

Topic 7: Ethics and Ownership

Frameworks for reasoning about ethics, how bias creeps into software, and two case studies that changed the profession.

Optional stretch material. This goes beyond the 9618 syllabus, so you will not be asked about it directly in the exam. Its purpose is to deepen ideas you already know. **Builds on:** 7.1 Ethics and ownership. Try the challenge questions on the last pages before you read the answers.

Four lenses for an ethical question

High-scoring ethics answers do more than say "it's wrong". They apply a way of reasoning. Four common frameworks, applied to *"Should a school use facial recognition to take the register?"*:

Framework	Asks	Applied to the example
Utilitarian	Which choice gives the greatest overall benefit?	Saves teacher time and speeds up registration, but weigh against the harm of constant surveillance of children.
Duty-based (deontological)	Which rules or rights must never be broken?	Pupils have a right to privacy and consent; collecting biometric data may be wrong regardless of benefits.
Virtue ethics	What would a person of good character do?	Would a trustworthy school act openly and proportionately?
Rights / justice	Is everyone treated fairly?	Does the system work equally well for all skin tones and for pupils with disabilities?

The frameworks can disagree, and recognising that, then justifying which you weigh most heavily, is the mark of a mature answer.

Algorithmic bias

A machine-learning system learns from past data, so it can inherit and amplify the biases in that data. Reported examples include a recruiting tool that learned to downgrade CVs mentioning women's organisations because most past hires were men, and facial-recognition systems with higher error rates for darker skin tones.

- **Biased training data:** the sample under-represents some groups, or records past unfair decisions.
- **Proxy variables:** removing "ethnicity" does not help if postcode, name or school still predict it.
- **Feedback loops:** predictive policing sends officers to areas with more recorded crime, which records more crime there.
- **Mitigations:** representative data, testing performance per group, human oversight, and explanations for decisions.

Two case studies

Therac-25 (1980s). A radiation-therapy machine whose software replaced hardware safety interlocks. A race condition meant that fast operator input could leave the machine in an unsafe mode; patients received massive overdoses and several were killed or seriously injured. Lessons: never rely on software alone for safety-critical protection, test timing and concurrency, and take user error reports seriously.

Post Office Horizon (1999–2015). Accounting software showed phantom shortfalls in branch accounts. The Post Office treated the system as reliable, and hundreds of sub-postmasters were prosecuted (more than 900 cases in total, including those brought by other bodies); many convictions have since been overturned. Lessons: software contains bugs, evidence from computers must be challengeable, and organisations need accountability when a system's output ruins lives.

Challenge questions

1. A council proposes software that predicts which families might need social-care support. Evaluate the proposal using two different ethical frameworks.
2. A hiring model is trained without a "gender" column, but still rejects women more often. Explain how this can happen.
3. Choose one of the case studies and explain how a professional code of conduct (such as the ACM/BCS codes) could have led to a different outcome.

Worked answers

Only look once you have had a proper go at each question.

1. Sample answer. *Utilitarian*: early help could prevent serious harm to many children, which outweighs the cost, provided false alarms do not cause widespread distress. *Duty-based*: families have a right to privacy and not to be profiled or treated as suspects on the basis of a prediction; this may rule the system out regardless of benefit. Credit any valid framework applied specifically, with a conclusion that weighs them.
2. Other columns act as proxies for gender (for example hobbies, previous job titles, career gaps, school). The model learns patterns from historic hiring decisions that favoured men, so it reproduces the same pattern using these correlated features.
3. Sample answer (Therac-25). Codes require members to avoid harm, to be competent, and to report risks honestly. Engineers following them would not have removed hardware interlocks without rigorous safety analysis, would have investigated the overdose reports instead of dismissing them, and would have tested the software for timing faults before use on patients.