

Poster: Seeing Is Believing? A Study on Trust, Authenticity, and Security When Using Synthetic Materials in Higher Education

MK Michael Cheung*, Miffy H. T. Cheung*, Peter H. F. Ng*, Eugene Yujun Fu†, Grace Ngai* and Youqian Zhang*

*Hong Kong Polytechnic University, Hong Kong, China

†Education University of Hong Kong, Hong Kong, China

Email: {man-ki-michael-mk.cheung, miffy-ho-ting.cheung, peter.nhf, grace.ngai, you-qian.zhang}@polyu.edu.hk, eugenefu@eduhk.hk

Abstract—Advances in generative AI, together with the growth of AI-enabled education, have made it increasingly feasible to create realistic lecture-style videos that replicate a professor’s appearance and voice. As university lecturers’ images, videos, and audio are often publicly available through institutional websites, recorded lectures and online events, attackers may readily repurpose these materials to generate synthetic content. This raises a significant security concern: trusted instructors can be impersonated to spread misleading or false academic information that appears legitimate. In more severe cases, such misuse may extend beyond educational misinformation to scams, fraudulent requests, and other forms of identity-based deception that exploit both the professor’s persona and the credibility of the institution.

This work presents an in-progress human-centred security study of synthetic instructor impersonation in a high-trust academic environment. Using AI-generated instructor-style videos derived from consented source materials, we compare participant responses across contents in different accuracy level. We examine how synthetic instructor identity influences perceived authenticity, judgments of correctness, and susceptibility to misleading information, with the goal of identifying implications for provenance, disclosure, and safeguards for academic media.

1. Introduction

Recent advances in generative AI have made it increasingly feasible to produce realistic e-lectures and instructor-style videos at scale [1], [2], including videos that replicate the appearance and voice of a known professor. As these technologies become more capable, accessible and easier to use, convincing synthetic academic media [3] can now be generated with lower technical barriers, cost and effort. Since professors’ images, recorded lectures and voice materials are often publicly available through university websites, teaching platforms and academic events, such data can be easily collected and repurposed by malicious actors. While these capabilities create new opportunities for educational content production, they also introduce security risks for digital learning environments. Synthetic media can be used

to impersonate legitimate instructors and distribute misleading or false instructional content under the appearance of institutional legitimacy. In more severe cases, such misuse may also enable scams, fraudulent requests, or other forms of identity-based deception that exploit both the professor’s identity and institutional trust.

This risk is not limited to malicious attacks. As AI-assisted content generation becomes easier to use, synthetic educational materials may also be deployed through insufficiently governed workflows in which inaccuracies are not fully noticed before release. As a result, misleading content may still be delivered with the authority of a trusted instructor and the legitimacy of an academic setting. Educational environments are therefore a relevant security context, since instructor identity, institutional affiliation, and lecture format often serve as strong signals of legitimacy. When these signals can be synthetically reproduced, inaccurate content may become difficult to distinguish from authentic instructional material. In this work, we present an in-progress human-centred security study of synthetic instructor impersonation in academic settings, examining how such media may introduce inaccurate content into trusted environments and what this implies for provenance, disclosure, verification, and other safeguards for academic media. This study received IRB approval and complies with all ethical requirements.

2. Research Questions

In this work we consider a threat model in which an attacker, or an inadequately controlled institutional workflow, produces and distributes a synthetic lecture video that appears to originate from a legitimate professor. The synthetic video preserves recognisable identity cues such as facial appearance, voice, and lecture-style presentation, in order to increase its apparent legitimacy. The content may be fully accurate, partially inaccurate, or deliberately false. The attack objective is to exploit the authority of academic identity and the legitimacy of institutional communication channels to increase acceptance of misleading instructional content. Under this framing, our study investigates the following research questions:

- RQ1: To what extent are synthetic instructor videos perceived as authentic academic media?
- RQ2: Does synthetic instructor impersonation increase users' acceptance of inaccurate or misleading instructional content, and how effectively can users distinguish content accuracy levels?
- RQ3: What security risks do synthetic instructor videos pose for academic media?

3. Data Collection and Experiment Design

In this work, we investigate how students and professors evaluate AI-generated e-lecture videos, focusing on how different levels of content inaccuracy affect perceived reliability, error sensitivity, and affective response. We adopted a within-subjects design with three accuracy conditions: Accurate (A), Partially Accurate (pA), and Inaccurate (I). A total of 13 participants were recruited, including 11 university students and 2 academic staff members, enabling an initial comparison between learners and domain experts. Participants were informed that they would evaluate AI-generated lecture videos, but were not told that some contained inaccuracies, in order to reduce demand characteristics.

To construct the stimuli, two professors provided consent for the use of their publicly available profile images and short voice recordings. Each selected a lecture slide from their regular teaching materials and recorded approximately one minute of spoken explanation. Three synthetic e-lecture videos were generated using Deepfake Maker by HeyGen Lab [4]. All videos were standardised in length, pacing, layout, and structure. The key manipulation was content accuracy: the A condition was fully correct, the pA condition contained minor plausible inaccuracies, and the I condition contained more salient errors while remaining fluent.

Each participant viewed all three videos in randomised order. After each video, participants completed a questionnaire assessing perceived trustworthiness [5], [6], perceived accuracy [7], confidence in judgment and affective reactions. After all videos, participants took part in a semi-structured interview exploring evaluation strategies, cues used, suspicion, and intended real-world responses. Questionnaire data will be analysed using repeated-measures comparisons, and interview data will be analysed thematically.

4. Results

Across the three conditions, participants' evaluations remained broadly stable, suggesting that differences in content accuracy did not substantially change how the synthetic lecture videos were received. Many participants appeared to rely more on the portrayed instructor's professional identity, delivery style, and lecture format than on careful verification of the content itself. This indicates that synthetic instructor-style videos may retain perceived legitimacy even when the material is partially or clearly inaccurate.

Interview findings reinforce this concern. Participants often treated familiar academic identity and formal presentation as signals of correctness, and many reported that

they would only question the content when obvious contradictions appeared. Importantly, this vulnerability was not limited to students. Responses from academic staff suggest that even professors overlook subtle inaccuracies in AI-generated educational content if it is not carefully reviewed. Taken together, these findings highlight a broader security issue: misleading content may be delivered under trusted academic identities through both malicious impersonation and insufficiently verified production workflows. This suggests that regular content checking, clear review procedures, and source verification are essential safeguards for synthetic educational media.

5. Conclusion

This work suggests that synthetic instructor contents should be viewed as a complex phenomenon in educational settings, rather than as inherently harmful. In our findings, many participants appeared receptive to this mode of lecture delivery and willing to accept it as a legitimate educational format. Meanwhile, participants often relied on instructor identity, presentation style and lecture format as cues of legitimacy, even when the content was partially inaccurate or clearly false. The security concern therefore lies not in the medium alone, but in its potential misuse or insufficient review, through which misleading content may be delivered with the authority of a trusted instructor. These findings underscore the need for stronger safeguards for synthetic educational media, including content verification, clear review processes and mechanisms for provenance and source authentication. Future work should include a larger and more diverse participant pool to support stronger statistical comparisons and a broader understanding of how different audiences respond to synthetic educational content.

References

- [1] B. Anić, "Artificial intelligence and deepfake learning in higher education," *Journal of Baltic Science Education*, vol. 24, no. 1, pp. 4–6, 2025.
- [2] H. Sharif, A. Atif, and A. A. Nagra, "Deepfake-style ai tutors in higher education: A mixed-methods review and governance framework for sustainable digital education," *Sustainability*, vol. 17, no. 21, p. 9793, 2025.
- [3] S. Siehl, K. Kammler-Sücker, S. Guldner, Y. Janvier, R. Zohair, and F. Nees, "To trust or not to trust? face and voice modulation of virtual avatars," *Frontiers in Virtual Reality*, vol. 5, p. 1301322, 2024.
- [4] L. C. Rahmawati and F. Sikouonmeu, "Ai-generated content as the future of marketing strategies: Selecting the right generative ai tools for enhanced campaign effectiveness," in *Generative Künstliche Intelligenz in Marketing und Sales: Innovative Unternehmenspraxis: Insights, Strategien und Impulse*. Springer, 2024, pp. 11–23.
- [5] P. B. Forsyth, C. M. Adams, and W. K. Hoy, *Collective trust: Why schools can't improve without it*. ERIC, 2011.
- [6] J.-Y. Jian, A. M. Bisantz, and C. G. Drury, "Foundations for an empirically determined scale of trust in automated systems," *International journal of cognitive ergonomics*, vol. 4, no. 1, pp. 53–71, 2000.
- [7] G. Pennycook, J. A. Cheyne, N. Barr, D. J. Koehler, and J. A. Fugelsang, "On the reception and detection of pseudo-profound bullshit," *Judgment and Decision making*, vol. 10, no. 6, pp. 549–563, 2015.

Seeing Is Believing? A Study on Trust, Authenticity, and Security When Using Synthetic Materials in Higher Education

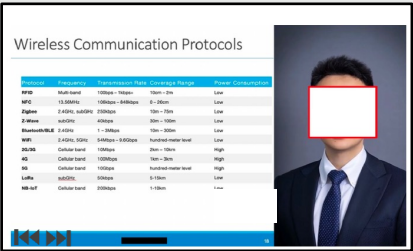
MK Michael Cheung¹; Miffy H.T. Cheung¹; Peter H.F. Ng¹; Eugene Yujun Fu²; Grace Ngai¹; Youqian Zhang³

1. Department of Computing, The Hong Kong Polytechnic University, Hong Kong, China
2. Centre for Learning, Teaching and Technology, The Education University of Hong Kong, Hong Kong, China
3. Department of Electrical and Electronic Engineering, The Hong Kong Polytechnic University, Hong Kong, China

BACKGROUND

Recent generative AI advances make it easy to create realistic synthetic lecture videos. In higher education, where teaching materials and instructor media are often publicly available, it creates both new teaching opportunities and security risks, as misleading content can appear credible when presented through instructor-like videos. This work explores the following:

- Examines how AI-generated lecturer videos are perceived in higher education contexts.
- Investigates how different levels of content accuracy influence trust, authenticity, and reliability judgments.
- Highlights the security risk that misleading educational content can gain legitimacy when presented through realistic synthetic instructor identities.



[Fig. 1] Example of realistic AI-generated lecture

Ethics

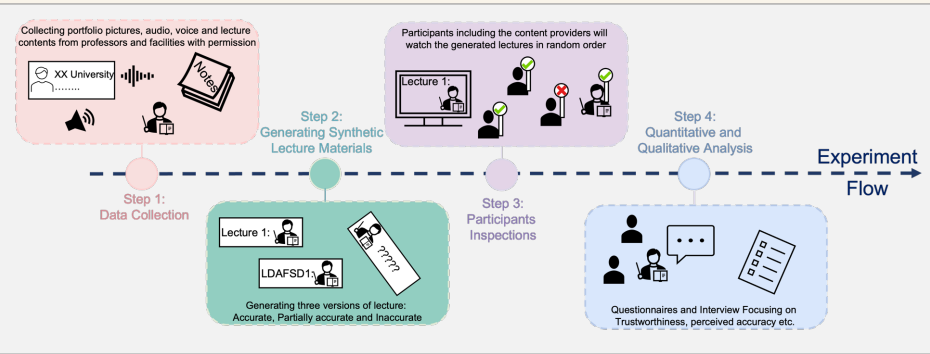
The study was ethically handled through IRB approval, the use of consented professor materials, and clear participant information that they would evaluate AI-generated videos. Limited withholding about inaccuracies was used only to reduce bias in responses.

METHOD AND RESULTS

Subjects:

This exploratory pilot study involved 13 participants, including 11 students and 2 academic staff, allowing an initial comparison between learners and domain experts. As all participants regularly engage with academic lecture content, they were well suited to evaluate the credibility, reliability, and accuracy of AI-generated e-lecture videos. Participants were told they would assess synthetic videos, but not that some contained inaccuracies, to reduce demand characteristics. Although the sample was small, it is appropriate for a pilot within-subjects study, and the findings should be treated as preliminary and informative for future research.

Experiment Design:



[Fig. 2] Experiment design of the study

Results:

- Responses were broadly similar across the Accurate, Partially Accurate, and Inaccurate conditions.
- No substantial differences were found in trust, authenticity, or content evaluation.
- Partially accurate and inaccurate videos were often judged similarly to the accurate video.
- Participants appeared to rely more on instructor identity, presentation style, and lecture format than on content accuracy.
- This suggests synthetic instructor videos may increase the credibility of misleading educational content.

CONCLUSION

This study highlights both the promise and the risk of synthetic instructor videos in education. While they may be accepted as a legitimate teaching format, they can also make misleading content appear credible. The key takeaways of this work are as follows:

- Synthetic instructor videos were often perceived as legitimate educational media.
- Familiar cues, such as instructor identity and lecture style, reduced sensitivity to inaccuracies.
- Safeguards and further research are needed to reduce misuse and improve trust in academic media.

REFERENCES

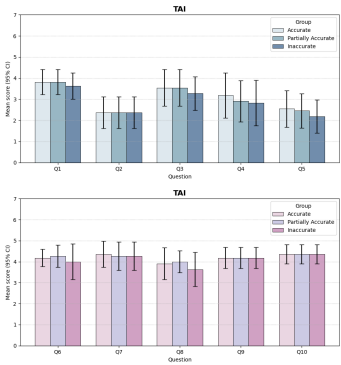
- [1] B. Ani'c, "Artificial intelligence and deepfake learning in higher education," Journal of Baltic Science Education, vol. 24, no. 1, pp. 4–6, 2025.
[2] H. Sharif, A. Atif, and A. A. Nagra, "Deepfake-style ai tutors in higher education: A mixed-methods review and governance framework for sustainable digital education," Sustainability, vol. 17, no. 21, p. 9793, 2025.
[3] S. Siehl, K. Kammler-Sücker, S. Guldner, Y. Janvier, R. Zohair, and F. Nees, "To trust or not to trust? face and voice modulation of virtual avatars," Frontiers in Virtual Reality, vol. 5, p. 1301322, 2024.

ACKNOWLEDGEMENT

This project was approved by the university's Institutional Review Board (IRB) with the reference number: HSEARS20260210005.

CONTACT

MK Michael Cheung
The Hong Kong Polytechnic University
Email: man-ki-michael-mk.cheung@polyu.edu.hk



[Fig. 3] Trust in Automation scale (TAI) questionnaire statistic results from participants