THE UNIVERSITY OF ARIZONA
College of Information Science

AI-Enabled Fraud and What It Means for Audit

ISACA Tucson · November 19, 2026 · Ryan Straight, Ph.D., University of Arizona

The Arup Deepfake Call

HK$200 million

About US$25.6M, moved in 15 transfers after one deepfake video call. No internal system was compromised. [1]

A Hong Kong finance employee joined a video call in which the CFO and colleagues were deepfakes. Seeing and hearing familiar faces overrode his suspicion of the original email. He checked with head office about a week later. The person on the call was the control the attack aimed at.

CHEAP PERSUASION

54%

Clicked an AI-written phish, as many as for human experts. Generic: 12%. About 25 volunteers per group. A 7,741-person field study: 2.4 times generic [15]. [2]

$893M

Losses in FBI complaints that mentioned AI, 2025, across 22,364 complaints. Probably an undercount; over $632M of it is investment scams. [3]

$3.05B

Business email compromise losses, 2025, up from $2.77B in 2024. [3]

4 of 6

Voice-cloning products lacking basic safeguards against cloning someone else's voice. [4]

The VOLTA infostealer sends each victim's stolen data to a commercial AI model, prompted as "a forensic intelligence analyst drafting a target assessment," which ranks what the victim is worth [5].

LIMITS OF DETECTION

61.3%

Human-written essays by non-native English writers flagged as AI, average of seven detectors. US eighth-graders: 5.1%. [7]

70% → 3.3%

AI-essay detection after one rewrite prompt. 31 essays; proof of concept. [7]

3,593 vs 3,087

Clicks with a voluntary training page after a failed simulation vs without one. 14,733 employees at one company, 15 months. [8]

HUMAN IN THE LOOP

86%

How often experts went with an AI agent's hidden, randomly assigned lean between two defensible options. Chance: 50%. [9]

15%

Experts who mentioned the agent at all. [9]

"Institutions often treat an expert's sign-off as evidence of independent judgment, and as AI does more of the work, that assumption becomes shakier." [9]

AUDIT RESPONSES

Payments	Sample bank-detail changes and urgent payments. Was the callback made on a channel the requester didn't choose, to a number on file before the request, before the money moved? Look for a call log or hold record, not a checkbox.	COBIT 2019 DSS06
Reviews	For AI-assisted work, was AI used, which alternatives were considered, who chose, and is the interaction record kept? Which decisions need a person's own informed judgment [9]? A missing record is an exception.	NIST AI RMF GOVERN 3.2, MAP 3.5
Detection	"The AI flagged nothing" is a claim to test. What is each tool's false-negative rate, how was it measured, on whose data, and when was it last re-tested [2, 7]?	NIST AI RMF MEASURE, where the tool is AI
Awareness	Completion rates measure activity. What happened to clicks and reports after training [8]? Is there a report button, and is its use tracked?	COBIT 2019 APO07
Exposure	If an infostealer took one employee's browser sessions today, what would an AI rank as worth selling [5]? Whose single machine could move money [10]?	COBIT 2019 APO12, DSS05

US$6.34 billion

Taken from Binance, October 29, 2026; about US$2.85B frozen (Binance's figures). [10]

The NIST AI RMF covers an organization's own AI, so it fits Reviews and Detection; the rest are COBIT controls. ISACA's AAIA covers auditing your own AI [12]; NIST's draft Cyber AI Profile addresses AI-enabled attacks [16].

Reporting and Protection Tools

- The FBI's Internet Crime Complaint Center, ic3.gov, takes fraud reports. Its annual report [3] comes from those complaints.
- haveibeenpwned.com shows whether an email address appears in known data breaches, and can notify you of new ones.
- The FBI advises families to agree on a secret word or phrase to verify identity against voice-clone scams [11]. A pre-agreed callback code for payment approvers applies the same idea at work (my suggestion, not the FBI's).
- Chrome Enhanced Protection sends suspicious pages and copied commands for analysis [6]. Its "ten times safer" figure counts warnings shown, not attacks caught.
- Your own report button. In a 15-month field study, employees kept reporting all the way through, and about 10% of reports arrived within five minutes [8].

Slides, handout, sources and tools

ryanstraight.com

Sources

1. Magramo, K. "British engineering giant Arup revealed as $25 million deepfake scam victim." CNN Business, May 16, 2024. https://www.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk/index.html
2. Heiding, F., Lermen, S., Kao, A., Mayrink Verdun, C., Schneier, B., and Vishwanath, A. "Evaluating large language models' ability to automate spear phishing." *Expert Systems with Applications* 314 (2026), 131546. Preprint: https://arxiv.org/abs/2412.00586
3. FBI Internet Crime Complaint Center. *2025 Internet Crime Report* (April 2026), pp. 26, 39–40, 61. "AI-related" means the complaint mentions AI. https://www.ic3.gov/AnnualReport/Reports/2025_IC3Report.pdf
4. Gedye, G. "Do these 6 AI voice cloning companies do enough to prevent misuse?" Consumer Reports, March 10, 2025. https://innovation.consumerreports.org/new-report-do-these-6-ai-voice-cloning-companies-do-enough-to-prevent-misuse/
5. Fetterman, R. "Appraisal by AI: model-driven victim prioritization in the VOLTA infostealer." Cisco Talos, October 31, 2026. https://blog.talosintelligence.com/appraisal-by-ai-model-driven-victim-prioritization-in-the-volta-infostealer/
6. Bawa, J. "Our latest defenses against evolving web scams." Google, October 20, 2026. https://blog.google/security/our-latest-defenses-against-evolving-web-scams/
7. Liang, W., Yuksekgonul, M., Mao, Y., Wu, E., and Zou, J. "GPT detectors are biased against non-native English writers." *Patterns* 4(7), 100779 (2023). https://doi.org/10.1016/j.patter.2023.100779
8. Lain, D., Kostiainen, K., and Čapkun, S. "Phishing in organizations: findings from a large-scale and long-term study." IEEE Symposium on Security and Privacy, 2022. https://doi.org/10.1109/SP46214.2022.9833766
9. Herdağdelen, A., Joren, H., and Tamkin, A. "Who's aligned to whom? How AI agents shape experts' research judgment calls." Anthropic, October 27, 2026. https://www.anthropic.com/research/agents-judgment-calls
10. Binance. "29 October 2026 Binance Security Incident: Initial Findings and Response Update." November 3, 2026; initial findings, subject to further independent review. https://www.binance.com/en/blog/community/3229153443759566763
11. FBI IC3. "Criminals use generative artificial intelligence to facilitate financial fraud." Public service announcement I-120324-PSA, December 3, 2024. https://www.ic3.gov/PSA/2024/PSA241203
12. ISACA. "AAIA exam content outline." https://www.isaca.org/credentialing/aaia/aaia-exam-content-outline
13. Palmer, K. "AI detectors are out, new assessments are in." Inside Higher Ed, August 5, 2026. https://www.insidehighered.com/news/tech-innovation/artificial-intelligence/2026/08/05/ai-detectors-are-out-new-approaches-are
14. Gray, C. M., Kou, Y., Battles, B., Hoggatt, J., and Toombs, A. L. "The dark (patterns) side of UX design." CHI 2018. https://doi.org/10.1145/3173574.3174108
15. Czybik, S., Kouam, A. J., Heubl, P., Nold, J. M., and Rieck, K. "A large-scale study of personalized phishing using large language models." USENIX Security Symposium, 2026. https://www.usenix.org/conference/usenixsecurity26/presentation/czybik
16. NIST. *Cyber AI Profile*, NIST IR 8596, initial preliminary draft, December 16, 2025. https://csrc.nist.gov/pubs/ir/8596/iprd

Further Reading

- The Anthropic study [9] is a short read and the source of the sign-off findings in this talk.
- Straight, R. "So the Detectors Are Out." ryanstraight.com, October 13, 2026. https://ryanstraight.com/posts/2026-10-13-detectors-out/
- Straight, R. "The Phishing Training Page That Made Things Worse." ryanstraight.com, November 11, 2026. https://ryanstraight.com/posts/2026-11-11-phishing-training/
- Straight, R. "The New CSI." ryanstraight.com, October 29, 2026. https://ryanstraight.com/posts/2026-10-29-new-csi/
- Straight, R. "Adversarial intentionality." *Techné: Research in Philosophy and Technology*, accepted 2026.