Hiring-AI Risk Classification Standard

Consistent method for classifying employment AI as prohibited, high, moderate, or low risk and assigning mandatory controls.

SIMULATED OPERATING DOCUMENT

Prepared for a fictional employer using synthetic systems and synthetic applicant data. This document is a governance demonstration, not legal advice.

1. Classification rule

Every system is classified before purchase, pilot, feature activation, or production use. Classification uses (a) mandatory policy overrides and (b) a weighted risk score. The higher outcome governs. A material change requires reclassification.

2. Mandatory policy overrides

Override	Tier / response
Emotion, affect, honesty, personality, mental-state, or protected-trait inference from voice, face, video, biometric, or behavioral signals	PROHIBITED under HarborPoint policy; do not deploy.
Opaque fully automated rejection or other consequential decision without meaningful human review and candidate recourse	PROHIBITED unless a specific lawful use is approved through Charter amendment.
Intentional inference of race, ethnicity, disability, religion, age, sex, pregnancy, genetic information, or other protected trait for selection	PROHIBITED.
AI score/rank/assessment substantially influences hiring, promotion, termination, compensation, or another consequential employment decision	At least HIGH.
Covered NYC AEDT use	At least HIGH and Local Law 144 control overlay.

3. Weighted scoring factors

Factor	Weight	Scoring guidance
Employment-decision impact	3	1 = administrative; 4 = materially affects screening/selection
Automation level	2	1 = drafting; 4 = automatic consequential action
Protected-class outcome risk	3	1 = no plausible differential outcome; 4 = known/likely differential impact
Disability/accessibility risk	3	1 = no candidate interaction; 4 = assessment can screen out or is inaccessible
Sensitive data exposure	2	1 = no personal data; 4 = sensitive/biometric/health-related
Explainability	1	1 = clear deterministic rule; 4 = opaque model output
Vendor transparency	1	1 = full evidence and audit access; 4 = limited documentation
Scale	1	1 = <100 people/year; 4 = >10,000/year

Factor	Weight	Scoring guidance
Jurisdictional complexity	2	1 = no special overlay identified; 4 = multiple AI/employment notice/audit regimes
Recourse availability	2	1 = clear human review/alternative; 4 = no meaningful challenge route

Weighted score = sum(factor score x weight). Minimum = 20; maximum = 80. This is an internal triage score, not a legal classification.

4. Tier thresholds and controls

Tier	Score / trigger	Typical examples	Required governance response
PROHIBITED	Policy override	Emotion/personality inference; protected-trait inference; opaque automated rejection	Do not deploy; disable; investigate data exposure; close POC; remediate.
HIGH	>=50 or consequential-decision override	Résumé ranking, candidate scoring, shortlisting, job matching used to advance/reject	Formal AIA; dataset/feature review; validation/adverse-impact testing; Legal/HR approval; human review; notice/recourse; vendor controls; subgroup monitoring.
MODERATE	32-49	Recruiting chatbot; interview transcription/summarization; JD drafting with review	Privacy/security review; use limits; disclosure where appropriate; human verification; periodic review.
LOW	20-31	Scheduling; internal drafting with no candidate data	Approved-tool controls; staff training; baseline monitoring.

5. Reclassification triggers

- New scoring, ranking, recommendation, inference, or auto-reject feature.
- New candidate data source or feature set.
- Material model or threshold update.
- Deployment to new job family, geography, or candidate population.
- Complaint trend, adverse-impact signal, accessibility issue, or override spike.
- Change in vendor ownership, subprocessors, data use, retention, or security posture.
- Legal or regulatory change affecting the use case.