

**INTERNATIONAL JOURNAL OF  
INNOVATION AND  
INDUSTRIAL REVOLUTION  
(IJIREV)**[www.gaexcellence.com/ijirev](http://www.gaexcellence.com/ijirev)**ANSWERABLE FOR WHAT ONE CANNOT KNOW:  
ACCOUNTABILITY WITHOUT COMPREHENSION IN THE  
ALGORITHMIC UNIVERSITY**

Fahmi Abdul Rahim<sup>1\*</sup>, Noor Shuhailie Mohamed Noor<sup>2</sup>, Wan Hasmat Wan Hasan<sup>3</sup>, Fadhlina Izzah Saman<sup>4</sup>, Afzan Nor Talib<sup>5</sup>

<sup>1</sup>Faculty of Business Management, Universiti Teknologi MARA, Cawangan Melaka, 75300 Melaka, Malaysia

[fahmi029@uitm.edu.my](mailto:fahmi029@uitm.edu.my)<https://orcid.org/0000-0001-8715-793X>

<sup>2</sup>Politeknik Melaka, 75250 Melaka, Malaysia

[shuhailie@polimelaka.edu.my](mailto:shuhailie@polimelaka.edu.my)<https://orcid.org/0009-0005-1790-1088>

<sup>3</sup>Faculty of Business Management, Universiti Teknologi MARA, Cawangan Melaka, Kampus Bandaraya Melaka, 75300 Melaka, Malaysia

[hasmat@uitm.edu.my](mailto:hasmat@uitm.edu.my)<https://orcid.org/0000-0002-3828-5967>

<sup>4</sup>Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA, Cawangan Melaka, Kampus Jasin, 77300 Merlimau, Melaka, Malaysia

[fadhlina218@uitm.edu.my](mailto:fadhlina218@uitm.edu.my)<https://orcid.org/0009-0008-9445-2599>

<sup>5</sup>Faculty of Business Management, Universiti Teknologi MARA, Cawangan Melaka, Kampus Bandaraya Melaka, 75300 Melaka, Malaysia

[afzan150@uitm.edu.my](mailto:afzan150@uitm.edu.my)<https://orcid.org/0009-0007-4127-1578>

\*Corresponding Author

**Article Info:****Article history:**

Received date: 26.07.2026

Revised date: 24.08.2026

Accepted date: 11.09.2026

Published date: 30.09.2026

**Abstract:**

Universities increasingly rely on intelligent systems to screen admissions, detect academic misconduct, flag students at risk and support grading. The leader who signs such a decision remains answerable for it, yet cannot reconstruct the reasoning that produced it. Existing scholarship does not close this gap: work on algorithmic opacity explains why the reasoning is inaccessible, work on artificial intelligence in higher education asks that accountability be traced but not what it consists of, and accountability theory assumes that whoever answers can give reasons. The objective of this paper is to develop a conceptual framework of accountability under algorithmic opacity. It names the condition accountability without comprehension and specifies a form of leadership fitted to it, answerability under opacity, through five dimensions: retained answerability, refusing to displace responsibility onto the system or its supplier; procured comprehension, obtaining through testing, audit and contract the understanding a leader

**To cite this document:**

Rahim, F. A., Noor, N. S. M., Hasan, W. H. W., Saman, F. I., & Talib, A. N. (2026). Answerable For What One Cannot Know: Accountability Without Comprehension in The Algorithmic University. *International Journal of Innovation and Industrial Revolution*, 8(26), 909-923.

cannot hold personally; scope discipline, holding a system to the boundary within which it was validated; explanation owed to those affected, giving the student a usable account rather than a technical one; and reversion authority, the capacity to return a decision to human hands. Theoretically it extends accountability theory, which treats reason-giving as constitutive of the relationship, by showing the relationship survives when that element is removed. Accountability does not disappear under opacity but shifts, from justifying the decision to justifying the arrangements that produced it, so legitimacy rests on those arrangements. It separates substantive answerability from oversight that legitimates without constraining, and advances six propositions, boundary conditions and a research agenda. For innovation leadership it argues that the capacity to account for intelligent systems is part of the innovation: institutions adopting Industry 4.0 and Industry 5.0 without building it acquire the tools while losing the ability to explain what the tools decide.

**DOI:** 10.35631/IJIREV.826054

**Keywords:**

Accountability; Algorithmic Opacity; Artificial Intelligence Governance; Higher Education; Industry 4.0; Industry 5.0; Innovation Leadership



© The authors (2026). This is an Open Access article distributed under the terms of the Creative Commons Attribution (CC BY NC) (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [ijirev@gaexcellence.com](mailto:ijirev@gaexcellence.com).

**Introduction**

Universities now use intelligent systems in decisions that carry heavy consequences for individuals: screening applications, upholding academic misconduct findings, flagging students at risk of failure and, in some settings, supporting grading. The leader who signs such a decision remains answerable for it, to the student, to the senate, to the regulator and to the public. The reasoning behind it, however, is frequently inaccessible. It cannot be reconstructed in detail, cannot be interrogated, and in many cases cannot be inspected at all under the terms of the supply contract. The leader must therefore answer for an outcome whose grounds are unavailable to them.

This is not a technical inconvenience. It cuts at the joint of what accountability has always been taken to mean. Bovens (2007) defines accountability as a relationship in which an actor is obliged to explain and to justify conduct to a forum, which may pose questions and pass judgement. The obligation to explain assumes that the actor can explain. Strip that away and the relationship does not simply weaken; it changes character. The actor can still be blamed, still be summoned, still be removed, but the explanatory core of the relationship has been hollowed out.

Higher education is a revealing place to examine this, for three reasons. First, the consequences named above are durable ones, and they accumulate: admission, progression, misconduct,

graduation. Second, universities have unusually elaborate governance machinery, senates, appeals committees, external examiners, and academic boards, all of which are built on the premise that decisions can be examined and justified. Third, adoption has been fast, and the field's own reviews say the surrounding debate has not caught up with it. Bearman et al. (2023), reviewing how the term is used across higher education journals, find few and confusing definitions and identify a discourse of altering authority in which these systems are positioned as decentring the teacher and spreading authority across staff, machines, corporations and students; they close by calling for research that traces accountability in artificial-intelligence-mediated practices. Williamson and Eynon (2020) set the present wave in a longer history and note how much of the debate remains unsettled. Work on institutional governance has meanwhile concentrated on what universities can do to build legal awareness and ethical literacy around these systems among students (Oncioiu & Bularca, 2025), which is a different question from what an officeholder owes when a system has already decided. Critical work on the limits of these technologies in education warns that what can be modelled statistically is narrower than what education actually does, and that the losses incurred in reorganising practice around such systems are easy to overlook (Selwyn, 2024). The arrival of generative tools has sharpened this in assessment specifically, where institutions have had to rebuild the relationship between student work and evidence of learning at speed (Perkins et al., 2024).

The problem belongs to the study of innovation as much as to the study of education, and that is where its contribution is meant to land. Industry 4.0 refers to the automation of production and decision-making through cyber-physical systems, data analytics and machine learning; Industry 5.0, as the European Commission has framed it, adds a human-centric, sustainable and resilient orientation to those same technologies (European Commission, 2021; Xu et al., 2021). Adoption of both is usually assessed by what an organisation gains: speed, scale, consistency, cost. Less attention is paid to what an organisation may quietly give up in the same transaction. This paper treats the capacity to account for a decision as an institutional asset that adoption can erode and argues that innovation leadership in this setting consists partly in noticing that erosion and building against it before consequential decisions are handed over.

The gap this paper addresses can be stated precisely, because three literatures approach it and none of them enters it. Work on algorithmic opacity establishes why the reasoning of these systems is inaccessible and to whom, but it is written about systems rather than about the officeholder who must answer for their output. Work on artificial intelligence in higher education examines adoption, readiness, capability and change management: how leaders should plan for adoption, govern its risks and align technology with institutional mission (Khairullah et al., 2025). It registers accountability as a concern and asks that it be traced (Bearman et al., 2023), but it does not say what accountability consists of once the tracing arrives at a person who cannot know the grounds. Accountability theory, for its part, is built on the premise that the actor who answers can furnish reasons for the conduct in question (Bovens, 2007), and it has not asked what the relationship consists of when that premise fails. What none of the three works out is the position of a leader who is fully answerable for a decision and structurally unable to know why it was made. Bearman et al. (2023) name the question; this paper answers it. That is the unexplained condition this paper takes as its subject, and it is an innovation leadership problem before it is an educational one, because it is created by the act of adoption itself. The question is therefore narrower than the adoption question and prior to it: when an institution has adopted such systems, what happens to the accountability of the people who must answer for what those systems do?

The objective of this paper is to develop a conceptual framework of accountability under algorithmic opacity. It names the condition accountability without comprehension and specifies a form of leadership fitted to it, which it calls answerability under opacity. The argument proceeds in three moves. It shows that leadership and accountability scholarship carries an unstated comprehension assumption, and that the literature on algorithmic opacity gives good reason to think that assumption cannot be repaired by better explanation. It defines the construct, specifies five dimensions, and differentiates it from adjacent constructs so that it is a contribution rather than a relabelling. And it advances propositions and boundary conditions, with an agenda for empirical work. The central claim is that accountability under opacity does not vanish but shifts its object: what the leader must justify is no longer the decision itself but the arrangements that produced it. The quality of a leader is accordingly visible in the strength of those arrangements rather than in a confidence about reasons they cannot actually possess. The theoretical contribution follows from that shift. Accountability theory has treated the giving of reasons as constitutive of the relationship (Bovens, 2007); this paper argues that the relationship survives when that element is removed, and specifies what takes its place, which extends the theory rather than abandoning it.

### **Theoretical Basis: The Comprehension Assumption in Leadership and Accountability**

Accountability, in the sense used across public administration and higher education governance, is relational and communicative. Bovens (2007) sets out its structure: an actor has an obligation to explain and justify conduct to a forum; the forum can interrogate and judge; consequences may follow. Every element of that sequence presupposes that the actor holds the reasons. Interrogation without the possibility of answer is not accountability but ritual.

Leadership scholarship inherits the same premise. Responsible leadership casts the leader as one who must justify conduct to a web of stakeholders (Maak & Pless, 2006), and stewardship theory describes an actor who feels answerable for what has been entrusted to them and acts on that felt responsibility (Davis et al., 1997; Hernandez, 2012). Both traditions assume the steward or the leader can give an account of what was done and why. Neither anticipates an actor who is fully responsible and structurally unable to explain.

Higher education governance embeds the assumption in its procedures. Academic appeal exists so that a decision can be re-examined on its reasons. External examining exists so that judgement can be checked against independent judgement. Misconduct hearings turn on evidence that can be shown and contested. These are not incidental arrangements; they are the mechanisms through which a university discharges its obligation to explain. When the substantive reason for a decision sit inside a system that no one in the room can inspect, these mechanisms continue to run but they no longer do the work they were designed to do.

It is worth being precise about what is missing. The leader is not merely uninformed in the ordinary way that any senior person is uninformed about operational detail. That kind of ignorance is remediable: the leader can ask, and someone in the organisation knows. Under algorithmic opacity there may be no one who knows. The gap is not in the leader's briefing; it is in the institution's epistemic position as a whole.

## Theoretical Basis: Why Better Explanation Does Not Close the Gap

The obvious response is to demand transparency, and it is the response most institutional policies reach for. The literature on algorithmic accountability suggests it is insufficient, for reasons worth stating carefully.

Burrell (2016) distinguishes three forms of opacity: opacity as intentional corporate or state secrecy, opacity as technical illiteracy on the part of those who encounter the system, and opacity that arises from the mismatch between the scale and logic of machine learning and the demands of human reasoning. Only the first two are addressed by disclosure and training. The third persists even when code and data are fully available, because the operations that produce a classification do not decompose into reasons of the sort a human forum can weigh.

Ananny and Crawford (2018) press the point further, arguing that the transparency ideal itself is limited: seeing inside a system does not confer the ability to understand or govern it, and can produce the appearance of accountability without its substance. Kemper and Kolkman (2019) add a condition that is directly relevant to universities: transparency accomplishes nothing without a critical audience capable of interpreting what is disclosed. A senate that receives a model card it cannot evaluate has been given disclosure, not accountability.

Nor does the technical field promise rescue. Rudin (2019) argues that post-hoc explanations of black-box models are approximations that can mislead, and that for high-stakes decisions interpretable models should be used instead of explained opaque ones. The argument matters here because it reframes the leadership question. If faithful explanation of an opaque system is not generally available, then a leader who insists on being able to explain every automated decision is not setting a governance standard; they are, in effect, deciding which systems the institution may use. That is a legitimate leadership choice, and this paper treats it as one of the choices the construct makes visible, rather than as a technical detail.

There is also reason to doubt the standard institutional remedy of human oversight. Green (2022), reviewing policies that require a human in the loop over government algorithms, finds two flaws: people are generally unable to perform the oversight functions the policies assume, and the presence of the requirement legitimates the use of flawed systems without constraining them. The clinical literature on automation bias points the same way, showing that decision-makers systematically defer to automated advice, including when it is wrong (Goddard et al., 2012). Regulation has nonetheless settled on oversight as its instrument: the European Union's artificial intelligence regulation requires that high-risk systems be capable of effective oversight by natural persons with the competence, training and authority to exercise it (European Parliament & Council of the European Union, 2024). The requirement is reasonable and the evidence suggests it is hard to satisfy. That gap between the formal requirement and the practical capacity is precisely where leadership becomes consequential, and where the difference between substantive and ceremonial answerability is decided. The term ceremonial is used throughout in this specific sense: an arrangement is ceremonial when it has the form of accountability, a committee, a sign-off, an appeal route, without the authority, competence or information needed to change an outcome.

Wieringa (2020), reviewing the algorithmic accountability literature systematically, shows that accounts of accountability in this field are fragmented across actors, forums and stages of the system life cycle, and that responsibility is easily diffused among vendors, developers,

procurers and users. Diffusion is the danger this paper takes seriously. When no single actor holds the reasons, the path of least resistance is for each to point at another, and for the institution to answer for nothing at all.

Philosophical work on responsibility gaps sets out what is at stake with more precision. Santoni de Sio and Mecacci (2021) argue that the problem is not one gap but several: a culpability gap, a moral accountability gap, a public accountability gap, and a gap in active responsibility, each with different technical, organisational and legal sources. Their public accountability gap, in which those affected by a decision cannot obtain an explanation from the body that made it, names precisely the exposure a university carries when a student asks why. Their analysis is directed at how such gaps might be closed through design, regulation and the allocation of duties across a system. The question this paper takes up begins where that leaves off: what a named institutional leader is to do while the gap remains open, as it will for most universities using procured systems for the foreseeable future.

### **The Construct: Defining Answerability Under Opacity**

Accountability without comprehension is defined here as the condition of an institutional leader who remains formally and morally answerable for decisions whose substantive reasoning is inaccessible to them and, frequently, to the institution as a whole. Answerability under opacity is the form of leadership fitted to that condition: the work of sustaining an institution's capacity to answer for automated and semi-automated decisions when the reasons behind those decisions cannot be personally held, exercised through the arrangements that surround the system rather than through knowledge of its internal logic.

Three elements of the definition carry the argument. The first is that answerability is retained rather than transferred. The leader does not become less responsible because a machine decided; the obligation is unchanged while the capacity to discharge it in the usual way is gone. This is what distinguishes the condition from ordinary delegation, where responsibility is shared with an agent who can themselves be asked to explain.

The second is that accountability migrates rather than disappears. Migration here means a change in what must be justified, not a change in who is answerable; the officeholder does not move, the object of their justification does. If the leader cannot justify the decision, what remains available is justification of the arrangements: how the system was chosen and validated, what it is permitted to decide, what independent evidence exists about how it performs, what route an affected person has to challenge it, and under what conditions a human being takes the decision back. This is a real and demanding form of accountability, and it is not the same as the one the governance machinery was built for.

The third is that the arrangements must bind. An arrangement that cannot constrain the system is decoration. The distinction between substantive and ceremonial answerability runs through the whole construct: the same procedural furniture, an oversight committee, a human sign-off, an appeals route, can either restrain a system or launder it (Green, 2022). What separates them is not the existence of the procedure but whether it carries the authority, the competence and the information to say no.

## **The Construct: Dimensions of Answerability Under Opacity**

The construct can be specified through five dimensions. Together they describe how an institution keeps its ability to answer when its reasons are no longer its own.

### ***Retained Answerability***

The first dimension is a refusal to displace responsibility onto the system or its supplier. Under conditions of diffusion (Wieringa, 2020) the available excuses are plentiful: the vendor built it, the algorithm decided, the data said so. Retained answerability is the leadership stance that treats the institution as the author of any decision issued in its name, whatever produced it. It is the point at which this form of leadership either exists or does not, because every subsequent dimension is an expression of it.

### ***Procured Comprehension***

The second dimension follows from the impossibility of personal understanding. If the leader cannot know how the system reasons, the institution must acquire what independent knowledge is available: external audit, validation on local data, error analysis disaggregated by student group, and documented evidence of performance in the conditions where it is actually used. This is not the leader understanding the model; it is the institution buying the best available account of its behaviour from parties competent to give one. Its value depends on the critical audience condition (Kemper & Kolkman, 2019): evidence procured but not intelligible to the body receiving it produces disclosure without accountability.

### ***Scope Discipline***

The third dimension is the work of holding a system to the boundary within which it was validated. Systems migrate: a tool procured to triage enquiries is used to rank applicants; a similarity detector becomes a determination of misconduct; a risk flag becomes a reason for exclusion. Each migration extends consequence beyond the evidence base, and each is usually undertaken for reasons of convenience rather than by decision. Scope discipline is the refusal to let a system decide questions it was never shown to be competent to decide, and it is among the few protections that remain fully within the institution's control.

### ***Explanation Owed to Those Affected***

The fourth dimension turns on the person at the receiving end. A student refused admission or found to have cheated is owed something, and the honest position is that the institution cannot always supply a full account of the reasoning. What it can supply is candour about the system's role, an account of the arrangements under which it was used, and a route of challenge that reaches a human being with authority to decide differently. Ananny and Crawford (2018) warn that visibility without understanding can substitute for accountability; the obligation here is not to perform transparency but to preserve a real avenue of contest.

### ***Reversion Authority***

The fifth dimension is reversion authority, by which is meant the practical capacity to return a decision to human hands and decide it differently. This means more than a nominal right of

override. It requires that someone holds the authority to set aside a machine output, that they have the competence and the information to judge, that the organisation does not penalise them for exercising it, and that defined triggers exist at which reversion is mandatory rather than discretionary. The evidence on automation bias (Goddard et al., 2012) and on the weakness of oversight requirements in practice (Green, 2022) suggests that reversion authority decays quietly unless it is deliberately maintained, and that its decay is the mechanism by which substantive answerability becomes ceremonial.

### ***Illustration From the Public Record***

A documented episode shows the dimensions together and shows what their absence costs. When examinations in England were cancelled in 2020, the regulator replaced them with grades produced by a standardisation algorithm applied to teacher assessments and historical centre performance. Kelly (2021), comparing the English and Irish cases, records the sequence: a model selected from several candidates on grounds that were not publicly deliberated, grades issued to individual students who could not be told why their own result had been lowered, appeal routes that initially returned the same calculated outcome, and a reversal only after protest, legal threat and political pressure. Ireland, which adopted a comparable approach, diverged in its handling and emerged with far less damage.

Read through the construct, the failures are legible. Answerability was formally retained by the regulator and the ministry, but the account they could give was about process rather than about any individual determination, which is the migration this paper describes, occurring by default rather than by design (retained answerability and procured comprehension). Scope discipline was absent in a particular way: a model built to hold aggregate distributions steady was used to issue determinations about named individuals, a purpose for which aggregate reasoning was never adequate (scope discipline). Explanation owed to those affected failed most visibly, since students received consequential outcomes with no account they could interrogate, and the route of challenge at first returned the same machine result (explanation owed to those affected). Reversion authority existed in principle and was exercised only at the point of political crisis, which is the difference between a trigger designed in advance and a capitulation extracted after the fact (reversion authority).

The case also illustrates the critical audience condition. What eventually broke the arrangement was not internal governance but external scrutiny from teachers, students, statisticians and journalists who could interrogate the model's effects, which is precisely the mechanism Kemper and Kolkman (2019) identify as the precondition of algorithmic accountability. Universities should not assume that such an audience will assemble for their own systems, which operate at smaller scale and attract less attention. The episode is offered as an illustration of the concept and not as evidence for it; the construct must still be tested in the ways set out below.

### ***Differentiating the Construct***

A construct earns its place only if it is not an existing one renamed. Answerability under opacity is related to, but distinct from, five neighbours, and from two other constraints on responsibility with which it might be confused.

From accountability theory (Bovens, 2007) it takes its relational structure, but that theory describes the obligation to explain and assumes the capacity to do so; the construct here specifies what leadership becomes when the capacity is absent while the obligation remains. From responsible leadership (Maak & Pless, 2006) it takes an ethical orientation toward those affected, but responsible leadership is a theory of relationships among stakeholders and does not address a leader's epistemic position. From stewardship theory (Davis et al., 1997; Hernandez, 2012) it takes the felt responsibility for what is entrusted and departs from it in the same respect: stewardship theorises motivation and posture, not the problem of answering for a judgement one cannot examine. And from the algorithmic accountability literature (Ananny & Crawford, 2018; Wieringa, 2020) it takes its diagnosis of opacity and diffusion, but that literature is written about systems, regulators and publics rather than about the institutional leader who must sign, and it stops short of specifying what such a leader should therefore do.

A fifth neighbour is the responsibility-gap literature itself (Santoni de Sio & Mecacci, 2021). It shares this paper's diagnosis and differs in its object and its addressee. That work asks how responsibility gaps might be closed, and addresses designers, regulators and policymakers who can change the conditions under which systems are built and deployed. This paper asks how an institutional leader should act while the gap stays open, and addresses the person who must sign. The difference is not merely one of audience. A theory written for those who can close the gap can treat the open gap as a defect to be remedied; a theory written for those who must work inside it has to specify what legitimate practice looks like when the remedy is unavailable, which is what the five dimensions set out to do.

The construct must also be distinguished from two adjacent conditions in which responsibility outruns control. Responsibility can be constrained by jurisdiction, where a leader is answerable for people or assets outside the legal reach of their authority. It can be constrained by hierarchy, where a leader is answerable for outcomes they lack the formal power to command. Answerability under opacity is constrained by neither. The leader here has both jurisdiction and authority; what they lack is epistemic access. That distinction matters because the remedies differ: brokering and negotiation address those two constraints, and neither does anything for this one.

It is fair to press harder, because a label is worth little if it only relabels a familiar worry about technology. What makes this more than a restatement of the black-box problem is the mechanism it names. In answerability under opacity the leader cannot ground legitimacy in the reasons for a decision, so legitimacy must be constructed out of the arrangements surrounding it, and the institution's account shifts from why this decision was right to why this system was permitted to make it, within these limits, checked in these ways, reversible on these triggers. Accountability is thereby displaced one level upward, from decision to arrangement. The algorithmic accountability literature has described the loss of first-order explanation; the leadership literature has not treated the deliberate construction of second-order justification as the act by which leadership over opaque systems is exercised. That displacement, and the standard it implies, that a leader is judged by the bindingness of arrangements rather than by confidence in reasons, is what is genuinely new here.

## Propositions

The dimensions imply relationships that can be examined empirically. The propositions below are offered as conjectures for testing rather than as demonstrated claims, and each is stated so that it could be found wrong.

P1. Where leaders retain answerability rather than attributing decisions to the system or its supplier, institutions are more likely to establish binding arrangements around automated decisions, and less likely to rely on procedural formalities that cannot constrain the system.

P2. Institutions that procure independent evidence about a system's behaviour, and that place that evidence before a body competent to interpret it, sustain a stronger account of automated decisions than institutions that rely on vendor assurance, and the difference widens as the opacity of the system increases.

P3. Systems used beyond the scope in which they were validated generate more contested decisions and more downstream reversals than systems held within scope, and scope migration is more likely where no named actor holds authority over permitted use.

P4. Where those affected have a route of challenge that reaches a human being with authority to decide differently, institutional legitimacy is better sustained under opacity than where challenge routes return the original automated determination, even when the substantive outcomes are similar.

P5. Reversion authority decays over time unless it is actively maintained through defined triggers, protected discretion and competence support, and the rate of decay is higher where staff are evaluated on throughput than where they are evaluated on judgement.

P6. Human oversight arrangements that lack authority, competence or information function ceremonially, legitimating the system without constraining it, and their presence is associated with weaker rather than stronger protection for those affected.

## Scope and Boundary Conditions

The construct is not a claim about all technology in all organisations. It applies where three conditions hold together: a decision carries material consequence for an identifiable person; the institution is formally obliged to account for that decision; and the reasoning behind it is inaccessible to those who must answer for it. Admissions, academic misconduct, progression and appeals are the paradigm cases in higher education. Timetabling, room allocation and routine administrative automation fall outside it, not because they are unimportant but because little turns on explanation.

It bears less directly on interpretable systems, where reasons can be reconstructed and ordinary accountability still functions, and this is a deliberate boundary rather than a hedge: it implies that choosing an interpretable system is itself a way of avoiding the condition (Rudin, 2019). It weakens where the institution genuinely holds internal expertise sufficient to interrogate the system, and it intensifies where systems are procured from external vendors under commercial confidentiality. It is also shaped by regulatory setting: where oversight obligations are legally specified (European Parliament & Council of the European Union, 2024), the risk of

ceremonial compliance is higher, because the institution can satisfy the letter of the requirement without acquiring the capacity it assumes.

## **Discussion: Contributions and Implications**

The paper makes five contributions. To leadership scholarship it identifies a constraint on responsibility that the field has not theorised: not distance, not the limits of formal authority, but the absence of epistemic access, and it specifies a form of leadership fitted to it. To accountability theory it shows that the explanatory core of the accountability relationship (Bovens, 2007) can be severed while the relationship persists, and that what fills the space is justification of arrangements rather than justification of conduct. To the algorithmic accountability literature (Ananny & Crawford, 2018; Burrell, 2016; Wieringa, 2020) it supplies the institutional actor that literature largely leaves out, moving from the question of whether systems can be held accountable to the question of how a leader who must answer for them should act. And to research on artificial intelligence in higher education (Bearman et al., 2023; Khairullah et al., 2025; Oncioiu & Bularca, 2025; Williamson & Eynon, 2020) it offers a governance vocabulary grounded in the university's own obligations to students rather than in generic readiness and adoption. A fifth contribution returns to where the argument began. To innovation leadership it identifies a cost of adopting intelligent systems that adoption frameworks rarely price, namely the loss of an organisation's ability to explain its own decisions, and it argues that protecting that ability is a leadership task rather than a technical one. Taken together, the five position the capacity to account for an intelligent system as part of the innovation rather than as a constraint upon it: an institution that acquires Industry 4.0 and Industry 5.0 capability without acquiring this one has innovated in the narrow sense and lost ground in the broad one.

For practice the implications are concrete. Institutions should name, for every consequential automated decision, the person who answers for it, and should resist arrangements in which that person is nominal. They should commission independent evidence about system behaviour and route it to a body able to read it. They should record the validated scope of each system and treat extension beyond it as a decision requiring authorisation rather than as a matter of convenience. They should ensure that a challenge from a student reaches a human being empowered to decide differently. And they should protect reversion authority explicitly, since the evidence suggests it will otherwise erode. The common thread is that leadership under opacity is exercised in advance, through the design of arrangements, rather than at the moment of decision, when it is already too late to understand.

## **Research Agenda**

Because the paper is conceptual, its purpose is to open inquiry. The dimensions await operationalisation: instruments could be developed to capture retained answerability, procured comprehension, scope discipline, explanation owed and reversion authority, and to distinguish substantive from ceremonial arrangements, which is the distinction on which most of the argument turns. Comparative case studies of universities that have adopted admissions, integrity or risk systems would show how these arrangements are actually made, by whom, and with what authority, and document work is a natural starting point, since policies, committee minutes and appeal records leave a trace of where explanation was demanded and what was supplied.

The propositions invite testing across settings that differ in regulation, in institutional autonomy, and in whether systems are built in-house or procured. Studies of appeal and complaint outcomes could examine whether challenge routes function as the construct requires. Longitudinal work is needed on the decay of reversion authority, since the claim is about erosion over time rather than about a state. Comparative work between higher education and other high-stakes public settings would test whether the construct travels. None of this is settled by argument alone, and the construct is put forward in the expectation that empirical work will sharpen it, qualify it, or show it to be wrong.

## **Conclusion**

Universities are being asked to answer for decisions they cannot fully explain, and the pace of adoption suggests the gap will widen before it narrows. The response cannot be to insist on a comprehension that is not available, nor to accept that responsibility dissolves because a machine was involved. This paper has argued for a third position: that accountability under opacity migrates from the decision to the arrangements that produce it, and that the work of building arrangements that genuinely bind, rather than arrangements that merely satisfy a policy, is leadership work of a recognisable and demanding kind. Naming it matters, because what is not named is not appraised, not taught, and not required of the people who sign. The propositions offered here are an invitation to test whether the concept holds; the wager of the paper is that a task so consequential, and so quietly arriving, is worth the theorising.

- 
- Acknowledgements:** The authors thank Universiti Teknologi MARA and Politeknik Melaka for institutional support, and colleagues who commented on earlier drafts of this manuscript.
- Funding Statement:** No Funding.
- Conflict of Interest Statement:** The authors declare that there is no conflict of interest regarding the publication of this paper. All authors have contributed to this work and approved the final version of the manuscript for submission to the International Journal of Innovation and Industrial Revolution (IJIREV).
- Ethics Statement:** This is a conceptual paper. It involved no human participants, no animal subjects, and no collection of primary data, and therefore did not require institutional ethics review. The illustrative material is drawn entirely from published public-record sources, which are cited in the text.
- Author Contribution Statement:** All authors contributed significantly to the development of this manuscript. Fahmi Abdul Rahim was responsible for conceptualisation, theory development and drafting of the manuscript. Noor Shuhailie M Mohamed Noor contributed to the literature review and to the specification of the construct dimensions. Wan Hasmat Wan Hasan contributed to the theoretical framing and critical revision. Fadhlina Izzah Saman contributed technical review of the algorithmic opacity material. Afzan Nor Talib contributed to the propositions and to manuscript review. All authors read and approved the final manuscript.
- Declaration on the Use of Artificial Intelligence (AI)** In preparing this manuscript the authors used a generative AI assistant (Anthropic's Claude) only for language editing and proofreading of their own original text. It was not used to generate the argument, the conceptual framework, the propositions or the interpretations, which are the authors' own. The authors reviewed and take full responsibility for the content, originality and integrity of the paper. No AI tool is credited as an author, consistent with COPE guidance.
-

## References

- Ananny, M., & Crawford, K. (2018). Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. *New Media & Society*, 20(3), 973-989. <https://doi.org/10.1177/1461444816676645>
- Bearman, M., Ryan, J., & Ajjawi, R. (2023). Discourses of artificial intelligence in higher education: A critical literature review. *Higher Education*, 86(2), 369-385. <https://doi.org/10.1007/s10734-022-00937-2>
- Bovens, M. (2007). Analysing and assessing accountability: A conceptual framework. *European Law Journal*, 13(4), 447-468. <https://doi.org/10.1111/j.1468-0386.2007.00378.x>
- Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, 3(1), 1-12. <https://doi.org/10.1177/2053951715622512>
- Davis, J. H., Schoorman, F. D., & Donaldson, L. (1997). Toward a stewardship theory of management. *Academy of Management Review*, 22(1), 20-47. <https://doi.org/10.5465/amr.1997.9707180258>
- European Commission, Directorate-General for Research and Innovation. (2021). Industry 5.0: Towards a sustainable, human-centric and resilient European industry. Publications Office of the European Union. <https://data.europa.eu/doi/10.2777/308407>
- European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121-127. <https://doi.org/10.1136/amiajnl-2011-000089>
- Green, B. (2022). The flaws of policies requiring human oversight of government algorithms. *Computer Law & Security Review*, 45, 105681. <https://doi.org/10.1016/j.clsr.2022.105681>
- Hernandez, M. (2012). Toward an understanding of the psychology of stewardship. *Academy of Management Review*, 37(2), 172-193. <https://doi.org/10.5465/amr.2010.0363>
- Kelly, A. (2021). A tale of two algorithms: The appeal and repeal of calculated grades systems in England and Ireland in 2020. *British Educational Research Journal*, 47(3), 725-741. <https://doi.org/10.1002/berj.3705>
- Kemper, J., & Kolkman, D. (2019). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, 22(14), 2081-2096. <https://doi.org/10.1080/1369118X.2018.1477967>
- Khairullah, S. A., Harris, S., Hadi, H. J., Sandhu, R. A., Ahmad, N., & Alshara, M. A. (2025). Implementing artificial intelligence in academic and administrative processes through responsible strategic leadership in the higher education institutions. *Frontiers in Education*, 10, 1548104. <https://doi.org/10.3389/feduc.2025.1548104>
- Maak, T., & Pless, N. M. (2006). Responsible leadership in a stakeholder society: A relational perspective. *Journal of Business Ethics*, 66(1), 99-115. <https://doi.org/10.1007/s10551-006-9047-z>
- Oncioiu, I., & Bularca, A. R. (2025). Artificial intelligence governance in higher education: The role of knowledge-based strategies in fostering legal awareness and ethical artificial intelligence literacy. *Societies*, 15(6), 144. <https://doi.org/10.3390/soc15060144>

- Perkins, M., Furze, L., Roe, J., & MacVaugh, J. (2024). The Artificial Intelligence Assessment Scale (AIAS): A framework for ethical integration of generative AI in educational assessment. *Journal of University Teaching and Learning Practice*, 21(6), 49-66. <https://doi.org/10.53761/q3azde36>
- Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, 1(5), 206-215. <https://doi.org/10.1038/s42256-019-0048-x>
- Santoni de Sio, F., & Mecacci, G. (2021). Four responsibility gaps with artificial intelligence: Why they matter and how to address them. *Philosophy & Technology*, 34(4), 1057-1084. <https://doi.org/10.1007/s13347-021-00450-x>
- Selwyn, N. (2024). On the limits of artificial intelligence (AI) in education. *Nordisk Tidsskrift for Pedagogikk og Kritik*, 10(1), 3-14. <https://doi.org/10.23865/ntpk.v10.6062>
- Wieringa, M. (2020). What to account for when accounting for algorithms: A systematic literature review on algorithmic accountability. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency* (pp. 1-18). Association for Computing Machinery. <https://doi.org/10.1145/3351095.3372833>
- Williamson, B., & Eynon, R. (2020). Historical threads, missing links, and future directions in AI in education. *Learning, Media and Technology*, 45(3), 223-235. <https://doi.org/10.1080/17439884.2020.1798995>
- Xu, X., Lu, Y., Vogel-Heuser, B., & Wang, L. (2021). Industry 4.0 and Industry 5.0: Inception, conception and perception. *Journal of Manufacturing Systems*, 61, 530-535. <https://doi.org/10.1016/j.jmsy.2021.10.006>