# Hyroo — AI hiring governance

How automated candidate scoring is controlled, recorded and audited

Our AI scores applicants to help a recruiter decide reading order. It never advances, rejects, hires or messages anybody. Every score it produces — including the ones produced with no AI model at all — is written to a decision log that records which model and which version of the scoring instructions made it, and every time a human decided against it.

## The tool is advisory, and cannot act

The score changes what a recruiter reads first. It changes nothing about a candidate's status.

**No automated action, in any configuration** — No stage change, rejection, hire or message is ever driven by a score. Every state transition is performed by a named person. There is no setting, plan or switch that makes the tool advance or reject anyone.

**A score is a similarity judgement, and is described as one** — It compares a résumé against a job description. It is not validated as a predictor of job performance, tenure or retention, and we make no such claim.

**Disagreement with the tool is recorded, not discouraged** — Advancing someone the tool did not recommend, rejecting someone it did, hiring against it, or an interview verdict contradicting it — each is captured with the actor and the before/after state, at every decision point in the pipeline.

## What the model is never given

The controls are structural: most of what could bias a score does not exist on the record to leak.

**Not the candidate's name** — A well-known proxy for gender, ethnicity and national origin. The scoring prompt renders the applicant's identity as withheld. The name exists on the record — it is simply never put in front of the model. Stated precisely, because the distinction matters: the control is on the prompt, not on the text. A name written inside the résumé itself can still reach the model, and is mitigated by instructing the model to ignore it rather than prevented structurally.

**Not a protected attribute — there is none to give it** — There is no gender, race, age, date-of-birth, religion, disability, veteran, marital or photo field on the candidate record. Where a workspace chooses to collect voluntary self-identification for fairness testing, it is stored in a separate table that the scorer has no path to reach, and it is never shown to anyone assessing the candidate.

**Not contact details** — Email and phone are not part of the scoring input.

**And the prompt says so out loud** — Every scoring prompt instructs the model to judge job-relevant merit only and to ignore any protected signal that leaks through free text. This mitigates indirect proxies in a résumé — a university, a country, a career gap — it does not eliminate them, which is exactly why the statistical testing below exists.

## What is recorded about every decision

The log is not a feature you buy; it is written for every workspace, always.

**One record per automated decision, written once and never rewritten** — Which candidate, which job, which model, which prompt version, the score, the band, the merit signals the model actually saw, and its rationale with contact details and names stripped out. No part of the product updates a decision once it is written; the only path that touches one again is erasure, which deletes it with the candidate.

**Including the decisions made with no AI at all** — When the AI is switched off, the spend cap is reached or the model call fails, a deterministic fallback produces the score. That is still an automated employment decision, so it is logged as one — recorded as having had no model and no prompt.

**No off switch, no plan tier, no flag** — An audit trail with gaps in it cannot answer the only question that matters: how was this person scored. So there is nothing to turn off.

**The prompt version is a fingerprint of the prompt's wording** — Not a number somebody remembers to increment. Change one word and past decisions stay attributable to the wording that produced them. That is attribution, not replay: the filled prompt and the model's answer are not kept, so a past score can be traced to its wording but not re-run.

**It commits with the score** — The explanation is written in the same database transaction as the score, so a score can never exist without one.

## What the fairness audit measures

Two halves, because only one of them works without asking candidates for demographic data.

**Process fairness — available to every workspace, on day one** — The distribution of scores, which model and prompt version produced them, the share produced with no model, and the rate at which humans decided against the tool. No demographic data is required for any of it. A near-zero override rate is itself a finding: it suggests the humans are rubber-stamping.

# Hyroo — AI hiring governance

How automated candidate scoring is controlled, recorded and audited

**The report itself is a module a workspace turns on** — The decision and override logs are written for everybody, always. The report that reads them is behind a flag that is off by default, and the statistical half needs data that has to be collected separately. A workspace that has enabled neither still has the log; it does not yet have the report.

**Statistical adverse impact — opt-in, and off by default** — The four-fifths test compares selection rates between demographic groups, and we hold no demographic data. Inferring it from names or photos would be both unlawful and the exact bias the name-blind scoring prevents. So a workspace may choose to invite voluntary, consented self-identification — a decision to take with legal advice, and one some jurisdictions constrain.

**Small groups publish nothing** — A group of fewer than 5 people has its counts, rates and ratio removed entirely, and is excluded from comparison. A ratio computed on three people swings on a single decision and can re-identify individuals; a number there is worse than no number, because it reads as evidence.

**And the surrounding totals are coarsened too** — When any group is withheld, the headcounts beside it are withheld and coverage is banded, because an exact total published next to the reported groups lets a reader recover the withheld one by subtraction.

**Every ratio is read against its coverage** — Self-identification is voluntary, so each table states what share of the scored population it was computed over, and low coverage is flagged on the report page. Where there was no scored population at all, that share is withheld rather than published as 0%, because a zero there would read as "we asked and nobody answered". A ratio over a small fraction of applicants is not evidence, and presenting one without its coverage is the easiest way to mislead with this report.

## What a candidate can ask for

The rights the log exists to make answerable.

**An explanation of a decision** — Every automated decision made about them, the model and prompt version behind it, the merit signals it used, and the human decisions that followed.

**Erasure that reaches the audit trail** — Decisions, overrides and any self-identification are erased with the candidate. An audit record's instinct is to survive; here it is deliberately overridden, because every one of those rows is about the data subject.

**An alternative process** — Where the automated-tool notice is shown at application time, it tells candidates they may request an alternative selection process or an accommodation.

## Where this maps

What the product produces against each obligation — not a claim that the obligation is discharged.

| Framework | Obligation | What we provide |
|---|---|---|
| EU AI Act | Art. 12 — automatically generated logs | One record per automated scoring decision, for every workspace, ungated |
| EU AI Act | Art. 13 — transparency to deployers | A published model card and fairness methodology, plus the model/prompt provenance on screen |
| EU AI Act | Art. 14 — human oversight | The override log and the published override rate; the tool takes no action of its own |
| EU AI Act | Art. 15 — accuracy and robustness | A deterministic fallback on every failure path, with its share published rather than averaged away |
| EU AI Act | Art. 10 — data governance | No protected attributes in the scoring input; self-identification isolated and consented |
| NYC Local Law 144 | Annual bias audit by an independent auditor | The exported workbook is the input an auditor works from — selection rates and impact ratios by category, with suppression applied. Engaging the independent auditor stays with the employer |
| NYC Local Law 144 | Published summary of results | The workbook's summary carries the scope, window, selection rule, threshold and the finding |
| NYC Local Law 144 | Candidate notice before use | An automated-tool notice at application time, enumerating what the tool is given. The lead time and the choice of jurisdiction-appropriate wording stay with the employer |

# Hyroo — AI hiring governance

Generated 2026-09-18 00:00Z

How automated candidate scoring is controlled, recorded and audited

## What we do not claim

Every statement above is bounded by these. The four-fifths threshold is 0.80 and the minimum reportable group size is 5.

- We do not certify the tool as unbiased, and the report is built so that it cannot say so. The strongest finding it can publish is: "No reportable group fell below the four-fifths (80%) threshold in this window." The absence of a signal is not proof of the absence of bias.
- We are not your independent bias auditor, and this does not replace one. NYC Local Law 144 requires an independent audit; we produce the evidence it is performed on.
- Showing a candidate notice at application time does not by itself satisfy a notice-period requirement, and the wording we ship is written for one jurisdiction. Timing and jurisdiction are the employer's duty.
- We do not claim the score predicts job performance, tenure, culture fit or salary. It has never been validated against on-the-job outcomes.
- We do not claim to eliminate bias. Name-blind scoring and the fairness directive reduce it; a résumé is free text and can still carry indirect proxies we cannot fully neutralise. That residual risk is the reason the statistical testing exists.
- This governance layer measures the scoring; it does not change it. The same prompts, the same model and the same scores as before it was added.
- Adverse-impact testing does not work out of the box. Without voluntary self-identification no four-fifths ratio can be computed at all — and the report says exactly that, rather than showing a clean result.
- Per-job adverse-impact testing usually returns "insufficient data". Most individual requisitions never reach the minimum reportable group size. That is honest, not a defect.
- Self-identification is voluntary, so respondents may not be representative of the applicant pool. Read every ratio against its coverage.
- A model or prompt change alters scores. Recording the model and prompt version on every decision makes that visible in the audit; it does not prevent it.
- Not every AI ranking in the product is an automated employment decision. Two are deliberately excluded from the decision log: copying a score that was already recorded, and the display ordering of external sourcing prospects, who have no candidate record for a decision to be about.

**This is evidence you can hand to an auditor or a regulator. It is not the audit, and it does not make anyone compliant: the independence of a bias auditor, the timing of candidate notices, and the decision to collect demographic data at all remain the employer's, under their own legal advice.**

Prepared by JStar Technologies Pty Ltd. The methodology, the limits and the residual risks summarised here are published in full in our model card and fairness methodology, which we will share on request.