CASE 04

AI-Generated Media & Public Trust

Should AI-Generated Content Be Clearly Labelled?

STUDENT HANDOUT — ENGLISH

REAL-WORLD CONTEXT

Generative AI can now create realistic images, voices, videos and text that may be difficult to distinguish from authentic human-made content. These tools can support creativity, education and accessibility, but they can also be used to mislead audiences. Governments and media platforms are therefore considering whether AI-generated or substantially modified public content should carry clear labels.

The Case

Hong Kong is considering a rule requiring public media content that is generated or substantially modified by AI to carry a clear label.

Supporters argue that people have a right to know when what they are seeing or hearing has been created or altered by AI. They believe clear labelling can reduce deception, protect public trust and help people judge the reliability of information.

Critics argue that a blanket rule could be too broad. AI-generated content is not automatically false, and human-created content is not automatically true. A label may also be unnecessary in obvious fictional, artistic or satirical contexts. Some creators worry that mandatory labels could discourage innovation or cause audiences to distrust legitimate content simply because AI tools were used.

There is also a deeper problem: as deepfakes become more common, people may begin to dismiss genuine recordings as fake. This is sometimes called the “liar’s dividend.”

CORE QUESTION

Should all AI-generated or substantially modified public media content be clearly labelled? Defend your judgment using evidence from the case and seriously consider a strong position different from your own.

Choose a Learning / Assessment Task

- Class discussion: Which kinds of AI-generated content most clearly require a label? Are there cases where a label should not be required?
- Short reflection (250–350 words): Decide whether mandatory labelling is justified. Give reasons, address one strong counterargument and state any exceptions you would allow.
- Structured argument (500–700 words): Present the strongest case for and against mandatory AI-content labels, then defend your own judgment.
- Small-group debate: One group defends a broad labelling rule; another argues for a narrower, risk-based approach. Write an individual final judgment afterward.
- Policy task: Design a labelling policy that distinguishes high-risk deceptive uses from low-risk creative uses and explain who should be responsible for compliance.

Before You Decide

Ask yourself: Does authenticity matter even when the content is factually correct? Is disclosure enough if most people do not notice or understand the label? Who should be responsible—the creator, platform, advertiser, publisher or AI provider? 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 04

AI 生成媒體與公共信任

AI 生成內容是否應清楚標示？

學生講義 — 繁體中文

現實背景

生成式 AI 現在可以製作非常逼真的圖像、聲音、影片及文字，有時很難與真實的人類創作分辨。這些工具可以促進創意、教育及無障礙應用，但亦可能被用來誤導公眾。因此，各地政府及媒體平台正考慮：由 AI 生成或經 AI 大幅修改、並向公眾發布的內容，是否應該有清楚標示。

個案

香港正考慮制定規則，要求所有由 AI 生成或經 AI 大幅修改、並向公眾發布的媒體內容，都必須有清楚標示。

支持者認為，公眾有權知道自己正在觀看或聆聽的內容是否由 AI 生成或修改。他們相信，清楚標示可以減少欺騙、保障公共信任，並幫助人們判斷資訊的可靠程度。

反對者則認為，一刀切的規定可能過於廣泛。AI 生成的內容並不一定是虛假，而由人創作的內容亦不一定是真實。在明顯屬於虛構、藝術或諷刺的內容中，標示可能並無必要。有些創作者亦擔心，強制標示會阻礙創新，或令公眾僅因使用了 AI 工具便不信任本來可信的內容。

此外，還有一個更深層的問題：當 Deepfake 愈來愈普遍，人們可能開始把真正的錄音或影片也說成是假的。這種現象有時被稱為「說謊者紅利」（liar's dividend）。

核心問題

所有由 AI 生成或經 AI 大幅修改、並向公眾發布的媒體內容，是否都應該清楚標示？請運用個案資料為你的判斷提出理據，並認真回應一個與你立場不同的有力論點。

選擇一項學習／評核任務

- 課堂討論：哪些類型的 AI 生成內容最明顯需要標示？有沒有一些情況可以不用標示？
- 短文反思（250–350 字）：判斷是否應強制標示，提出理由、回應一個有力反方論點，並說明你容許的例外。
- 結構化論證（500–700 字）：分別提出支持及反對強制標示的最有力理據，然後為自己的判斷辯護。
- 小組辯論：一組支持廣泛的標示規定；另一組主張較狹窄、按風險分級的做法。辯論後各自寫出最終判斷。
- 政策任務：設計一套標示政策，區分高風險欺騙用途與低風險創意用途，並說明誰應負責遵守規定。

作出判斷前

想一想：即使內容在事實上正確，「真實性」仍然重要嗎？如果大部分人沒有留意或理解標籤，單純披露是否足夠？責任應由創作者、平台、廣告商、出版者還是 AI 供應商承擔？你的世界觀中哪些核心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 04

AI-Generated Media & Public Trust

FACILITATOR GUIDE — ENGLISH

WHY THIS MATTERS NOW

Generative AI is lowering the cost of producing convincing synthetic media. The ethical challenge is not simply whether AI-made content is true or false, but whether audiences can understand what they are seeing, whether disclosure is meaningful, and who carries responsibility when synthetic media misleads. The case also introduces the "liar's dividend": a world in which genuine evidence may be dismissed as fake.

Learning Objectives

- Distinguish truthfulness, authenticity and disclosure as separate ethical questions.
- Identify the exact decision, stakeholders and ethically significant facts in a media-policy case.
- Reason about freedom of expression, the public's right to know, deception, responsibility and proportionality.
- Surface worldview assumptions about truth, trust, freedom, human dignity and social responsibility.
- Present 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	First judgment	Students choose: label all, label some, or no mandatory labels. Require one reason.
5–15 min	Case analysis	Pairs identify stakeholders and three facts that matter ethically.
15–27 min	Concept distinction	Separate truth, authenticity and disclosure. Ask whether an AI-made statement can be true yet still require disclosure.
27–38 min	Competing positions	Groups build the strongest case for broad labelling and the strongest case for a risk-based approach.
38–47 min	Worldview exploration	Discuss truth, trust, freedom, dignity, creativity and responsibility.
47–56 min	Counterargument + policy repair	Students answer the strongest objection to their view and design safeguards or exceptions.
56–60 min	Final judgment	Students revise or reaffirm their position and explain why.

Key Tensions / Competing Values

- Truthfulness vs authenticity
- Freedom of expression vs protection from deception
- Creative innovation vs public trust
- Disclosure vs effective disclosure
- AI-generated ≠ automatically false; human-generated ≠ automatically true
- Uniform rules vs proportional / risk-based regulation
- Creator freedom vs platform and publisher responsibility / the "liar's dividend"

Worldview Exploration Prompts

- Why does truth matter: because it produces good outcomes, respects persons, sustains trust, reflects reality, or for some other reason?
- When freedom of expression and protection from deception conflict, which should have priority and under what conditions?
- What does your worldview say about responsibility for maintaining a trustworthy public information environment?

GOSPEL-INFORMED PERSPECTIVE — FOR FACILITATOR USE

A Gospel-informed reading gives strong weight to truth because truthful speech respects God and neighbour. Human creativity, including AI-assisted creativity, can be welcomed as a gift, but freedom is not freedom to deceive. Imago Dei supports respect for persons whose image or voice may be manipulated, while love of neighbour asks how communication affects vulnerable audiences. Justice, stewardship and wisdom therefore support proportionate safeguards that protect truth and trust without treating all AI-assisted creativity as morally suspect.

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

CASE 04

AI-Generated Media & Public Trust

FACILITATOR GUIDE — REASONING & ASSESSMENT

Weak vs Strong Reasoning

Weak / Incomplete	Stronger Reasoning
Weak: “AI content should always be labelled because AI is fake.”	Stronger: “AI-generated content is not automatically false. The stronger case for labelling is that audiences may reasonably need to know when realistic media has been synthetically generated or substantially altered, especially when deception or impersonation is plausible.”
Weak: “A label solves the problem because people have been informed.”	Stronger: “Disclosure is ethically meaningful only if it is noticeable, understandable and timely. A tiny or ambiguous label may satisfy a rule while still leaving audiences misled.”
Weak: “Creators have freedom of expression, so labelling should never be required.”	Stronger: “Freedom of expression matters, but it does not automatically include a right to conceal material information when synthetic content could deceive viewers or damage another person’s reputation.”
Weak: “All AI content is dangerous.”	Stronger: “Risk varies by context. A fictional artwork, an educational simulation and a deepfake political endorsement create different levels of potential harm and may justify different labelling requirements.”

Counterargument Requirement

A strong response must present the best argument against the student’s own position before defending, qualifying or revising the final judgment. The opposing view should be represented fairly. Assess the quality of the reasoning, not whether students favour broad regulation, narrow regulation or no mandatory rule.

Alternative Assessment Uses

- TESOL / oral communication: structured discussion using claim → reason → evidence → counterargument → response.
- Composition: 500–700 word argumentative essay on universal versus risk-based labelling.
- Formative assessment: classify several hypothetical media examples by labelling risk and justify each decision.
- Summative assessment: design a public-media AI labelling policy with exceptions, enforcement responsibilities and safeguards.

Compact 4-Level Ethical Reasoning Rubric

Level	Descriptor
4 — Strong	Defines the precise decision; uses relevant case evidence; weighs competing values and stakeholders; distinguishes truth, authenticity and disclosure; makes worldview/value assumptions explicit; 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 weighing, distinctions or worldview analysis may be limited.
2 — Developing	States a position with some reasons but relies on assertion or one-sided analysis; limited use of case facts; weak counterargument; little explanation of value priorities or policy conditions.
1 — Beginning	Mostly gives opinion or reaction; oversimplifies AI content as simply true or false; 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 Media Ethics. It examines ethical issues raised by AI-generated or substantially modified public media, including authenticity, deception, disclosure, freedom of expression, public trust, responsibility and the “liar’s dividend.” The scenario is designed for ethical reasoning practice rather than legal 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-ME-E4Q2-001
Primary PSI area: Media Ethics | Cross-cutting context: AI-generated media / public trust
© ers999.com 2026. All rights reserved.