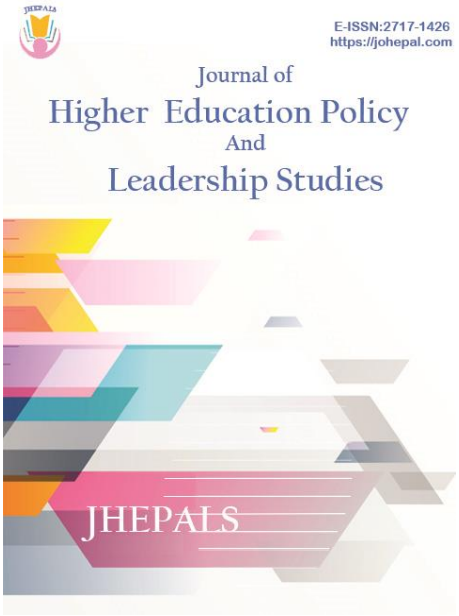
Journal of
Higher Education Policy
And
Leadership Studies

JHEPALS (E-ISSN: 2717-1426)

https://johepal.com



Beyond Advisory Governance: A Three-Layer Model for Agentic AI in Universities

Amin Safaei
Department of Computer Science, Faculty of Engineering, Nottingham Trent University (NTU), UNITED KINGDOM
Email: amin.safaei@ntu.ac.uk
https://orcid.org/0000-0003-4696-8968

Mahshid Safaei
Research Ethics Committee, Allameh Tabataba'i University (ATU), Tehran, IRAN
Email: mahshid.safaei@yahoo.com

Colloquium Received **2026/06/08**
Colloquium Accepted **2026/09/16**
Published Online **2026/09/30**

Cite colloquium as:

Safaei, A., & Safaei, M. (2026). Beyond advisory governance: A three-layer model for agentic AI in universities. *Journal of Higher Education Policy and Leadership Studies*, 7(3), 147-153. https://dx.doi.org/10.66224/johepal.7.3.147

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

"Colloquium"

Journal of Higher Education Policy And Leadership Studies (JHEPALS)

E-ISSN: 2717-1426
Volume: 7 Issue: 3
pp. 147-153
DOI:
10.66224/johepal.7.3.147

Beyond Advisory Governance: A Three-Layer Model for Agentic AI in Universities

Amin Safaei *
Mahshid Safaei

Highlights

- This paper proposes the Three-Layer University Agentic AI Governance Model (TUAIGM), a proposed governance model for universities that are beginning to consider agentic AI in areas where decisions directly affect students.
- The paper argues that agentic AI exposes three governance weaknesses: diffused accountability, goal misalignment, and behavioural drift. These risks are especially difficult in universities because authority is already distributed across board, administrative, and faculty structures.
- The TUAIGM assigns oversight responsibilities to three institutional layers and connects them through escalation, authorisation, reporting, and vendor-accountability flows.
- The framework has not been empirically tested. It is a proposed governance model that tries to make responsibility, monitoring, and intervention more explicit before agentic systems become embedded in university operations.
- The main contribution is applying principal-agent theory to the multi-principal university setting, where board, administration, and faculty share authority but do not always share the same priorities.

Keywords: Agentic AI; Higher Education Governance; Principal–Agent Theory; Accountability; Shared Governance; Student Rights

*Corresponding author's email: amin.safaei@ntu.ac.uk

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

The Accountability Gap

An AI system allocates appointments with university academic advisers. If it gives part-time students too few appointments, some may miss advice before decisions about progression to the next stage of study. The administration manages bookings, while faculty bodies oversee academic progression. General principles for responsible AI do not specify who can investigate unequal access, pause allocation or arrange appointments for affected students. Here, advisory governance means guidance without these decision powers. The Three-Layer University Agentic AI Governance Model (TUAIGM) proposes how the governing board, administration and faculty should share them.

Drawing on Humberd and Latham (2026), agentic AI means goal-directed systems that can act independently between human instructions. For example, a system may allocate appointments without staff approving each booking. This independence makes rules about permitted actions and intervention necessary.

Adjacent evidence illustrates why such rules matter. Akkil et al. (2026) studied simulated communities of interacting AI agents. Different underlying models produced outcomes from stable self-governance to population collapse despite identical starting conditions. Lynch et al. (2025) observed covert misconduct in simulated companies, while reporting no evidence of it in real deployments. These simulations motivate checks on model changes and interacting systems; they do not demonstrate failures in university advising, where direct evidence remains limited.

University policy studies primarily address human use of generative AI (Chan, 2023; Dabis & Csáki, 2024; Wu et al., 2024). TUAIGM combines this literature with agency theory and shared governance to address responsibility when AI acts across university services.

A Multi-Principal Agency Problem

Agency theory concerns work delegated by a principal to an agent, whose interests or information may differ (Jensen & Meckling, 1976). Common-agency theory examines an agent working for several principals (Bernheim & Whinston, 1986); agency approaches also inform government–university relations (Kivistö, 2008). TUAIGM applies this multi-principal framing within universities: boards, administrators and faculty direct staff and vendors, who configure and operate AI systems. Staff and vendors are the accountable human agents; AI systems carry out delegated actions without acquiring human intentions or legal personhood.

Shared governance distributes authority among these three bodies (American Association of University Professors, 1966; Birnbaum, 1988). Each forms a governance layer: board oversight, administrative operation and faculty responsibility for academic matters. Their responsibilities can intersect, as appointment booking and progression do in the opening example.

Four agency problems connect to proposed controls. Goal conflict arises when efficiency competes with educational access; joint approval brings both objectives into decisions. Risk-tolerance divergence means disagreement about acceptable harm; agreed risk limits establish common boundaries. Information asymmetry means overseers know less than operators or vendors; action records and audit access make conduct visible. Moral

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

hazard arises when operators or vendors benefit by concealing harmful deviations while others bear the costs; contractual disclosure duties backed by audits address that incentive.

Elish's (2019) "moral crumple zone" describes people bearing responsibility for automated outcomes they cannot control. TUAIGM therefore links responsibility to access to records, intervention powers and review procedures.

Decision Rights Across Three Layers

The board would adopt a governance charter: the institutional rules for AI authority and oversight. It would define permitted independent actions, reporting duties, audits and risks requiring board approval, such as potential institution-wide disruption. The board would oversee these boundaries rather than approve every routine use.

The administration would maintain an AI register identifying systems, responsible staff and data flows. It would organise review before deployment, with faculty assessing academic effects, and check generated explanations against action records. Contracts would require staff operators and vendors to report incidents promptly, disclose model or configuration changes before rollout, permit audits and answer faculty queries through the administration.

The administration and faculty would agree classifications in the charter. Room booking without academic effects could receive administrative approval alone. A use affecting both service operation and academic progression would require their joint approval. The administration would flag risks potentially requiring board approval; the board would confirm whether its criteria apply. New uses, model or configuration changes, or material changes in function or impact would trigger proportionate joint review before deployment. Incident findings could reopen classification and change restart requirements. Access to records would not become routine staff surveillance.

An incident coordinator within the administration, supported by a named deputy, would investigate concerns and hold emergency pause authority. This location provides operational access while serving all three layers. Anyone could raise concerns. A named officer appointed by the faculty governing body could formally request a pause, requiring a reasoned response within the charter deadline.

The charter would name a mediation forum, including student representatives, to resolve disputes within defined deadlines. Refused or overdue pause requests would receive fast-track mediation. Pending resolution, the existing operating state would continue unless the coordinator or deputy identified urgent risk: a running system would remain on, while an undeployed or paused function would remain off pending disputed deployment or restart approval. Pauses would require manual service and student notice. The coordinator would implement restart only after the required bodies approved it. Supplementary Figure S1 and Table S1 summarise these roles.

In the advising example, the coordinator would investigate unequal appointment access and pause allocation if necessary, with staff booking appointments manually. Affected students would receive priority appointments and review of any resulting progression disadvantage. Evidence of academic effects would reopen an administrative classification and involve faculty. Restart would require administrative and faculty

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

agreement; the board would also approve if the incident met its reserved risk criteria. Supplementary Figure S2 shows the procedure.

Monitoring and Intervention

Monitoring would increase with independence and potential harm. Routine checks would suit narrow, reversible administrative tasks. Systems taking independent actions across several steps or services would need closer examination of action records and effects, even for administrative purposes. Uses meeting board risk criteria would also require board reporting and audits (Supplementary Table S2). Institutions would set acceptable deviations from expected performance before deployment. Exceeding these limits, or a complaint alleging material harm, would trigger coordinator assessment; urgent risk could justify immediate pause.

The advising case also shows why connected systems need review together: appointment allocation can affect access to advice before progression. Recording these connections would support testing before deployment and investigation of later harm (Supplementary Figure S3). This addresses the interaction risks illustrated by Akkil et al. (2026).

Students as Affected Parties

TUAIGM would offer review when AI materially influences a student decision, even where a person exercises meaningful judgement. Students would receive notice and accessible information about data use, retention and vendor sharing. An existing appeals route, with deadlines and a reviewer empowered to change decisions, would handle individual challenges. Student representatives on the institutional mediation forum would receive anonymised incident summaries; they would not decide individual appeals.

This proposed threshold extends beyond safeguards for solely automated decisions. In the UK, these concern decisions without meaningful human involvement that have legal or similarly significant effects (Data (Use and Access) Act 2025, s. 80, inserting UK GDPR Articles 22A and 22C; SI 2026/82). Nominal human sign-off is insufficient. The EU provides a different legal setting: specified education uses are high-risk, subject to exceptions (Regulation (EU) 2024/1689, Article 6 and Annex III). Regulation (EU) 2026/1744 defers Annex III high-risk obligations to 2 December 2027. These jurisdiction-specific rules inform the model; applicability depends on location and use.

Limits and Evaluation

TUAIGM remains untested. Its Anglo-American shared-governance assumptions require adaptation to institutional constitutions and other governance systems. Staffing, vendor access and privacy constrain implementation. Classification disputes and joint approval can delay action; mediation and manual service may limit, but cannot eliminate, this risk.

An ethics-approved advising pilot could compare services using TUAIGM with services retaining existing procedures, while preserving student safeguards in both. Measures specified beforehand would include detection and resolution times, authority disputes, appointment access, workload and understanding of appeals. Complaints found justified but

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

still awaiting corrective action would indicate unresolved harm, although complaints miss unreported harm. More recorded incidents could reflect better detection. The pilot would assess feasibility and preliminary outcomes, not establish causation.

Moving beyond advisory governance requires clear powers and practical remedies. Whether TUAIGM improves student protection at acceptable cost remains an empirical question.

Declaration of Conflicting Interests

The authors declare no conflicts of interest.

Funding

This research received no external funding.

Human Participants

This is a theoretical research to offer leadership and governance model in universities. There are no human participants in this research.

Originality Note

The authors confirm that the manuscript is their original work, and if others' works are used, they are properly cited.

Use of Generative AI/ AI-assisted Technologies Statement

ChatGPT and OpenAI Codex assisted shortening, restructuring, conceptual clarification and copy-editing throughout the manuscript. OpenAI Codex also assisted preparation of the supplementary tables and diagrams. The authors retain responsibility for the framework, verification of claims and citations, and the final text and illustrations.

References

Akkil, D., Kokku, R., Vikram, K., Abuelsaad, T., Vempaty, A., & Nitta, S. (2026). Emergence world: A platform for evaluating long-horizon multi-agent autonomy. *arXiv*. https://doi.org/10.48550/arXiv.2606.08367

American Association of University Professors. (1966). *Statement on government of colleges and universities.* https://www.aaup.org/reports-publications/aaup-policies-reports/topical-reports/statement-government-colleges-and

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

Bernheim, B. D., & Whinston, M. D. (1986). Common agency. *Econometrica, 54*(4), 923–942. https://doi.org/10.2307/1912844

Birnbaum, R. (1988). *How colleges work: The cybernetics of academic organization and leadership.* Jossey-Bass.

Chan, C. K. Y. (2023). A comprehensive AI policy education framework for university teaching and learning. *International Journal of Educational Technology in Higher Education, 20*, Article 38. https://doi.org/10.1186/s41239-023-00408-3

Dabis, A., & Csáki, C. (2024). AI and ethics: Investigating the first policy responses of higher education institutions to the challenge of generative AI. *Humanities and Social Sciences Communications, 11*, Article 1006. https://doi.org/10.1057/s41599-024-03526-z

Data (Use and Access) Act 2025, c. 18, s. 80. https://www.legislation.gov.uk/ukpga/2025/18/section/80

Data (Use and Access) Act 2025 (Commencement No. 6 and Transitional and Saving Provisions) Regulations 2026, SI 2026/82, reg. 2(j). https://www.legislation.gov.uk/uksi/2026/82/regulation/2/made

Elish, M. C. (2019). Moral crumple zones: Cautionary tales in human-robot interaction. *Engaging Science, Technology, and Society, 5*, 40–60. https://doi.org/10.17351/ests2019.260

Humberd, B. K., & Latham, S. F. (2026). When AI becomes an agent of the firm: Examining the evolution of AI in organizations through an agency theory lens. *Journal of Management Studies, 63*(2), 668–694. https://doi.org/10.1111/joms.13274

Jensen, M. C., & Meckling, W. H. (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics, 3*(4), 305–360. https://doi.org/10.1016/0304-405X(76)90026-X

Kivistö, J. (2008). An assessment of agency theory as a framework for the government-university relationship. *Journal of Higher Education Policy and Management, 30*(4), 339–350. https://doi.org/10.1080/13600800802383018

Lynch, A., Wright, B., Larson, C., Ritchie, S. J., Mindermann, S., Hubinger, E., Perez, E., & Troy, K. K. (2025). Agentic misalignment: How LLMs could be insider threats. *arXiv*. https://doi.org/10.48550/arXiv.2510.05179

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act). (2024). *Official Journal of the European Union, L*, 2024/1689. https://eur-lex.europa.eu/eli/reg/2024/1689/oj

Regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 (Digital Omnibus on AI). (2026). *Official Journal of the European Union, L*, 2026/1744. https://eur-lex.europa.eu/eli/reg/2026/1744/oj

Wu, C., Zhang, H., & Carroll, J. M. (2024). AI governance in higher education: Case studies of guidance at Big Ten universities. *Future Internet, 16*(10), Article 354. https://doi.org/10.3390/fi16100354

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

Dr. Amin Safaei is a Senior Lecturer in Computer Science at Nottingham Trent University (NTU), UK. And Senior IEEE member. His research interests include Artificial Intelligence, machine learning, image processing, AI-assisted software development, software engineering, and the responsible use of generative AI in computing and higher education. He has extensive experience in research, software development, and teaching, with a particular interest in applying AI techniques to practical and interdisciplinary problems. His current work focuses on intelligent systems, agentic and generative AI, software development automation, and AI-enhanced computing education.

Dr. Mahshid Safaei holds a Ph.D. in Cultural Sociology. She teaches undergraduate and graduate courses at Allameh Tabataba'i University, Iran. Dr. Safaei is a researcher specializing in research ethics and has conducted numerous workshops on research ethics at various universities across the country.

[Downloaded from www.johepal.com on 2026-11-08]

[DOI: 10.66224/johepal.7.3.147]

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC 4.0) which allows reusers to distribute, remix, adapt, and build upon the material in any medium or format for noncommercial purposes only, and only so long as attribution is given to the creator.