EMPOWERING WOMEN
FOR DIGITAL & REMOTE HR CAREERS

MODULE 3 · HR Data, AI Literacy, and Digital Systems · LESSON 3.6

# Interrogating AI outputs: bias, exclusion, and accountability

Learn to question AI outputs against the original HR requirement, identify possible bias and exclusion, and understand who remains accountable for decisions supported by AI.

1.5 training hours · Phase 1 · RemoteHeR



WHAT THIS LESSON COVERS

# From AI output to accountability

- **Evaluate the output against the requirement:** Check whether an AI result actually answers the task it was given.
- **Understand how bias enters:** Explore training data, proxy variables, and filter design.
- **Recognise disparate impact:** Identify patterns of unequal outcomes in an AI-supported shortlist.
- **Ask what candidates need to know:** Understand basic transparency requirements around automated decisions.
- **Keep accountability human:** Identify who remains responsible when a person approves an AI-supported recommendation.



# An AI output can sound convincing and still fail the original requirement.

The first question is not: “Does this sound good?”, it is: “Does this meet the requirement we gave the system?”

Input-based expectation management

# Start with the requirements

Before reviewing the AI output, return to the original task. Check:

- What was the AI asked to do?
- What constraints did the task include?
- What output is expected based on a set input?



# Compare Input and Output

## Input: Original Requirement

- Required qualification
- Relevant experience
- Maximum length
- Inclusive language
- Evidence from supplied data

## Output: AI-Product

- Missing
- Included
- Exceeded
- Some problematic wording
- Includes unsupported claim



# “Sounds good” is not a quality test as fluent language can hide a weak result.

An AI output may be: clear but incomplete, detailed but irrelevant, confident but unsupported, or polished but inconsistent with the original requirement.

Good writing ≠ good evidence ≠ good HR decision

A four-question output check

# Use four questions before accepting an AI output.

1. **Requirement**

   Does the output match the original task and criteria?

2. **Evidence**

   What information actually supports the output and are we seeing relevant evidence, or are assumptions being made from indirect signals?

3. **Exclusion**

   Who or what might be left out because of the way the system interprets or filters information?

4. **Impact**

   What happens if we use this output and could the recommendation create unequal outcomes or affect people differently?



**Bias is not necessarily created at the moment an AI makes a recommendation.**

It can enter through the data used to train or inform a model, variables that stand in for other characteristics, or the way filters and requirements are designed.

**A model can learn patterns from historical data, including problematic patterns.**

If previous decisions reflect unequal opportunities or selection patterns, a model trained on those data may reproduce or reinforce them. Ask “What patterns does the data teach the system to recognise?”





# A variable does not need to name a sensitive characteristic to be related to it.

A variable such as location, career history, or another background characteristic may correlate with information the organisation did not intend the system to use. Ask: “What else might this variable be standing in for?”



# A filter can exclude people because of how the requirement is defined.

“At least five years of continuous experience in the same role” may exclude candidates whose experience is relevant but structured differently, wherefore we need to clarify whether the filter measures the requirement we actually care about.

Entry points for bias

# Look for bias in three places.

### Training data

What patterns did the system learn previously?

### Proxy variables

What might the variables indirectly represent?

### Filter design

Who could be excluded by the way the requirement was defined?



# A process can apply the same rule to everyone and still affect groups differently.

The important question is not only: “Did everyone go through the same filter?”, but also: “What happened to different groups after the filter was applied?”

# A difference in outcomes does not automatically tell you why it occurred.

It may indicate a difference in the input population, a filter that affects groups differently, patterns learned from historical data, or another factor that has not yet been examined. The next step is investigation, not assumption.



Test for a possible disparity

# Compare outcomes at the same stage of the process

1. Count applicants in each group.
2. Count how many pass the relevant filter or reach the shortlist.
3. Calculate the rate for each group.
4. Compare the rates.
5. Investigate substantial differences and possible explanations.



# One stage can hide what happened at another stage.

A difference may appear during screening, disappear later, or become larger at another stage. Check the stage where the difference enters:
Application → Screening → Shortlist → Interview → Hire

What should a candidate be told?

# Role of automated systems influencing a decision

### Selective usage

Automated processing or decision support is being used, where relevant.

### Role in the process

What role does the system play?

### Information type

What information was considered?

### Human involvement

Where did human review or decision-making remain involved?

Transparency is more than “AI was used”

# “AI was used” does not make the process understandable

### What happened?

Role of the system in the process

### Why does it matter?

Effects of AI on the process

### Who decided?

Human and AI involvement



# “The system decided” is not an adequate explanation of human responsibility.

A useful explanation describes the process in plain language and the explanation should reflect what actually happened: “The system was used to identify applications matching defined criteria. A recruiter then reviewed the resulting shortlist.”



# An AI recommendation does not transfer accountability to the system.

If: AI recommends → Person reviews → Person approves,
the person and organisation responsible for the decision remain accountable for how that decision was made.

Recommendation is not responsibility

# The more consequential the decision, the more important human accountability becomes.

## AI may:

- identify patterns
- rank candidates
- flag information
- suggest an option

## But people remain responsible for:

- checking the recommendation
- questioning problematic results
- considering relevant evidence
- making or approving the decision



# Pause before accepting an AI recommendation.

If the output conflicts with the requirements, information is missing, groups are affected differently, recommendations are unclear, you find unsupported assumptions, or the decision could have significant consequences for a person, pause before accepting an AI recommendation.

From AI output to accountable decision

# A responsible process has four stages.

1. **Check**
   Does the output meet the requirement?

2. **Question**
   Could bias, exclusion or missing information affect it?

3. **Explain**
   Can the role of the system and the decision process be described?

4. **Decide**
   Can a responsible person justify the final decision?

## Before using an AI-supported HR result, ask about the process as a whole.

Within the process, the original requirement, influential data or patterns, variables acting as a proxy, possibly excluding filters, group-based differences in outcomes, the system's role, and human accountability in the final decision should always be taken into consideration.

NOW IT IS YOUR TURN

# Mark up the AI shortlist

## Evaluate an AI shortlist against the requirement

Review a fictional AI shortlist, mark evidence, exclusions and concerns, compare selection rates, and identify what should be checked before approval.

**It feeds into**

AI critique: evidence, exclusion risks, selection-rate analysis and accountability checks

**Read by your tutor**

No mark, no score and no ranking. Your tutor reads your work and may give you feedback.

RECAP

# Before you move on

- Evaluate an AI output against the original requirement, not simply against how convincing it sounds.
- Bias can enter through training data, proxy variables, and filter design.
- Different outcomes between groups are a reason to investigate the process and its effects.
- Candidates need understandable information about the role of automated systems in decisions that affect them.
- Using AI does not remove human accountability for the decision.

Complete “Mark up the AI shortlist” and then take the knowledge check.

EMPOWERING WOMEN FOR DIGITAL & REMOTE HR CAREERS

# Lesson 3.6 complete

You evaluate an AI-generated shortlist against the original recruitment requirement and identify possible exclusion, bias and accountability issues.

**Erasmus+ 2025-1-PT01-KA220-VET-000365831**

Co-funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

CC BY-SA 4.0