ASSESSING AI RISK IN HIRING & EMPLOYMENT

A Four-Factor Risk Framework

Risk in employment AI depends on many factors: the data a system draws on, how much independent judgment it exercises, how directly its output drives a decision, and the stakes of that decision. The same AI system may present very different risks depending on how it is configured and deployed, making a simple high-versus-low-risk classification insufficient. A tool used in a high-stakes process, such as an assistant that schedules interviews, is not necessarily high risk. Meanwhile, even tools considered low risk can shape important decisions down the line. Our four-factor framework helps organizations assess an AI system's overall risk profile and scale governance measures to match.



No single factor tells the whole story. Weighing all four together — and scaling safeguards to match — helps organizations build workplace AI that's fair, transparent, and worthy of trust.

FACTOR 1

Quality and Sensitivity of Data and Inferences (Input & Output)



CORE QUESTIONS

What data is used for training a model? What inputs does the system use, and what outputs or inferences does it produce?

KEY CONSIDERATIONS

Sources
Relevance and Accuracy
Inferences
Individual Expectations

LOWER RISK EXAMPLE

A generative AI system reviews employee writing and suggests improvements to grammar or persuasiveness.

HIGHER RISK EXAMPLE

A generative AI system analyzes employee writing to infer confidence, anxiety, or enthusiasm, flagging employees as "low engagement" or "high stress" for supervisors (potentially unexpected inferences affecting protected classes).

FACTOR 2

Degree of Autonomy, Discretion, and Action (Decisioning)



CORE QUESTION

To what extent does the AI system exercise independent judgment or discretion in reaching its output?

KEY CONSIDERATIONS

Decision Logic and Latitude
Transparency
Scope of Action
Output Constraints

LOWER RISK EXAMPLE

Automated candidate screening tools based on fixed rules, such as "Do you have a valid driver's license?"

HIGHER RISK EXAMPLE

Generative AI system holistically evaluating and summarizing a candidate's subjective fit ("Is this person likely to be a good driver?")

FACTOR 3

Proximity to the Decision (Output / How Output Is Used)



CORE QUESTION

How close is the system's output to the final decision?

KEY CONSIDERATIONS

Purpose of Use
Influence of Output
Human Review
Reversibility

LOWER RISK EXAMPLE

AI chatbot that answers questions about internal HR policies, such as vacation days or benefits.

HIGHER RISK EXAMPLE

AI chatbot with agentic capabilities responds to, and independently approves, requests for time off or other internal HR benefits.

FACTOR 4

Nature and Significance of Impact (Consequentiality)



CORE QUESTION

What kind of decision does the system inform, and how significant are the consequences for the individual?

KEY CONSIDERATIONS

Type of Decision
Material Impact
Severity and Reversibility
Breadth of Impact

LOWER RISK EXAMPLE

System that reviews past performance reviews and recommends to employees opportunities for professional development training.

HIGHER RISK EXAMPLE

System that reviews past performance reviews and recommends to supervisors candidates for termination.