SAFWAN ALMODHAYAN



OPINION AND ANALYSIS

Editorial date: October 31, 2026

The Late Signature: Who Owns a Decision Made by a Machine?

> In the summer of 2020, an examination grade looked like a mathematical answer to an educational question. But the search for who had “decided” that grade led to a chain of people, each holding part of the authority while none held the whole picture. This is not only a story about a computer error. It is about an institution that divided the decision and forgot to reunite responsibility.

The Grade No Student Wrote on an Examination Paper

On August 13, 2020, A-level results were announced in England. These grades shape thousands of students’ progression to university, but the examinations had not taken place that summer. On March 18, the Covid-19 pandemic had led the British government to cancel them. A replacement mechanism used schools’ estimates and student rankings, with statistical processing intended to align the distribution of results with previous years.[1][2]

The grade shown to a student was definite. The route to it was distributed: teachers submitted estimates and rankings; a regulator designed the standardization method; examination boards issued results under formal requirements; and a ministry set the policy objective. No robot sat in an examiner’s chair. There was no written examination to mark. The state was trying to create a substitute for an educational event that had not occurred.

Four days later, on August 17, the rule changed. Where relevant, the final result became the higher of the school estimate and calculated grade. In Ofqual's later analysis of a defined sample of 18-year-olds in England—**457,420 subject entries** from **246,110 students**—the calculated grade matched the school estimate in about **59 percent of entries**, was lower in **39 percent**, and higher in about **2 percent**. These were subject-entry rates, not student rates; it would be inaccurate to say that 39 percent of students had their grades reduced.[3]

A difference from the teacher estimate does not prove the teacher right or the computer unjust in every case. It proves that the system could issue a judgment with major personal effects, then change the basis of that judgment by institutional decision within days. Who owned the first decision? Who bore its consequences? Who could prevent it before it reached the student?

On September 2, 2020, Ofqual chair Roger Taylor wrote to Parliament's Education Committee that the regulator **accepted its full share of responsibility**. He explained that the decision to use statistically standardized teacher estimates followed ministerial direction, while implementation was the regulator's responsibility. He reached a telling conclusion: improving the algorithm alone could not make the results acceptable; limitations in the data and the nature of the process were part of the problem.[4]

Calling this an "algorithm scandal" and moving on is easy. The label strangely absolves humans. An equation does not issue ministerial direction, decide that fairness means similar grade distributions across two years, or choose the acceptable harm when an individual result differs from a student's expectation. Human decisions preceded the calculation.

A machine may execute a choice. It does not erase the choice's human history.

Four Authorities Hidden Inside One Word

Institutions say that "a decision was made" as though one clear owner existed. An institutional decision is not a point but a passage among rights: to propose an action, approve the decision rule, authorize the result, and stop or correct it.

In "Separation of Ownership and Control," published in 1983, Eugene Fama and Michael Jensen distinguished **decision management**—initiating and implementing choices—from **decision control**—ratifying choices and monitoring performance. Their subject was organizational structure and conflicts between owners and managers, not AI. The distinction helps reveal what a modern one-click result conceals.[5]

Consider a **hypothetical** bank. A model predicts an applicant's default risk. The technical team produces the estimate; risk staff choose the acceptance threshold; management approves the tolerable loss; the operating system sends notice; customer service explains it; and the licensed

bank is the entity in a legal and contractual relationship with the customer. Who declined the application?

Two answers can be correct to two questions. **External legal responsibility** follows applicable law, contract, and facts, and does not disappear because a model came from a vendor. **Internal causal capacity** may belong to whoever approved a variable, set a threshold, or authorized operation without review. Neither description replaces the other; a serious investigation needs both.

Layer	Question	Required authority	What does not prove it
Define the purpose	What interest may the system optimize?	Approve objectives, constraints, and non-tradable rights	Buying a “high-accuracy model”
Build the means	What data, assumptions, and costs turn purpose into calculation?	Test the model, alternatives, and limits	Showing confidence scores
Give effect	When does a recommendation become an influential decision?	Approve thresholds, permissions, and automation limits	An automated departmental signature
Correct and remedy	Who can stop, annul, or reconsider the effect?	Effective authority, time, resources, and an appeal path	An objection inbox whose owner can change nothing

This is not a new list of legal duties in any jurisdiction. It is a proposed operating map for tracing them. Failure arises when the four powers are separated without clear connections: the model reviewer cannot change the purpose; the person signing the purpose does not understand the equation; the person facing the customer can only read a predetermined reason.

Written Authority Is Not Always Real Authority

Philippe Aghion and Jean Tirole’s 1997 paper “Formal and Real Authority in Organizations” develops a simple, deep idea: holding the formal right to decide does not ensure the practical capacity to choose, especially when subordinates possess information that a principal cannot absorb in time.[6]

Complex models widen that separation. A risk director may approve policy but not know what removing one variable does across thousands of cases. A data scientist may know the effect but lack authority to decide whether it is acceptable in credit or law. “We approved the model” can cover three different acts: viewing a presentation, understanding risk, and accepting consequences. Doing the first does not establish the other two.

The principal–agent problem then appears. A board delegates execution to management; management delegates technical work to a team; the team may delegate hosting or operation to a vendor. Each link pursues a legitimate local goal: deliver on time, reduce defaults, shorten response time, or raise approval rates. Agreement on local objectives does not guarantee that the final decision serves the institution's purpose.

The formal authority holder may become the last to know what that authority permits. This need not be conspiracy or neglect; it follows from uneven knowledge, time, and means of influence. Unless an institution deliberately returns real power to those who bear responsibility, "human review" may become a signature arriving after reversal is costly.

When the Equation Is Right and the Question Is Wrong

Returning to England, stabilizing the 2020 grade distribution against previous years was not a senseless objective. Accepting unadjusted school estimates could produce differences in generosity between schools, grade inflation, and distorted comparisons between student cohorts. Ofqual stated these concerns in its records. Conversely, statistical standardization turned some differences between schools into an effect on the grade of an individual student who had no opportunity to demonstrate their level in an actual examination.14

The model could therefore be consistent with its specified objective while the objective itself was insufficient to ground an acceptable decision. The difference between **a calculation being valid within a model** and **the legitimacy of using it to judge an individual** is not processor speed. It is a difference in the question first put to the processor.

The episode exposes a risk difficult to see from inside technical teams. Designers may measure orderly result distributions; government, educational continuity; universities, admission readiness; and students, a fair opportunity to demonstrate learning. These are not measurements of one thing but interests and rights that can conflict. One objective function cannot reconcile them automatically. Their order and permissible sacrifice require an explicit political or institutional decision.

After the crisis, the UK Office for Statistics Regulation published a review on March 2, 2021. It stressed distinguishing the responsibility of the body that commissions a model and defines its purpose from that of the developer, assessing fitness for purpose and public acceptability, and defining the basis and timing of challenge. Its recommendation was not to abandon models, but to stop reducing public decisions to what looked like a technical exercise.7

What Does a Model Actually Decide in a Loan Application?

For illustration only, suppose a bank uses a model predicting repayment probability $p(x)$, where x is available application information. Let B be additional return if the loan is repaid, L the loss if it defaults, and C the service cost. A simplified risk-neutral rule might be:

Accept the application if

$$p(x)B - (1 - p(x))L - C > 0.$$

This is neither a real banking rule nor credit advice. Actual lending is more complex and constrained by liquidity, capital, regulation, equal treatment, model risk, and customer rights. The equation reveals one point: a model may produce the **prediction**, but it does not necessarily determine B, L, or C, or whether the institution may define its objective this way. Choosing "expected profit" alone is itself contestable.

"Predicts repayment" also hides an epistemic issue. A prediction is not a promise; repayment probability is not moral capacity to fulfil an obligation, nor necessarily legal eligibility for finance. Unrepresentative data or credit histories formed under old standards may carry prior decisions into the number. Recognizing these limits does not remove the need to decide; it changes what may be claimed about the basis.

Imagine an estimated probability of 0.79 and a managerial threshold of 0.80. A data analyst cannot tell the customer merely that 0.79 was "the model's order." A human decision made the difference consequential. If capital cost or the risk budget changes, the threshold may move while model and customer data remain unchanged. An institution can create new decisions by moving a policy control, then misattribute them to system "intelligence."

This matters to applicants seeking a reason they can inspect and correct, risk staff asked to explain a choice they did not design, shareholders who need to know who changed risk appetite, and supervisors for whom a model name in a registry is insufficient.

Why "We Don't Understand the Black Box" Is Not an Excuse

US credit rules provide a useful distinction **within their jurisdiction only**. Section **§1002.9** of **Regulation B**, in its current published form, governs adverse-action notices and specific reasons under conditions varying by application type. Official interpretation says statements such as failing to achieve a qualifying score are insufficient by themselves; reasons must relate to factors actually used.[8]

The Consumer Financial Protection Bureau issued a 2022 circular saying algorithmic opacity was no excuse for failing to give reasons. That circular and a related 2023 circular were **withdrawn on May 12, 2025**. Its former interpretation should not be represented as current

policy. The point is to distinguish withdrawn guidance from the continuing regulation, not pronounce on a bank outside the United States.9

Feature-attribution methods may explain a model's prediction, but an explanation of calculation is not necessarily an explanation of an institutional decision. The first indicates which inputs contributed to a number under a method; the second must explain why the bank chose the number as a standard, how customer data were represented, and what policy gave the result effect. Approximate methods may disagree and do not always represent causal relations.

Conversely, a system can reveal every computational reason while the decision remains unacceptable. Who authorized historical data for this purpose? Who decided a particular false rejection should fall entirely on the customer? Transparency is an important window, but it does not make an unlawful decision lawful. Knowing why a door closed is not possessing its key.

The Institutional Error of Employing a "Human Approver"

Many systems promise a human in the loop. The device can be useful, but the name says nothing about the person's authority. Can they see relevant facts, understand model limits, depart from the recommendation without implicit penalty, take enough time, and prevent effect until review ends?

Oversight is not created by adding an "approve" box. Facing hundreds of similar recommendations, an employee may become a stamp rather than a judge. The institution inserts a human to prove one remains, then designs the work so disagreement is impossible.

NIST's voluntary 2023 **AI Risk Management Framework** addresses actual responsibility: **GOVERN 2.1** calls for documented roles and communication lines, and **GOVERN 2.3** assigns executive leadership responsibility for decisions about system development and deployment risk. The framework is **voluntary**, not a law conferring powers people otherwise lack.10

Article 14 of the **EU AI Act** says oversight of high-risk systems should, as context requires, understand capabilities and limits, disregard or reverse outputs, and intervene or stop the system. Application is phased. According to the European Commission after 2026 amendments, provisions for several Annex III uses begin **December 2, 2027**, and regulated-product systems **August 2, 2028**. It would be wrong to treat this as a current European duty of a Saudi bank or assume every clause covers every model.11

Beyond geography, the design lesson is that human oversight is not a worker's location on a flow chart. It is **practical power to object before objection loses value**.

Saudi Arabia: From Allocating Responsibilities to Testing Them

The Saudi Central Bank's **Key Principles of Governance in Financial Institutions Subject to its Oversight and Supervision**, issued June 30, 2021 and updated in a second edition on May 8, 2024, establish current roles for boards, executive management, committees, control, and risk management. They are not themselves a regulation specifically about signing AI-model decisions. Applying them requires translating their meaning into testable authority and procedure without inventing textual obligations.12

A Saudi bank board might ask in a proposed internal review what share of decisions "AI made." The sharper questions are: when does model output become an effect changing a right or benefit; who approved that stage; what loss can occur before an authorized employee intervenes; and can the recipient of a correction request change the result or only send another email?

If risk approves a model and product staff change its threshold without proper authorization, an accuracy test alone will not find the defect in **decision rights**. If a vendor can change the production version without suitable contractual and technical controls, the bank's name on a refusal does not show where real control went. If a system learns from new interactions, its behavior may change while the approval for an earlier version rests serenely in a compliance folder describing a world that no longer exists.

A December 2024 Financial Stability Institute/BIS study identifies financial-sector AI challenges including governance, model management, and third-party dependence. It does not judge every Saudi project, but shows that divided knowledge between bank and vendor is not a marginal local question.13

Executable Correction Authority: A Proposed Concept, Not a Ready Certificate

I propose **executable correction authority** as an analytical tool, **not a recognized academic theory or existing regulatory standard**. For a consequential decision, it joins three conditions: a named person responsible for review; enough information to understand the basis and limits; and practical and legal power to change or stop the result while correction still matters.

An institution can test it without inventing a sacred metric. Select a declared sample of real consequential decisions, replay a legitimate challenge safely, and ask: could the authorized official be reached; could they reconstruct material reasons without guessing; and, on proven

error, could they enforce correction within a stated time? Report the share passing **all three conditions**, distributed by decision type and severity rather than an average hiding the worst harms.

The measure **does not prove the original decision sound**, create a legal right to reverse the irreversible, or resist gaming through easy samples and permissive deadlines. It exposes one aspect of accountability: whether correction can take effect. The testable hypothesis is that systems defining this right early and testing it periodically will have fewer **wrong decisions that persist after discovery** than systems with only a nominal complaints route. Controlled field evidence is required before treating that as a finding.

Four questions are more honest than “is there a human in the loop?” Can the authority see decisive data, challenge policy as well as result, stop an urgent decision before an irreversible effect, and provide a challenge route intelligible without internal modeling language? Joint ICO and Alan Turing Institute guidance calls for explaining AI-assisted decisions to affected people; explanation is part of the institutional relationship, not merely a display of mathematics.14

The Strongest Objection: Humans Are Not Fair Merely Because They Are Human

The necessary alternative is not returning every file to an employee’s intuition. Humans can be inconsistent and tired; judgments vary with mood, experience, and workload. A well-tested model using consistent criteria can reduce some disparities and catch errors manual review misses. Human review of every case may itself harm applicants through delay. This is a serious argument.

Nor does 2020 prove every statistical decision invalid. The Office for Statistics Regulation acknowledged that officials sought accurate outcomes in good faith under exceptional time and data limits, and that abandoning statistical models generally could forfeit public benefit.7 Human decisions may also be inexplicable when reasons are undocumented.

The objection corrects implementation rather than defeating the thesis. Accountability does not require humans to repeat every calculation. Oversight can follow risk: clear automation for simple reversible cases; review at sensitive boundaries; mandatory stops beyond delegation; and independent sampling, including cases employees approved unchanged. This is not one prescription for all sectors, but a way to keep authority proportional to effect.

ICO governance guidance similarly requires identifying who determines the purposes and means of data processing and who processes for another; legal roles may vary by stage and relationship. A contract calling a vendor a “processor” cannot settle every responsibility when

actual control and legal competence differ. This is data-protection guidance within scope, not a comprehensive ruling on civil or banking liability for all AI outputs.15

The decisive test is not whose finger touches approval, but who can say “this rule is unacceptable” and cause an effective institutional result. Without that capacity, many signatures will not save a decision system. At worst, signing becomes a ceremony for allocating blame after authority has already been distributed.

Before the Next Result

England reissued that summer’s grades on a different basis. Students could not be returned to examination halls on August 13, and recalculation could not restore days lived under the first result. Reversal was important and limited. A decision is not merely a database number. It enters people’s lives at a particular hour, then may keep moving after amendment.

Ofqual’s deeper lesson is not the easy warning of an “unfair algorithm.” It is that a judgment can be clear on paper and obscure on the authority map. As models write recommendations, identify risks, and dispatch decisions, fewer workflows will pause for an explicit human view. A board chair should ask not “which model decided?” but “which decision of mine allowed this model to create this effect, and who can stop it if we were wrong?”

The grade printed in August 2020 carried few letters. Behind each stood a long chain of decisions. A modern institution that learns from it will not merely find a signature at the chain’s end. It will look for who held the pen at its beginning.

References

1. Awarding GCSE, AS & A levels in summer 2020: interim report, **Ofqual**, August 13, 2020; primary interim regulatory report describing the mechanism, context, and limits in England.
2. Requirements for the calculation of results in summer 2020, **Ofqual**, issued July 7 and amended through August 20, 2020; primary requirements whose page includes versions after policy changed.
3. Grading gaps in summer 2020, **Ofqual**; later official analysis of 457,420 subject entries from 246,110 students in a defined age and geographic scope, not a measure of every individual estimate’s truth or the whole UK.
4. Written statement from Chair of Ofqual to the Education Select Committee, **Roger Taylor, Ofqual**, September 2, 2020; primary statement identifying ministerial direction and regulatory responsibility.
5. Separation of Ownership and Control, **Eugene F. Fama and Michael C. Jensen**, *Journal of Law and Economics*, June 1, 1983; foundational academic distinction between decision management and control, not an AI experiment.
6. Formal and Real Authority in Organizations, **Philippe Aghion and Jean Tirole**, *Journal of Political Economy*, 1997; theoretical study of formal authority and real influence under distributed knowledge.

7. Ensuring statistical models command public confidence, Office for Statistics Regulation, March 2, 2021; relatively independent official review of 2020 grading lessons across UK countries, not a judicial ruling.
8. 12 CFR §1002.9: Notifications, Consumer Financial Protection Bureau, current US regulation checked October 31, 2026; primary text on adverse-action reasons within its scope.
9. Consumer Financial Protection Circular 2022-03, May/June 2022, historical interpretive statement; the agency's Withdrawn Guidance records its and Circular 2023-03's withdrawal on May 12, 2025. They are not presented as current guidance.
10. AI Risk Management Framework 1.0: Core, GOVERN, NIST, 2023; voluntary framework for leadership roles and model-risk management.
11. Article 14: Human Oversight, European Commission, legislative text in the consolidated July 27, 2026 version; see AI Act: Regulatory framework for the 2026 official timetable distinguishing enactment from phased application.
12. Key Principles of Governance in Financial Institutions Subject to Saudi Central Bank's Oversight and Supervision, Saudi Central Bank, June 30, 2021, second edition May 8, 2024, listed as in force; primary financial-governance source, not a rule specifically for signing AI decisions.
13. Regulating AI in the financial sector: recent developments and main challenges, Financial Stability Institute/BIS, December 12, 2024; comparative institutional study, not a measurement of particular Saudi banks.
14. Explaining decisions made with AI, ICO and Alan Turing Institute via GOV.UK, January 27, 2025; practical explanation guidance, not a field experiment.
15. What are the accountability and governance implications of AI?, ICO, official guidance available at review; explains controller and processor roles in data protection, not every form of legal liability.