

ALGORITHMIC POWER LAB

Field Guide: A print-friendly guide to the Decision Rule Auditor, algorithmic governance, accountability, contestability, and institutional responsibility.

Fictional educational simulation. No real decisions or predictions. References listed below are bibliographic leads awaiting independent URL verification.

The rule is a political choice

An automated decision system is not simply a calculator. Someone defines the goal, chooses the data, decides which errors matter, and establishes who can appeal. These choices distribute attention, opportunity, and burden. Algorithmic governance refers to the use of computational systems in processes that shape decisions, institutions, and public or private power. Accountability asks who can explain, contest, change, and remedy those decisions. The question is not whether computers are inherently fair or unfair. It is whether the surrounding institution can justify its decisions to the people affected.

What the auditor actually models

This exhibit is a fictional allocation exercise, not a real credit, employment, policing, housing, or benefits system. Six imaginary applicants have an experience score, a need score, and an access flag. You choose weights for experience and need, decide whether to penalize an access barrier, and set the number of available places. The auditor ranks candidates with a transparent linear score, then allocates the available places. Because the arithmetic is visible, you can test whether a seemingly small rule change changes the people selected. No generated outcome establishes what should happen in an actual institution.

A decision pipeline has many owners

A useful audit starts before the model. Who defined the problem? Who decided which population would be measured? Who assembled the records? Which categories are proxies for other attributes? Who approved the thresholds? Who monitors drift and complaints? Who has authority to stop the system? A vendor can build a technically accurate tool while an agency makes an unjust policy choice. Conversely, a sensible policy can be undermined by incomplete data, rushed deployment, or inaccessible appeal processes. Responsibility must be mapped across procurement, design, operations, and governance.

Accuracy is not the whole story

Accuracy describes how often a prediction matches a chosen target, given a chosen dataset and definition of correctness. It does not tell us whether the target is legitimate. A model predicting past hiring decisions might reproduce historical exclusion rather than identify future potential. A model predicting which applications were previously approved could simply learn the habits of earlier reviewers. Even a high aggregate accuracy can conceal worse errors for smaller groups. Evaluation therefore needs both technical metrics and scrutiny of the social question being answered.

The threshold is policy, not nature

A threshold converts a continuous score into a consequential yes-or-no result. Moving it can change the number of people accepted and the mix of errors. A fixed number of places, as used in this fictional lab, produces competition: one person entering can push another out. That makes capacity a separate policy variable. Increasing capacity can change outcomes without changing anybody's measured qualities. Institutions should disclose whether a threshold reflects evidence, resource limits, a legal requirement, or administrative convenience.

Proxy variables and missingness

A proxy is a measured feature that stands in for something else, sometimes unintentionally. Postcode, school attended, device type, and gaps in records may correlate with opportunity or structural barriers. Removing a protected characteristic does not guarantee that a system stops reproducing related disparities. Missing records can also be systematically uneven. If some people have less documentation because services were harder to access, treating missingness as a negative signal can compound that barrier. The lab's access flag illustrates this design hazard, but it does not model real population distributions.

Fairness is plural

Different fairness definitions can conflict. One view prioritizes consistent treatment of individuals with similar relevant evidence. Another emphasizes group-level outcomes or error rates. A third asks whether the underlying objective and allocation of resources are defensible at all. These are not interchangeable. Equal acceptance rates may conceal differences in need; identical score thresholds may preserve earlier inequalities. The right approach depends on the domain, law, evidence, institutional purpose, and people affected. A dashboard cannot resolve the normative dispute on its own.

Explanations versus meaningful recourse

Showing a score formula is helpful but insufficient. A person also needs to know what information was used, whether it is accurate, which policy was applied, and how to request correction or review. An explanation may be mathematically faithful yet practically useless if the person cannot change a mistaken record or contact a responsible official. Recourse requires deadlines, accessible channels, human authority, documentation, and the possibility of an effective remedy. Contestability is a process, not a tooltip.

Two documented cases: lessons and limits

The Netherlands' childcare benefits scandal involved wrongful accusations and harsh recovery demands affecting many families. Parliamentary investigations and subsequent public debate highlighted institutional failures involving risk selection, enforcement, discrimination, and weak safeguards. It should not be reduced to the claim that one algorithm alone caused the harm; administrative decisions and policy design mattered. In a different domain, the UK's 2020 exam grading controversy showed how statistical standardization could produce consequential and contested outcomes when normal examinations were cancelled. Public pressure led to a reversal toward teacher-assessed grades. Neither case is replicated by the toy scoring exercise here. They demonstrate why oversight and remedies matter when automated or statistical rules meet real lives.

The case for and against automation

Supporters argue that carefully designed decision tools can improve consistency, document criteria, surface patterns, and help overloaded institutions allocate scarce attention. Some automated rules may be easier to inspect than undocumented discretionary judgments. Critics counter that scale magnifies errors, hidden objectives can become entrenched, and formal consistency can give weak policies a false appearance of neutrality. Both concerns can be valid. The relevant comparison is not an ideal machine against a flawed human, or an ideal human against a flawed machine. It is the actual alternative process, its evidence, safeguards, costs, and error distribution.

Auditing without false certainty

An audit can inspect documentation, sample records, test changes to rules, interview affected communities, and review appeal outcomes. But data access may be limited; small groups may be difficult to evaluate; historical labels may be contaminated; and causal claims need stronger designs than descriptive comparisons. Simulation helps identify questions, not answer all of them. In this exhibit the scores are illustrative and unvalidated. You cannot infer real-world fairness, discrimination, safety, or legal compliance from the displayed selections. A responsible audit records what remains unknown.

Procurement, monitoring and exit

Governance should be designed into contracts and workflows. Institutions need a clear owner, purpose limitations, record retention rules, incident reporting, independent review where appropriate, and a route to suspend a system. They should specify what evidence would trigger a rollback and how affected people would be notified. Monitoring cannot mean merely checking whether a server is online. It must examine error reports, shifts in population, changes in input quality, and whether remedies actually work. Sometimes the best accountable decision is not to automate.

A guided reflection

First run the default rule and note the selected candidates. Increase the weight on need while holding capacity fixed. Identify whose position changes and which assumption you changed. Next remove the access penalty, if enabled, and compare. Finally increase capacity. Ask three questions: Which choices are arithmetic, which are policy, and which require evidence not present here? Write a one-paragraph public explanation of your chosen rule. Then write an appeal path for someone who believes their data are wrong. If you cannot articulate the reason and recourse, the rule is not ready for consequential use.

What this does not prove

This toy exercise does not validate any predictive model, establish causal discrimination, estimate population-level impacts, or demonstrate legal compliance. Its candidates are invented; its attributes and weights are not calibrated to a real setting. A change in selection is evidence only about the fictional arithmetic under the assumptions shown. It is not a diagnosis of people or institutions. Treat every result as a prompt for inquiry, not as a verdict.

Print exercise: audit the rule

Score = (experience weight × experience) + (need weight × need) – (access penalty × barrier). Default weights: 5, 5, 0; places: 3. Fictional applicants: Ada (8,3,0), Bina (5,8,1), Ciro (7,5,0), Dev (4,9,1), Ema (9,2,0), Farah (6,7,1). Calculate scores, rank descending, break ties alphabetically. Repeat with need weight 8, then access penalty 3. Record which people change positions and explain which policy assumption moved.

Reflection: Who chose the objective? What evidence would justify each input? Who can appeal? What would trigger suspension? Which fairness definition applies, and who gets to decide?

Frequently asked questions

Is an algorithm automatically objective?

No. A reproducible calculation can still implement a contestable goal, imperfect measurements, or an arbitrary threshold. Objectivity requires scrutiny of the task and evidence, not just repeatable arithmetic.

Why show fictional applicants?

Fictional records let visitors inspect a decision rule without using personal data or implying that a real person can be reduced to a score. They are deliberately unsuitable for real decisions.

Does removing a sensitive variable eliminate bias?

Not necessarily. Other variables may act as proxies, and past outcomes may encode unequal access. Evaluation must consider data generation and actual effects.

Is a human reviewer always safer?

No. Humans can be inconsistent or biased too. A meaningful review requires training, authority, reasons, documentation, and an ability to reverse mistakes.

Can different fairness goals conflict?

Yes. For example, equal selection rates and equal thresholds answer different questions. The appropriate objective depends on purpose, context, and legitimate constraints.

What would make a real audit credible?

Clear scope, reliable and representative evidence, stakeholder participation, documented limitations, independent scrutiny where warranted, and actionable remediation pathways.

Further reading — bibliographic leads; verification pending

Cathy O'Neil: Weapons of Math Destruction (2016 — Crown)

Virginia Eubanks: Automating Inequality (2018 — St. Martin's Press)

Safiya Umoja Noble: Algorithms of Oppression (2018 — NYU Press)

Solon Barocas, Moritz Hardt, Arvind Narayanan: Fairness and Machine Learning (2023 — MIT Press)

NIST: AI Risk Management Framework (AI RMF 1.0) (2023 — NIST AI 100-1)

OECD: OECD Principles on Artificial Intelligence (2019; updated 2024)

European Union: Regulation (EU) 2024/1689 (AI Act) (2024 — Official Journal)

UK House of Commons Education Committee: Getting the grades they've earned: Covid-19 and the 2020 exams fiasco (2020 — parliamentary report)

Dutch Parliamentary Inquiry Committee: Ongekend onrecht (Unprecedented Injustice) (2020 — parliamentary report)

Danielle Citron: Technological Due Process (2008 — Washington University Law Review)

Sahilverse · sahiltayal.com · Write to Sahil: sahil@sahiltayal.com