CASE 06

AI Fraud Detection

How Much Privacy Should We Trade for Security?

STUDENT HANDOUT — ENGLISH

REAL-WORLD CONTEXT

Banks increasingly use AI to detect fraud by analysing large amounts of customer data. Beyond normal transaction records, such systems may draw on device information, location patterns, online behaviour and other digital signals. These tools may improve security, but they also raise questions about privacy, proportionality, false suspicion and whether people should be able to challenge automated restrictions.

The Case

A Hong Kong bank introduces an AI fraud-risk system to identify suspicious customer activity.

The system analyses a wide range of information, including transaction history, spending patterns, GPS/location data, login and device information, online shopping behaviour, credit information and selected social-media or network signals. The bank argues that combining many data points allows the AI to detect fraud earlier than traditional methods.

Customers judged “high risk” may face delayed transfers, temporary account restrictions or additional checks before accessing funds. In some cases, these controls may be applied before any human investigation has taken place.

The bank says the system protects customers and the financial system from fraud. Critics argue that risk is not the same as guilt, that some data may be irrelevant or excessively intrusive, and that customers may not know why they have been flagged or how to challenge the decision.

The bank must decide how much data it should collect and how much decision-making authority should be given to the AI system.

CORE QUESTION

How much personal data should a bank be allowed to use for AI fraud detection, and when should human review be required before restricting a customer's access to money? Defend your judgment using evidence from the case and seriously consider a strong opposing view.

Choose a Learning / Assessment Task

- Class discussion: Which types of data seem necessary for fraud prevention, and which seem excessive? Explain why.
- Short reflection (250–350 words): Decide whether the bank's system is ethically acceptable. Give reasons, address one strong counterargument and state any safeguards you would require.
- Structured argument (500–700 words): Present the strongest case for broad AI fraud monitoring and the strongest case for privacy and due process, then defend your own judgment.
- Small-group debate: One group argues that security justifies broad data use; another argues that privacy and human review should sharply limit the system.
- Policy task: Design a bank policy covering data minimisation, risk thresholds, human review, explanation, appeal and temporary restrictions.

Before You Decide

Ask yourself: Is data available to a bank necessarily data it should use? Is a high-risk score evidence of wrongdoing or only a reason to investigate? When does fraud prevention become excessive surveillance? What values from your worldview shape your answer?

Adapted from the HKDSE Ethics and Religious Studies (ERS) Mock Exams | © ers999.com 2026. All rights reserved.

CASE 06

AI 詐騙偵測

為了安全，我們應該犧牲多少私隱？

學生講義 — 繁體中文

現實背景

銀行愈來愈多使用 AI 分析大量客戶資料以偵測詐騙。除了交易紀錄外，系統亦可能分析裝置資料、位置模式、網上行為及其他數碼訊號。這些工具可能提升安全，但亦帶來私隱、比例原則、錯誤懷疑，以及客戶能否挑戰自動限制等倫理問題。

個案

香港一間銀行引入 AI 詐騙風險系統，用來識別可疑的客戶活動。

系統分析多種資料，包括交易紀錄、消費模式、GPS / 位置資料、登入和裝置資料、網上購物行為、信貸資料，以及部分社交媒體或人際網絡訊號。銀行認為，把大量資料結合分析，可以比傳統方法更早偵測詐騙。

被系統評為「高風險」的客戶，可能面對轉帳延遲、暫時限制帳戶使用，或在取用資金前接受額外審查。有些限制甚至可能在真人調查之前便已自動執行。

銀行認為，這套系統可以保障客戶和金融體系免受詐騙。批評者則指出，「高風險」並不等於「有罪」；部分資料可能與詐騙無關，或過度侵犯私隱；而且客戶未必知道自己為何被標記，也未必有清晰渠道提出異議。

銀行因此必須決定：應收集多少個人資料，以及應給予 AI 多大的決策權。

核心問題

銀行為了利用 AI 偵測詐騙，應獲准使用多少個人資料？在限制客戶取用金錢之前，甚麼情況下必須由真人覆核？請運用個案資料為你的判斷提出理據，並認真回應一個有力的反方觀點。

選擇一項學習／評核任務

- 課堂討論：哪些資料對防止詐騙看來是必要的？哪些可能屬於過度收集？請解釋理由。
- 短文反思（250–350 字）：判斷銀行的制度是否可接受，提出理由、回應一個有力反方論點，並說明你要求的保障措施。
- 結構化論證（500–700 字）：分別提出支持廣泛 AI 監察及支持私隱與程序公義的最強理據，然後為自己的判斷辯護。
- 小組辯論：一組主張安全需要廣泛資料使用；另一組主張私隱與真人覆核應嚴格限制系統。
- 政策任務：為銀行設計政策，包括資料最少化、風險門檻、真人覆核、解釋權、上訴機制及暫時限制。

作出判斷前

想一想：銀行能取得的資料，是否就代表銀行應該使用？「高風險」分數是錯誤行為的證據，還是只代表需要進一步調查？防止詐騙在甚麼時候會變成過度監察？你的世界觀中哪些價值正在塑造你的答案？

Adapted from the HKDSE Ethics and Religious Studies (ERS) Mock Exams | © ers999.com 2026. All rights reserved.

CASE 06

AI Fraud Detection

FACILITATOR GUIDE — ENGLISH

WHY THIS MATTERS NOW

AI fraud systems illustrate a central problem of automated governance: a tool designed to protect people can also constrain them. The ethical issue is not simply privacy versus security. Students must ask whether the data used are necessary, whether risk scores are being treated as proof, whether restrictions are proportionate and whether meaningful human review and appeal exist.

Learning Objectives

- Distinguish fraud prevention from automatic suspicion or punishment.
- Identify the exact decision, stakeholders and ethically significant facts.
- Reason about privacy, security, proportionality, due process, fairness and responsibility.
- Evaluate whether data availability is the same as data necessity.
- Engage a serious counterargument and defend a final or conditional judgment.

Recurring Ethical Reasoning Capabilities

Define the decision → Identify stakeholders → Recognize competing values → Use case evidence → Examine outcomes → Consider rights / duties / fairness → Surface worldview assumptions → Present the strongest counterargument → Consider alternatives → Defend a reasoned judgment

Suggested 50–60 Minute Flow

Time	Activity	Facilitator Move
0–5 min	Initial judgment	Students choose: broad AI monitoring, limited AI monitoring, or monitoring only with strict safeguards.
5–15 min	Data sorting	Pairs classify case data as clearly relevant, possibly relevant or difficult to justify.
15–27 min	Risk vs guilt	Discuss why a risk score may justify investigation but not necessarily restriction or punishment.
27–38 min	Competing positions	Groups build the strongest argument for security and the strongest argument for privacy and due process.
38–47 min	Worldview exploration	Discuss dignity, freedom, trust, responsibility, justice and the common good.
47–56 min	Policy repair	Students propose safeguards: minimisation, explanation, human review, appeal, time limits and proportional restrictions.
56–60 min	Final judgment	Students revise or defend their position with one explicit principle and one case fact.

Key Tensions / Competing Values

- Security vs privacy
- Risk prediction vs presumption of innocence
- Data availability vs data necessity
- Efficiency vs due process
- Automated restriction vs human accountability
- Fraud prevention vs proportionality
- System-wide safety vs individual hardship

Worldview Exploration Prompts

- How should a society balance the common good against individual privacy?
- Does suspicion justify restriction, or only further investigation?
- What duties do institutions have when their protective systems can wrongly harm innocent people?
- What does your worldview say about truth, justice, trust and the proper limits of institutional power?

GOSPEL-INFORMED PERSPECTIVE — FOR FACILITATOR USE

A Gospel-informed reading begins with the dignity of persons rather than treating people merely as risk profiles. Justice requires truthfulness, proportionality and fair treatment; love of neighbour supports protecting people from fraud; stewardship supports prudent use of technology; and awareness of human fallibility cautions against giving either humans or algorithms unchecked power. A high-risk prediction should therefore not be confused with guilt,

Adapted from the HKDSE Ethics and Religious Studies (ERS) Mock Exams | © ers999.com 2026. All rights reserved.

and systems that can significantly harm people should include meaningful accountability and avenues for correction. These ideas can be explored alongside other worldviews without requiring a Christian conclusion.

Adapted from the HKDSE Ethics and Religious Studies (ERS) Mock Exams | © ers999.com 2026. All rights reserved.

CASE 06

AI Fraud Detection

FACILITATOR GUIDE — REASONING & ASSESSMENT

Weak vs Strong Reasoning

Weak / Incomplete	Stronger Reasoning
Weak: “If the bank can access the data, it should use all of it to stop fraud.”	Stronger: “Fraud prevention is important, but data use should be limited to information that is relevant and reasonably necessary. More data may increase surveillance without proportionate security benefit.”
Weak: “If the AI says someone is high risk, the bank should freeze the account.”	Stronger: “A risk score may justify additional investigation, but risk is not proof of wrongdoing. Serious restrictions should require proportionate evidence and, where feasible, human review.”
Weak: “Privacy matters more than security.”	Stronger: “Privacy has strong moral weight, but banks also have duties to protect customers from fraud. The question is whether the data collected and restrictions imposed are necessary, proportionate and contestable.”
Weak: “Using the same algorithm for everyone is fair.”	Stronger: “A uniform process can still be unfair if the model relies on weak proxies, produces unequal errors or gives people no meaningful way to understand and challenge harmful decisions.”

Counterargument Requirement

A strong response must present the best argument against the student's own position before defending, qualifying or revising the final judgment. Students should distinguish inconvenience from serious harm and should avoid treating either privacy or security as automatically absolute.

Alternative Assessment Uses

- TESOL / oral communication: structured discussion using claim → reason → evidence → counterargument → response.
- Composition: 500–700 word argumentative essay on privacy, security and human review.
- Business / Finance: write an ethical policy for AI fraud monitoring and account restrictions.
- Formative assessment: initial position followed by a revised judgment after data-minimisation and proportionality analysis.

Compact 4-Level Ethical Reasoning Rubric

Level	Descriptor
4 — Strong	Defines the precise decision; uses relevant case evidence; weighs privacy, security, fairness and stakeholder interests; distinguishes risk from guilt; presents a serious counterargument; considers alternatives or safeguards; defends a coherent final or conditional judgment.
3 — Competent	Identifies the main ethical issue and relevant evidence; considers more than one perspective; gives reasons for a judgment; engages a counterargument, though proportionality or safeguard analysis may be limited.
2 — Developing	States a position with some reasons but relies on assertion or one-sided analysis; limited case evidence; weak counterargument; little attention to necessity, proportionality or human review.
1 — Beginning	Mostly gives opinion or preference; oversimplifies the decision; minimal use of case evidence; ignores competing perspectives; conclusion is unsupported.

Very Short Source / Context Note

This teaching case is adapted from an ERS Paper 1-B mock question on Human Rights. It uses AI-assisted fraud detection to examine privacy, security, proportionality, due process and the ethical limits of automated restrictions. The scenario is designed for ethical reasoning practice rather than legal or financial advice.

ERS ATTRIBUTION
Adapted from the HKDSE Ethics and Religious Studies (ERS) Paper 1-B Mock Assessment Series.
Original Mock Exam Question ID: P1B-MT-2025Q6-E2Q3-001
Primary PSI area: Human Rights | Cross-cutting context: AI / financial technology
© ers999.com 2026. All rights reserved.