

A correct calculation can produce a harmful decision

Whose definition of success could differ from the platform's?

AI conceptual illustration · fictional scene, not case evidence



Replacing and supporting human judgement

REPLACE A DECISION

Rule produces final outcome

Effective review needs useful evidence, time and authority.

Automation can apply a rule consistently at scale

SUPPORT A REVIEWER

Rule provides information;
reviewer checks context and
can overturn

Bias can enter at several points

Who is represented?

What proxy or label?

Which objective?

Whose results differ?

Bias can enter at several points. Compare effects for relevant groups instead of relying only on a whole-dataset average.

Data may omit or misrepresent a group

Missing examples can change who succeeds

Examples · teal: Invented training examples · common scale 0–90



A training set contains 90 examples from Group A and 10 from B

A proxy can misrepresent the intended quality

TARGET QUALITY

Reliability for a media event

A student with no earlier opportunity cannot build the same count. Investigate access before inferring ability.

Desired quality: reliable access to a task

RECORDED PROXY

Number of earlier events
completed

A feedback loop can reinforce an early advantage

1. Exposure opportunity →

2. Recorded interaction →

3. Updated input/score →

4. Next exposure

The output affects later input. An early visibility advantage can be reinforced; an unmeasured group may remain hidden.

Exposure gives an item opportunities for interaction

Human judgement can erode through over-reliance

UNCRITICAL ACCEPTANCE

Score → accept → repeat

INDEPENDENT CHECKING

Score → inspect input → give
reason → allow challenge

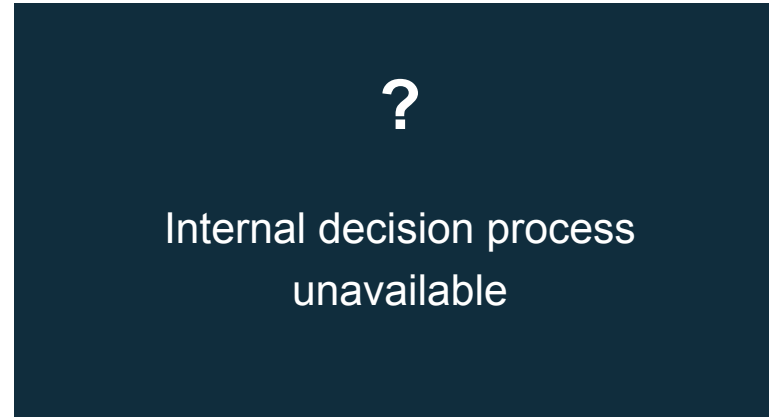
Numbers can appear authoritative. A challenge route is a safeguard to evaluate, not a guarantee.

A score may seem authoritative because it is numerical

A black box limits explanation

VISIBLE INPUT

History
Duration
Language



VISIBLE OUTPUT

Ranked
recommendations

Seeing the result does not explain the internal reasoning.

Inputs and outputs may be visible

Transparency makes parts of the decision understandable

Rule is disclosed

Individual result
explained

Limitations made clear

Visibility can help understanding; it does not by itself assign responsibility or provide remedy.

Explain the objective and significant inputs

Accountability assigns responsibility and consequences

1. Responsible actor →

2. Challenge route →

3. Review with authority →

4. Repair or consequence

A route must be usable and capable of affecting a decision; merely adding a human name is insufficient.

Name who must explain the decision

A case claim needs a boundary

2019

NIST · independent
technical evaluation

Source paraphrase · read the cited original

SUPPORTS THIS CLAIM

Demographic performance differences in tested face-recognition algorithms.

DOES NOT ESTABLISH

Not a test of every current system; not proof of one cause or of criminal intent.

Technical failure, causal explanation and real-world harm are different claims.

Real case A: a provider explains recommendations

2020

TikTok · provider account

Source paraphrase · read the cited original

SUPPORTS THIS CLAIM

Signals and a broad account of recommendation logic.

DOES NOT ESTABLISH

Current exact weights, independent wellbeing outcomes or creator earnings.

TikTok describes interaction and video-information signals

Real case C: a regulator examines design risks

6 FEB 2026

European Commission ·
preliminary finding

Source paraphrase · read the cited original

SUPPORTS THIS CLAIM

A regulator's concern about addictive design and safeguards.

DOES NOT ESTABLISH

A final legal determination or the exact ranking rule.

The European Commission issued preliminary findings in February 2026

Observation, explanation and intention

1. Recorded observation →

2. Proposed mechanism →

3. Contextual consequence →

4. Qualified judgement

State which links the source establishes and which are your inferences. Give a condition that could change the judgement.

Observation: a measured pattern in the results

A conditional judgement weighs an intervention

PROPOSED INTERVENTION

Give new items an initial exposure opportunity

TEST THE TRADE-OFF

New creators: chance to be measured

Viewers: relevant and varied choices?

Platform: objective and appeal route?

CONDITIONAL JUDGEMENT

Recommend only if the evidence supports the criterion and the safeguard works.

Criterion: fair access to visibility

Paper 1 and Paper 2 use the same knowledge differently

PAPER 1-STYLE PRACTICE

Explain a digital mechanism;
analyse people and context;
reach a supported judgement.

PAPER 2-STYLE PRACTICE

Use stimulus evidence;
compare sources where
required; integrate knowledge
into a qualified synthesis.

These are short formative rehearsals, not complete examination papers or predicted grades.

Paper 1: apply knowledge to a specified situation

Chapter 3.2: delayed explanation check

PAUSE · INDIVIDUAL FIRST ATTEMPT · 2 MINUTES

What makes an evaluative conclusion stronger?

Predict / retrieve

Justify with an operation

Check → correct

Save your first response below. Explain the correction on a fresh example; practice is not a mark.

Explain one mechanism without notes