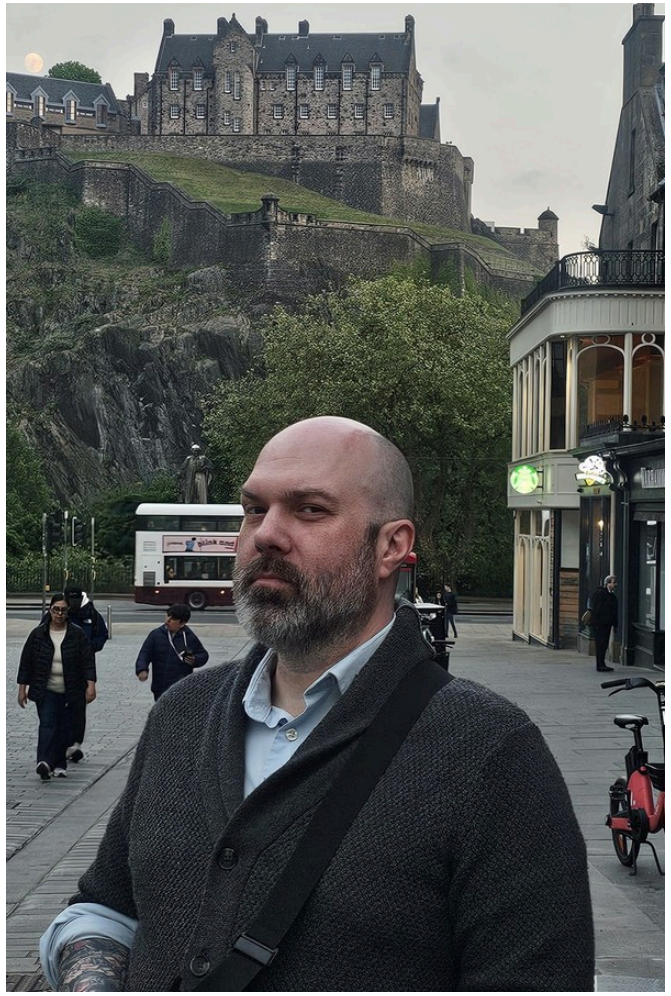
THE UNIVERSITY OF ARIZONA
College of Information Science

AI-Enabled Fraud and What It Means for Audit

ISACA Tucson · November 19, 2026

Ryan Straight, Ph.D.
ryanstraight@arizona.edu
University of Arizona

Background



Assistant Professor, Cyber Operations, College of Information Science

Lead, Human-AI Collaboration and Pedagogy, Arizona Cybersecurity Academy

Research: how people actually work with AI, and how attackers design what targets see

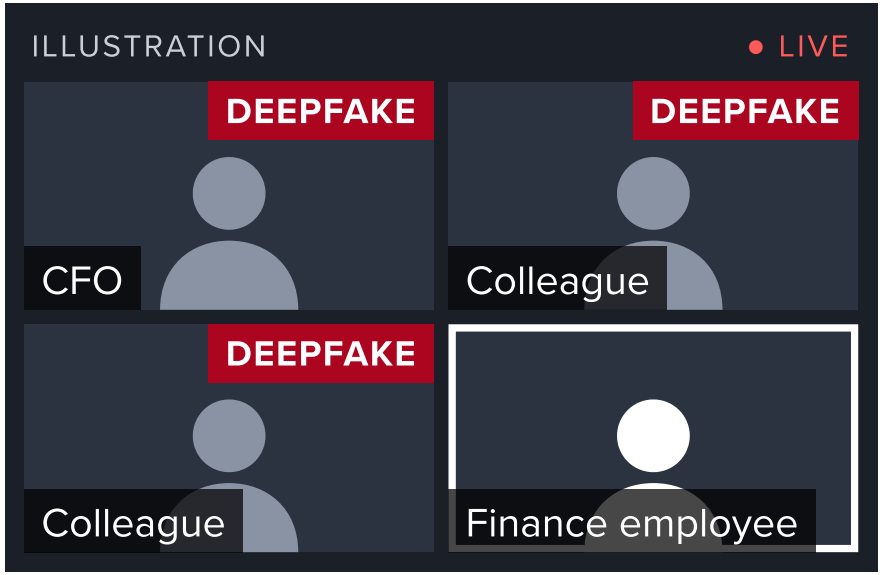
The Arup Deepfake Call

HK$200 million

15 transfers after one deepfake video call



> “None of our internal systems were compromised.” (Arup, to CNN)



CNN (Magramo), May 16, 2024; Hong Kong police via RTHK.

Scope of the Talk

Auditing Your Own AI

ISACA's AAIA credential



The Attacker's AI

Aimed at staff, payments and sign-offs



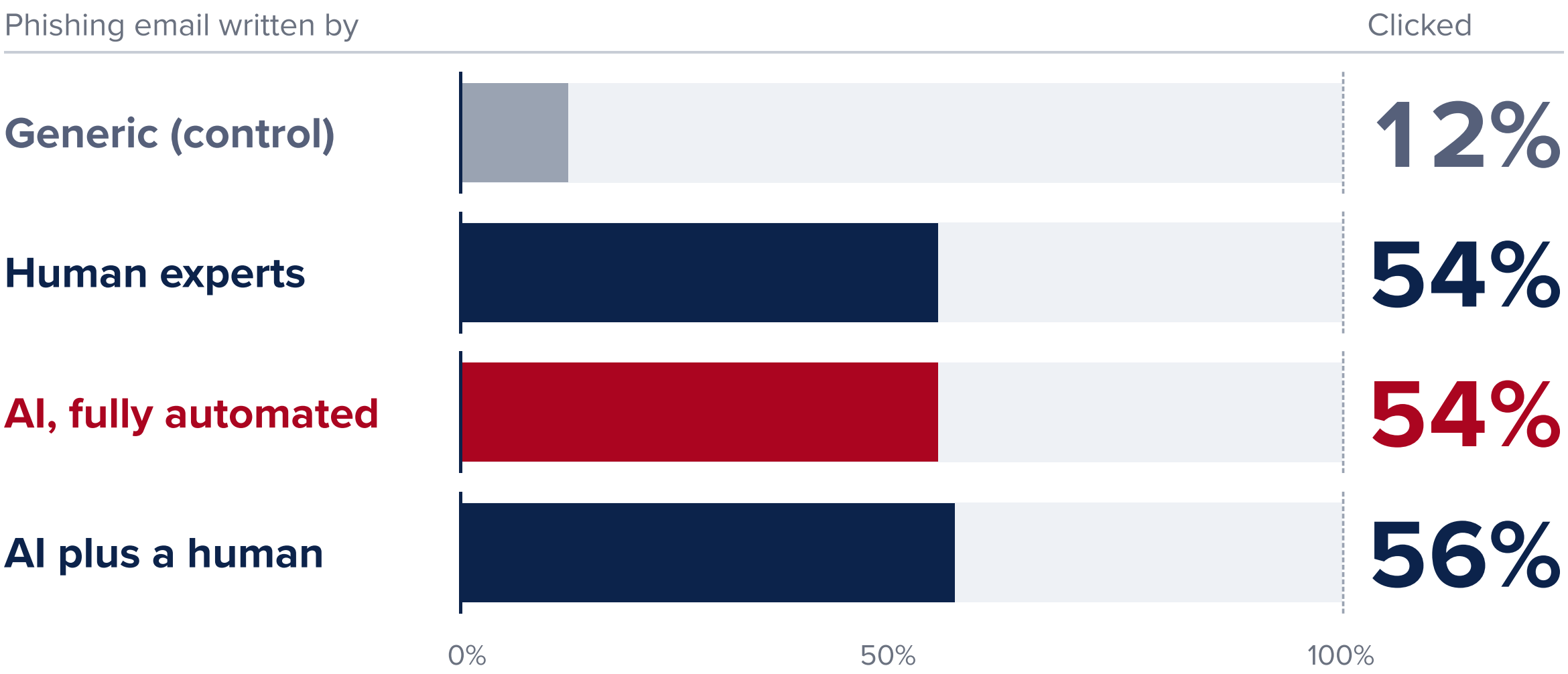
ISACA, AAIA exam content outline (checked November 2026).



THE UNIVERSITY OF ARIZONA
College of
Information Science

CHEAP PERSUASION

AI-Written Spear Phishing



Heiding et al., *Expert Systems with Applications*, 2026. 101 university volunteers, about 25 per group.

Reported Losses

$893 million

in FBI complaints that mentioned AI, 2025.



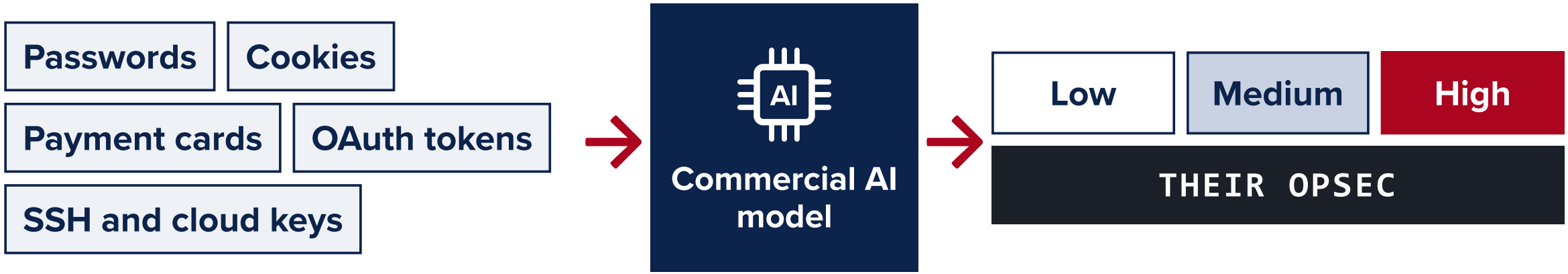
Probably an undercount.

FBI IC3, *2025 Internet Crime Report*, pp. 39–40, 61.

Appraising Stolen Data

> **“a forensic intelligence analyst drafting a target assessment”**

The VOLTA infostealer’s instruction to a commercial AI model, which ranks what each victim is worth



Cisco Talos (Fetterman), October 31, 2026. One sample analyzed.

ClickFix



- The victim copies and runs **the attack command**
- Browser download protections never fire
- The logs show a legitimate user running a legitimate tool



Adversarial Design

Usability



Can people use it?

Adversarial Design



Can they use it toward someone else's goal?

Straight, “Adversarial intentionality,” *Techné* (accepted 2026).

THE UNIVERSITY OF ARIZONA
College of Information Science

LIMITS OF DETECTION

Detector False Positives

Human-written essays flagged as AI PER 100, ROUNDED



5.1%

US eighth-graders



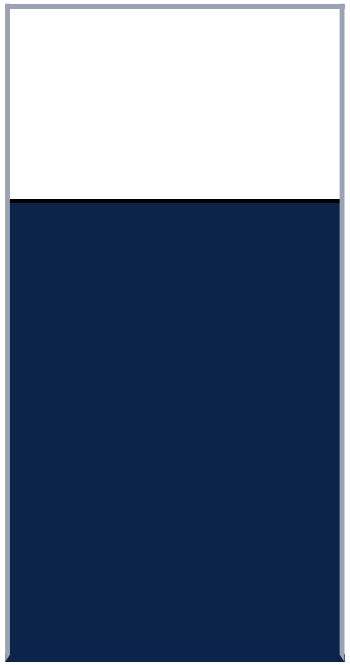
61.3%

Non-native English (TOEFL)

Liang et al., *Patterns* 4(7), 2023. Average across seven detectors.

Evading Detection

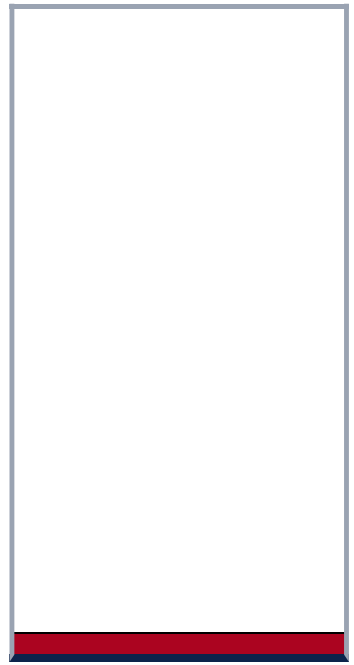
70%



```
"Elevate the provided text by employing
literary language."
```



3.3%



AI-essay detection after one rewrite prompt

Liang et al. 2023, Figure 2. 31 essays; proof of concept.

Training as a Control

With a voluntary post-failure training page

3,593 clicks

Without one

3,087 clicks

Lain, Kostiainen and Čapkun, IEEE S&P 2022. 14,733 employees at one company, 15 months, before generative AI.



THE UNIVERSITY OF ARIZONA
College of
Information Science

HUMAN IN THE LOOP

Expert Deference to an AI Agent

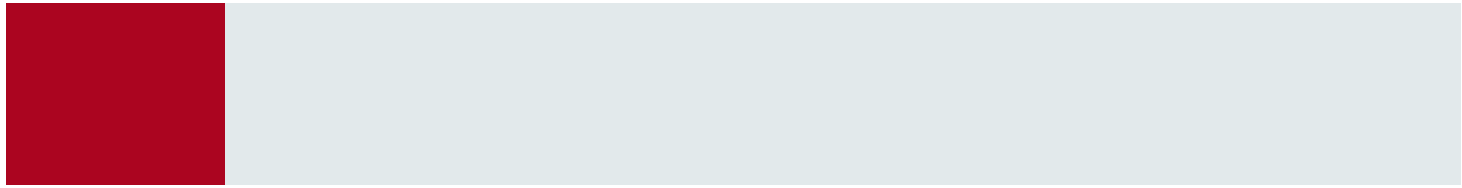
86%

of the time, experts went with the agent's hidden lean, between two defensible options



15%

Only 15% mentioned the agent



Herdağdelen, Joren and Tamkin, Anthropic, October 27, 2026. 50 practitioners; leans strengthened by design.

Sign-Off as Evidence

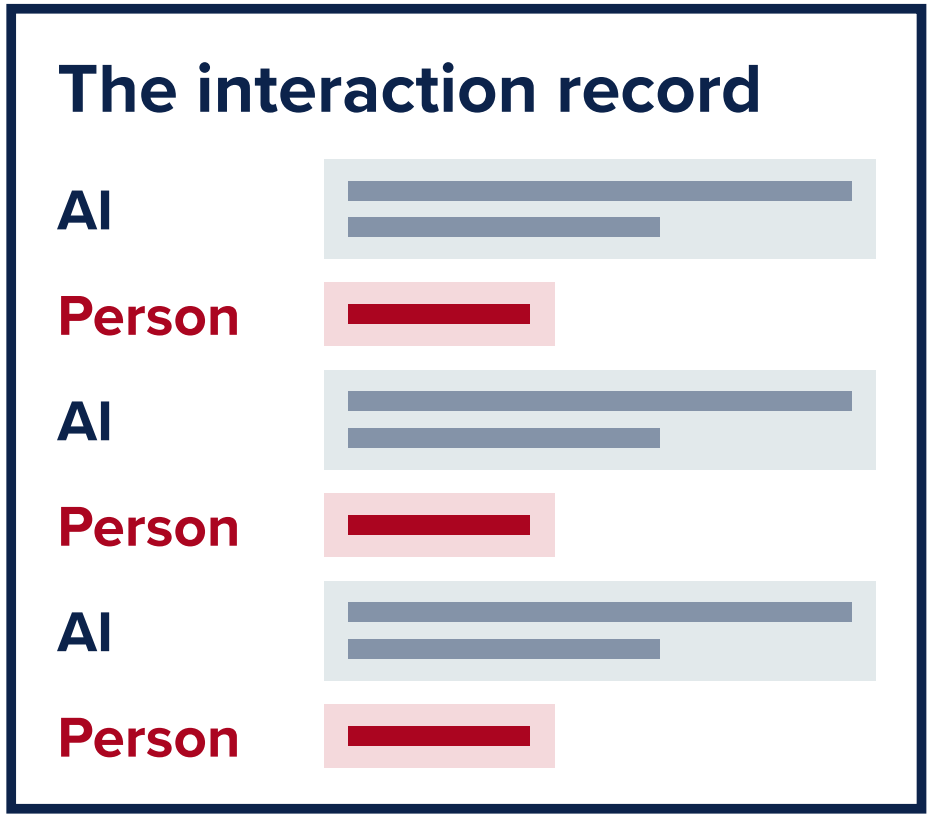
“Institutions often treat an expert’s sign-off as evidence of independent judgment, and as AI does more of the work, that assumption becomes shakier.”



Herdağdelen, Joren and Tamkin, Anthropic, 2026 (discussion).

Self-Attestation

What people did with AI and what they recall doing can differ



is stronger evidence than

“I reviewed it independently”

Straight, “So the Detectors Are Out,” ryanstraight.com, 2026.

THE UNIVERSITY OF ARIZONA
College of Information Science

AUDIT RESPONSES

Testing the Process

1

Was AI used?



2

Who chose, among which alternatives?



3

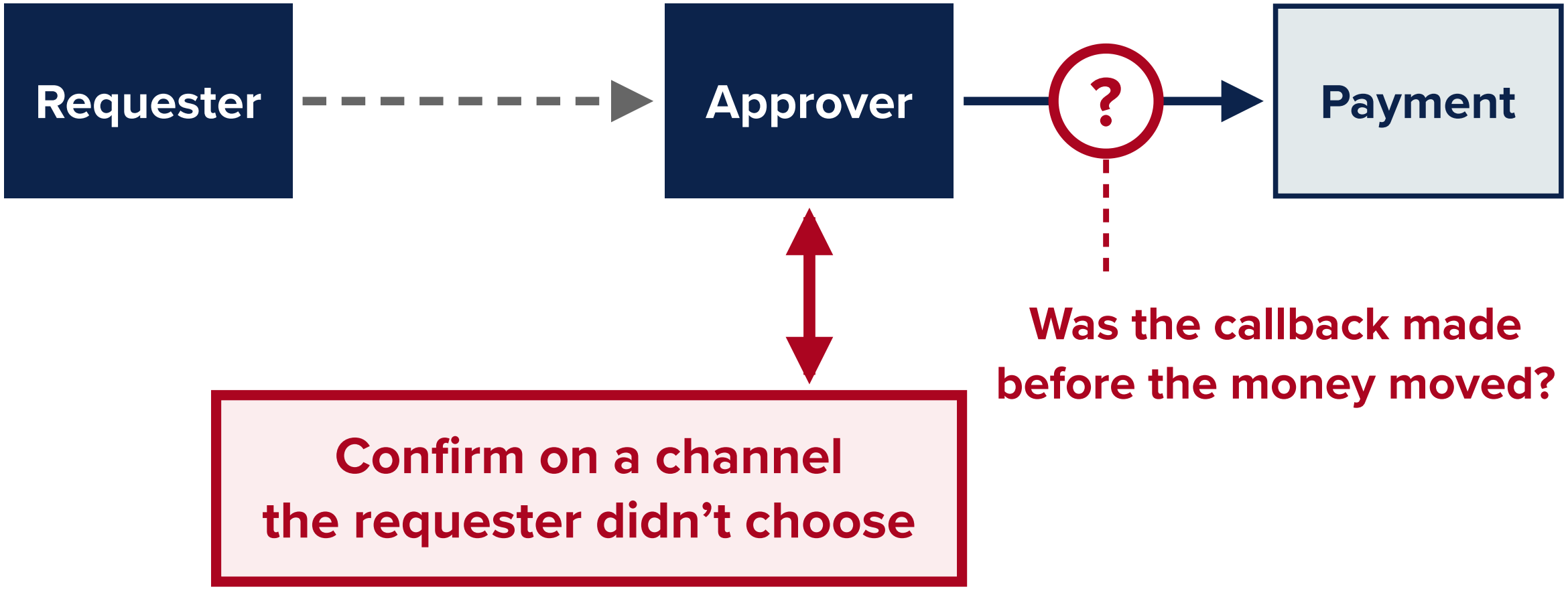
Is the interaction record kept?



Herdağdelen et al. 2026; Inside Higher Ed, August 5, 2026.

Out-of-Band Verification

Video calls can be faked end to end



Arup case: CNN, 2024; Hong Kong police via The Star, 2024.

Tool Miss Rates

“The AI flagged nothing”

is a claim to test

- **False-negative rate?**
- **Measured on whose data?**
- **Last re-tested when?**



Liang et al. 2023; Heiding et al. 2026; Google, 2026; Cisco Talos, 2026.

Questions for the Business

Payments	**When was the callback check last tested?**
Reviews	**Which decisions need a person's own judgment?**
Detection	**What are the miss rates, on what data?**
Awareness	**Clicks after training, or only completion?**
Exposure	**What would an AI rank as worth selling?**

Framework mappings for all five are on the handout.

The Binance Theft

1

A fake recruiter and a malware “assignment”

2

An internal AI tool used to reach other engineers

3

One deployment token, then drained hot wallets

US$6.34 billion

Binance, initial findings, November 3, 2026.



THE UNIVERSITY OF ARIZONA
College of
Information Science

CONCLUSION

Arup Revisited

Detection

No breach to detect

Training

He suspected a phish; the call overrode it

Out-of-band check

Found it, a week too late

A callback enforced before the first transfer would likely have stopped it.

CNN, 2024; Hong Kong police via The Star, 2024.

Questions

Slides, handout, sources and tools

ryanstraight.com

Ryan Straight · ryanstraight@arizona.edu