



The use of AI to perpetrate and prevent fraud

Kristen Moore

Acting Group Leader: Cybersecurity & Quantum Systems

CSIRO's Data61, Australia



OFFICIAL

Agenda

- 1 Where AI is helping (and hurting) fraud prevention today
- 2 Implementation realities: What actually works
- 3 Using AI responsibly: Managing risks and maintaining trust

OFFICIAL



Why cyber-enabled fraud looks different now

- AI amplifies fraud at both ends:
 - Mass-scale, low-cost scams
 - High-credibility, targeted impersonation
- Fraud tools — deepfakes, synthetic IDs, forged documents — are now in every fraudster's toolkit
- Traditional controls are strained by the speed, scale, and creativity of attacks



1. Where AI is helping (and hurting) fraud prevention today






Where AI is Already Helping

- **Fraud pattern recognition at scale**
 - Detection anomalies in grants, welfare, and payroll before disbursement
- **Smarter procurement oversight**
 - Analyse text to flag conflicts of interest & bid-rigging
- **Network & relationship mapping**
 - Detect shell companies & hidden relationships
- **Insider threat & misuse detection**
 - Behavioural anomaly detection for access abuse/data theft
- **Strengthening Digital ID & onboarding**
 - Detect deepfakes in biometric & document verification





Fraud vectors enabled by AI-generated media

Identity fraud:

-  Bypass biometric verification (face/voice spoofing)
-  Create synthetic IDs and forged documents
-  Fake records to access benefits or entitlements
Impersonate executives to authorise fraudulent transfers

Impersonation & manipulation:

-  Video impersonation of executives to authorise fraudulent transfers
-  Clone voices to manipulate staff in real time

Real-world example:

Feb 2025 – Italy: AI-cloned voice of Defence Minister demanded €1M “ransom” from prominent business leaders, convincing some to pay.



Case Study: Synthetic Job Applicants

The Stats

Gartner

1 in 4 candidate profiles fake by 2028

20% rise in video deepfake fraud (2022-2024)

90% rise In candidate fraud since shift to remote work

36% Australians targeted by deepfake scams

The Attack Path

Identity Creation



Generated resume, LinkedIn profile, and work history



Interview Deception



Deepfake video/voice for remote interview



Hiring & Access



Role obtained – remote access granted



Malicious Activity

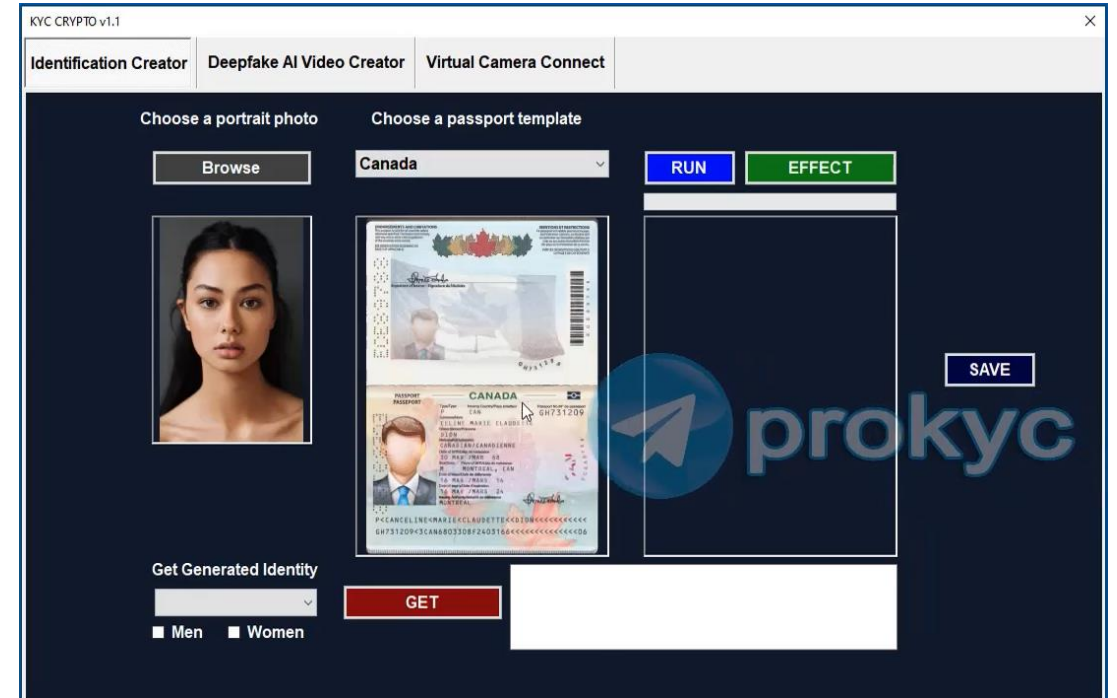


Espionage, data theft, or financial fraud.



ProKYC: Fraud-as-a-Service for Onboarding

- Offers custom AI-generated IDs and forged documents
- Deepfake liveness bypass for biometric checks
- Synthetic proof-of-income and utility bills
- Enables criminals to pass KYC remotely at scale





KYC Risks & Strengthening KYC

The Problem:

- KYC assumes documents, photos, and videos are genuine.
- Deepfakes break KYC at onboarding, authentication, and ongoing monitoring.

The Risks:

- Regulatory, Reputational, Operational

The Solution:



Layered authentication



Advanced liveness + data triangulation



Post-onboarding monitoring + red-teaming”



Staff training



OFFICIAL

2. How to make AI work in practice

OFFICIAL



Making AI Work in Practice



Essentials for success

- **Data Readiness:** Start with clean, representative, and risk-relevant data
- **Workflow Integration:** Integrate AI into workflows — don't bolt it on as a separate tool
- **Cross-functional Collaboration:** Deliver with cross-functional teams: fraud, data, tech, compliance

Our work in combating deepfakes

Systematic evaluation of leading detectors

- Assessed robustness & failure points
- Why accuracy drops in deployment:
 - 1) Preprocessing mismatch
 - 2) Deepfake generator diversity gap
- Tested on public datasets + fakes found in the wild
- **No detector excelled across all scenarios**



Input image



Face detected from engine #1



Face detected from engine #2



3. Using AI responsibly: Managing risks and maintaining trust



Managing Risk & Building Trust

Risks

Bias in training data

Model drift over time

Opaque, black-box decisions

Over-automation without oversight

Best Practices

Explainable AI models that show their reasoning

Red-teaming & adversarial testing

Human-in-the-loop for critical decisions

Cross-functional governance from design to deployment

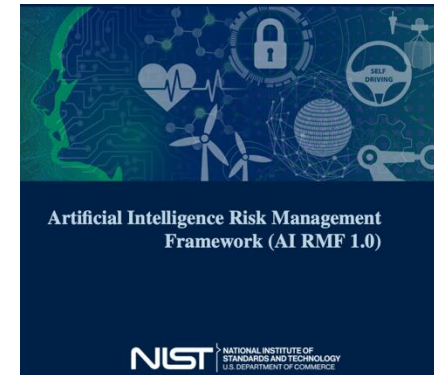


Aligning with AI Assurance Frameworks

1. Strategic Foundations

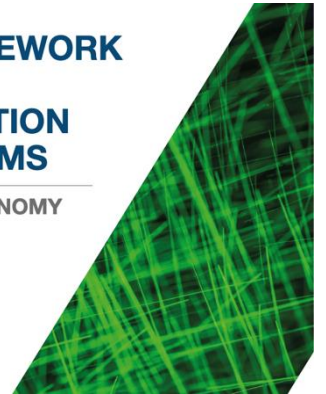
Define the scale, type, and context of AI risk before deployment

- NIST AI Risk Management Framework



OECD FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS

OECD DIGITAL ECONOMY PAPERS
February 2022 No. 323



2. Testing & Preparedness

Stress-test AI systems under real-world and adversarial conditions

- CDEI's Assurance Toolkit – Tools for testing AI robustness, reliability, and fairness

3. Transparency & Assurance

Prove your AI is fair, explainable, and accountable

- Model Cards for model documentation
- Australian National AI Assurance Framework



Human-Centric Fraud Prevention with AI

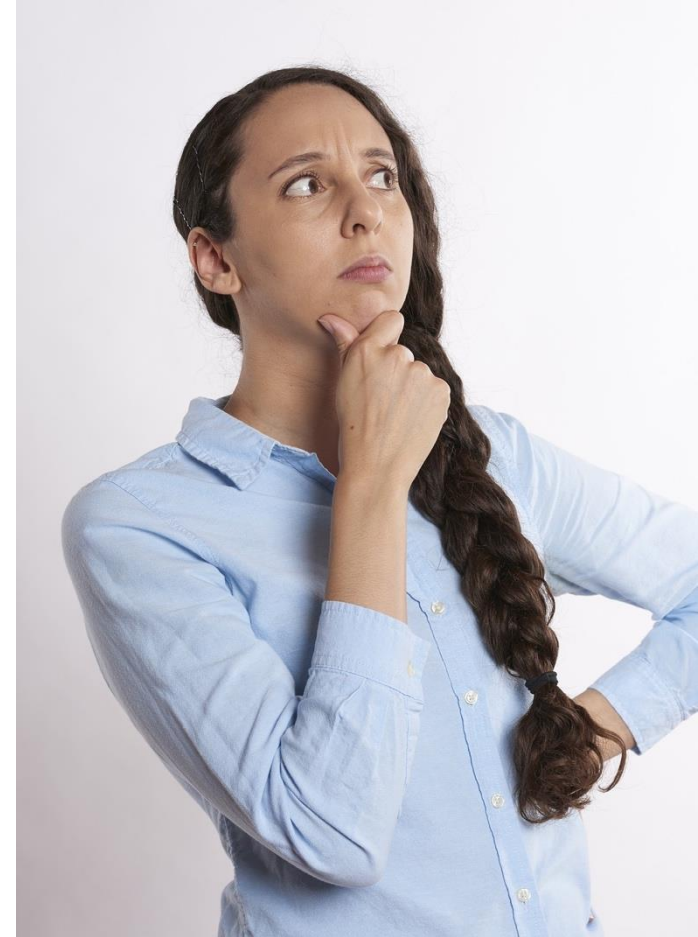
Reality: flag fraud, approve claims, escalate alerts — but humans still make the call.

The problem:

- People often misjudge deepfakes and are overconfident in their ability to spot them
- Leads to missed fraud, false approvals, and reputational harm

The solution:

- AI that explains its reasoning
- Supports – not replaces – human judgment
- Allows intervention, override, and challenge



Emerging AI Risk Domains

Threats



Multimodal deepfakes
(face + voice + text)



AI-assisted social
engineering



Adaptive synthetic IDs

Responses



Adaptive
detection
methods



Cross-sector
intelligence
sharing



Closing Thoughts

Key takeaways:

- AI can significantly enhance fraud detection — if implemented with care
- Deepfakes and synthetic content require new identity assurance methods
- Risk and assurance practices must evolve alongside capability
- Human–AI collaboration, not replacement, is the path to trusted systems

Our next challenge: building fraud prevention systems that adapt as fast as fraudsters innovate

Thank you

Please get in touch!

Email: kristen.moore@data61.csiro.au

LinkedIn: <https://www.linkedin.com/in/kristenlmoore/>

Webpage: <https://people.csiro.au/m/k/kristen-moore>

