectively manage the risks associated with these systems. For instance, an increase in the number of threats to the campus network will require an update to the detection systems. However, automatic responses to cyber threats can have a negative impact on the campus network. Thus, human intervention is required to manage these systems and ensure the safety of the network. In addition to the safety of the campus network, the rights of the campus community should also be considered in the management of these cybersecurity systems and procedures.

Table 3 provides a translation of the framework into specific control measures for each of the scenarios presented in this chapter. The pattern that emerges is that the more that an AI system impacts the rights, qualifications, discipline, employment, or personal information of individuals on the campus, the more control measures are required to manage those systems appropriately.

Table 3 Scenario-Control Matrix for Higher-Education AI Applications

Scenario	Permissible AI Role	Risk Tier	Core Controls	Required Redress
Student information service	Answer, translate, route, schedule	Low-Moderate	Verified knowledge base; identity disclosure; human escalation	Correction and staff review
Student-risk analytics	Flag patterns for supportive review	High	Data minimization; fairness tests; separation from punishment; case review	Human reconsideration and data correction
Teaching and formative feedback	Generate examples and non-binding feedback	Moderate	Course rules; literacy training; teacher verification; accessibility	Alternative support and complaint channel
Grading or academic integrity	Organize evidence; assist moderation	High	No sole-source judgment; documented academic decision; audit trail	Appeal to qualified academic body
Quality assurance	Summarize evidence; identify anomalies	Moderate-High	Mixed evidence; contextual interpretation; model/version documentation	Program response and independent review
Research administration	Match calls; screen completeness; summarize portfolios	Moderate	Confidentiality; IP protection; bias review; human committee judgment	Correction and committee reconsideration
Campus security and biometric systems	Limited support under exceptional necessity	High-Prohibited	Legal basis; necessity; proportionality; independent authorization	Rapid challenge, suspension, and external oversight

8 RESEARCH CONCLUSIONS AND PROSPECTS

8.1 Major Conclusions

Artificial intelligence-enabled digital governance in higher education is conceptualized as a socio-technical institutional arrangement in which data, models, platforms, rules, and humans interact to produce the decisions and services provided within the institution. There are four main mechanisms by which such an arrangement can operate: sensing, augmentation, execution, and learning. These mechanisms provide benefits to the institution, such as improved quality, equity, efficiency, and responsiveness. However, they also introduce changes to the visibility, discretion, and responsibility of various actors within the institution.

Several risks are associated with each of the mechanisms mentioned above. For example, the sensing mechanism can lead to data extraction from students and drift from the intended purpose of the AI-enabled digital governance system. Similarly, augmentation can lead to bias in the AI system and a lack of transparency in the decisions made. Execution can lead to an over-dependence on automation within the institution and a diffusion of responsibility for the decisions made. Finally, the learning mechanism can lead to a drift from the values of the institution as the AI systems adapt to perform better in their tasks.

To mitigate the risks associated with AI-enabled digital governance in higher education, a framework can be created to govern such systems. The framework includes aspects such as authorization based on the level of risk posed by each system, impact assessments for each system throughout its lifecycle, human oversight, traceability of responsibilities, audits, contestability of decisions made by the AI systems, and mechanisms for redress of any negative impacts caused by the AI systems. At the center of the framework is a dual closed loop. One loop asks whether the AI system is performing as intended; the other asks whether the decisions shaped by the system remain accountable and legitimate.

In general, AI-enabled digital governance should be used to support the judgment of people within the university, not to replace it. These systems should not turn into independent authorities, invisible administrative layers, or substitutes for

the educational purposes of the institution. Even when decisions are influenced by vendors, platforms, or embedded models, the university still bears responsibility for actions taken in its name. Legitimate adoption therefore depends on whether the institution can identify where AI is being used, explain its role, allow its use to be questioned, and stop it when necessary.

8.2 Theoretical and Practical Implications

From the theoretical perspective, the framework links digital-governance research with higher-education studies. The institutional significance of AI lies not only in automation but also in reconfiguring knowledge and discretion within institutions of higher education. The framework extends research into ethical principles of AI by translating concepts like fairness, transparency, privacy, and human agency into the framework of governance. The risk-boundary concept proposed in the framework explains the need for appropriate governance for AI in higher education that is proportionate to the consequences of the use of AI, rather than its label as an AI technology.

In practice, universities should establish an AI governance committee or equivalent body. They should maintain an AI register, adopt a risk taxonomy, require impact assessment for high-risk uses of AI, and define minimum contract terms for AI vendors. Universities should also train their staff and students in AI literacy. They should provide alternative channels for individuals who are unable or unwilling to use AI-mediated services, and evaluate equity alongside efficiency in their use of AI. The governance of AI in higher education should be integrated with other governance policies like data protection, cybersecurity, academic quality, ethics, procurement, and internal audit, rather than in a single technical unit.

Institutional policy also needs enough specificity to guide action. While broad declarations of responsible AI use within the institution are necessary, they are not sufficient. Policies should specify prohibited uses of AI, authorization thresholds, disclosure duties, standards for human review of outcomes, evidence requirements, retention periods for data, procedures for handling incidents involving AI, and rights of appeal [29]. Policies should also be reviewed on a regular basis as AI technologies and regulations develop.

8.3 Limitations and Future Research

This is a conceptual study and did not involve the observation of any one institution. The categories proposed in this study require empirical testing within a variety of higher-education institutions. Future research should focus on how universities assign responsibility for AI systems, how frontline staff interact with AI recommendations, and whether the impact assessments and appeal mechanisms in place actually change the outcomes of AI systems within universities.

Comparative research is also needed to determine the maturity of the policies that different institutions have in place regarding AI. Analyses of existing documents from several institutions have shown that there is considerable variation in the guidance that each institution provides regarding AI [11]. Future research could create measurable indicators of the maturity of these policies. Research could also examine whether the governance mechanisms in place remain effective as AI models and vendors change.

Another area of research that should be prioritized is the measurement of the value that AI systems provide to the public. While the performance of AI systems can be measured, it is much more difficult to measure values such as trust in AI systems, academic freedom, the development of intellectual capabilities, and distributive fairness. Research that uses both qualitative and quantitative methods could provide valuable information on these issues. Additionally, students and staff should be included in the design of AI systems that have a significant impact on these individuals' lives.

Finally, as AI systems become more generative and agentic, they will be able to perform multiple tasks within universities. This change in the capabilities of AI systems will require changes in the governance of these systems. Instead of having governance mechanisms that are specific to each tool, universities will need to have governance mechanisms in place to continuously control the capabilities of these AI systems. As a general principle, the more autonomous that an AI system is, the more responsibility that the institution must have for that system.

COMPETING INTERESTS

The author declares that there are no relevant financial or non-financial interests.

FUNDING

This work was supported by the National Social Science Foundation of China (Grant No. 24VSZ170).

REFERENCES

[1] Bullock J B. Artificial Intelligence, Discretion, and Bureaucracy. The American Review of Public Administration, 2019, 49(7): 751-761. DOI: 10.1177/0275074019856123.

[2] Mergel I, Edelmann N, Haug N. Defining digital transformation: Results from expert interviews. Government Information Quarterly, 2019, 36: 101385. DOI: 10.1016/j.giq.2019.06.002.

[3] Sun T Q, Medaglia R. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. Government Information Quarterly, 2019, 36: 368-383. doi:10.1016/j.giq.2018.09.008.

[4] Wirtz B W, Weyerer J C, Geyer C. Artificial Intelligence and the Public Sector—Applications and Challenges. International Journal of Public Administration, 2019, 42: 596-615. DOI: 10.1080/01900692.2018.1498103.
[5] Zuiderwijk A, Chen Y C, Salem F. Implications of the use of artificial intelligence in public governance: A systematic literature review and a research agenda. Government Information Quarterly, 2021, 38: 101577. DOI: 10.1016/j.giq.2021.101577.
[6] Ouyang F, Zheng L, Jiao P. Artificial intelligence in online higher education: A systematic review of empirical research from 2011 to 2020. Education and Information Technologies, 2022, 27: 7893-7925. DOI: 10.1007/s10639-022-10925-9.
[7] Zawacki-Richter O, Marín V I, Bond M, et al. Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 2019, 16: 39. DOI: 10.1186/s41239-019-0171-0.
[8] Kasneci E, Sessler K, Küchemann S, et al. ChatGPT for good? On opportunities and challenges of large language models for education. Learning and Individual Differences, 2023, 103: 102274. DOI: 10.1016/j.lindif.2023.102274.
[9] Holmes W, Porayska-Pomsta K, Holstein K, et al. Ethics of AI in Education: Towards a Community-Wide Framework. International Journal of Artificial Intelligence in Education, 2022, 32: 504-526. DOI: 10.1007/s40593-021-00239-1.
[10] Williamson B, Eynon R. Historical threads, missing links, and future directions in AI in education. Learning, Media and Technology, 2020, 45: 1-13. DOI: 10.1080/17439884.2020.1798995.
[11] Humble N. Higher Education AI Policies—A Document Analysis of University Guidelines. European Journal of Education, 2025, 60: e70214. DOI: 10.1111/ejed.70214.
[12] Floridi L, Cowls J, Beltrametti M, et al. AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 2018, 28: 689-707. DOI: 10.1007/s11023-018-9482-5.
[13] Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nature Machine Intelligence, 2019, 1: 389-399. DOI: 10.1038/s42256-019-0088-2.
[14] Rahwan I. Society-in-the-loop: programming the algorithmic social contract. Ethics and Information Technology, 2018, 20: 5-14. DOI: 10.1007/s10676-017-9430-8.
[15] Janssen M, Van Der Voort H. Agile and adaptive governance in crisis response: Lessons from the COVID-19 pandemic. International Journal of Information Management, 2020, 55: 102180. DOI: 10.1016/j.ijinfomgt.2020.102180.
[16] Gasser U, Almeida V A F. A Layered Model for AI Governance. IEEE Internet Computing, 2017, 21: 58-62. DOI: 10.1109/MIC.2017.4180835.
[17] Khosravi H, Shum S B, Chen G, et al. Explainable Artificial Intelligence in education. Computers and Education: Artificial Intelligence, 2022, 3: 100074. DOI: 10.1016/j.caeai.2022.100074.
[18] Baker R S, Hawn A. Algorithmic Bias in Education. International Journal of Artificial Intelligence in Education, 2022, 32: 1052-1092. DOI: 10.1007/s40593-021-00285-9.
[19] Pardo A, Siemens G. Ethical and privacy principles for learning analytics. British Journal of Educational Technology, 2014, 45: 438-450. DOI: 10.1111/bjet.12152.
[20] Selbst A D, Boyd D, Friedler S A, et al. Fairness and Abstraction in Sociotechnical Systems. Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA: Association for Computing Machinery, 2019: 59-68. DOI: 10.1145/3287560.3287598.
[21] De Fine Licht K, De Fine Licht J. Artificial intelligence, transparency, and public decision-making. AI & SOCIETY, 2020, 35: 917-926. DOI: 10.1007/s00146-020-00960-w.
[22] Kroll J A, Huey J, Barocas S, et al. Accountable Algorithms. Social Science Electronic Publishing, University of Pennsylvania Law Review, 2016, 165: 633-705.
[23] Raji I D, Smart A, White R N, et al. Closing the AI accountability gap: defining an end-to-end framework for internal algorithmic auditing. Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain: Association for Computing Machinery, 2020: 33-44. DOI: 10.1145/3351095.3372873.
[24] Coates H, Croucher G, Calderon A. Governing Academic Integrity: Ensuring the Authenticity of Higher Thinking in the Era of Generative Artificial Intelligence. Journal of Academic Ethics, 2025, 23: 2015-2028. DOI: 10.1007/s10805-025-09639-7.
[25] Yeung K. Algorithmic regulation: A critical interrogation. Regulation & Governance, 2018, 12: 505-523. DOI: 10.1111/rego.12158.
[26] Tlili A, Shehata B, Adarkwah M A, et al. What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. Smart Learning Environments, 2023, 10: 15. DOI: 10.1186/s40561-023-00237-x.
[27] Chan C K Y. A comprehensive AI policy education framework for university teaching and learning. International Journal of Educational Technology in Higher Education, 2023, 20: 38. DOI: 10.1186/s41239-023-00408-3.
[28] Isaifan R J, Hasna M O. Artificial intelligence for quality assurance in higher education: a policy-to-practice model from Qatar with global relevance. Quality in Higher Education, 2025, 31: 288-303. DOI: 10.1080/13538322.2025.2576326.
[29] Alfiras M I I, Emran A Q, Mohamed A M. Ethics and governance of generative AI in education: a systematic review on responsible adoption. Discover Education, 2025, 5: 37. DOI: 10.1007/s44217-025-01051-y.