

A4 MACHINE LEARNING · A4.4

Ethical considerations in AI

The implications of machine learning in society: **bias, privacy, transparency, accountability**, jobs, and the environment. Discuss both sides.

01 Key issues

- Bias** Models inherit bias from data.
- Privacy** Personal data, consent, surveillance.
- Transparency** Black box vs explainable decisions.
- Accountability** Who is responsible for mistakes.
- Jobs** Automation displaces and creates roles.
- Environment** Training uses lots of energy.

02 Responsible AI

- Fairness** Diverse data and fairness testing.
- Transparency** Explainable, documented models.
- Privacy** Consent and minimal data.
- Accountability** Clear responsibility for outcomes.
- Oversight** Humans in the loop for big decisions.
- Sustainability** Weigh the energy cost.

03 How bias enters a system

- | | | |
|------------------------|---|---------------|
| Biased data | Training data under-represents a group or reflects past prejudice. | CAUSE |
| Model learns it | The model fits the data and absorbs the bias along with the real pattern. | TRAIN |
| Unfair output | Predictions discriminate against the under-represented group. | EFFECT |
| Real harm | Consequences in hiring, lending, and criminal justice. | IMPACT |

04 Answering a "discuss"

Benefits Faster, consistent, cheaper, scalable.

Risks Bias, privacy, opacity, job loss.

Safeguards Pair each risk with a mitigation.

Conclusion A balanced judgement, not "good/bad".

Ground it Use a real application.

05 Worked example: AI CV shortlisting

Benefit Fast, consistent screening.

Risk Bias from past hiring data.

Risk No explanation for rejections.

Fix Audit data, explain, human reviewer.

06 Match each issue to a safeguard

Bias	Diverse, representative data; audit for skew; test fairness across groups.	FAIRNESS
Privacy	Informed consent, minimal data collection, and strong data protection.	CONSENT
Opacity	Explainable models and documentation so decisions can be understood.	TRANSPARENCY
Unclear blame	Clear accountability and human oversight, set before deployment.	ACCOUNTABILITY

FINAL PASS BEFORE THE EXAM

Rapid exam tips

Eight things that lose marks in Paper 1 if you slip on them. Skim before you walk in.

01

"Discuss" means weigh benefits and risks; never give just one side.

02

Name **specific issues** (bias, privacy, accountability), not "AI is bad".

03

Bias usually comes from the **data**, not the code.

04

Reduce bias with **diverse data**, fairness metrics, and human oversight.

05

Transparency lets a decision be explained, challenged, and corrected.

06

Set **accountability** before deployment, not after a failure.

07

Privacy needs informed consent and collecting only the data you need.

08

Note the **environmental** cost, and always ground the answer in a real application.