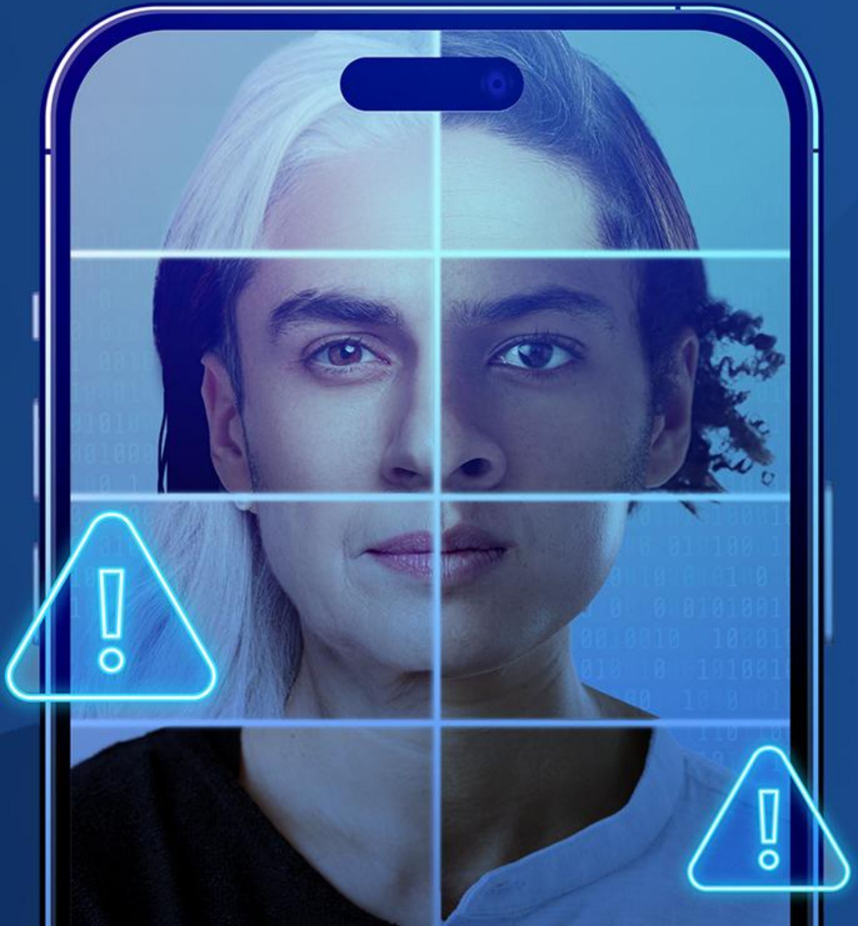
Identifying, Defending and Protecting Against Emerging Threats to Biometric Face Verification

Gemma Bird

Head of Biometric Platform

iProov



Emerging Threats to Biometric Face Verification

- A recognised risk
- What are they; injection vs presentation attacks
- Why do they matter?
- Defending against them

The threat from synthetic imagery is increasingly well understood by policymakers

enisa

EUROPEAN UNION AGENCY FOR CYBERSECURITY

THREATS 2030

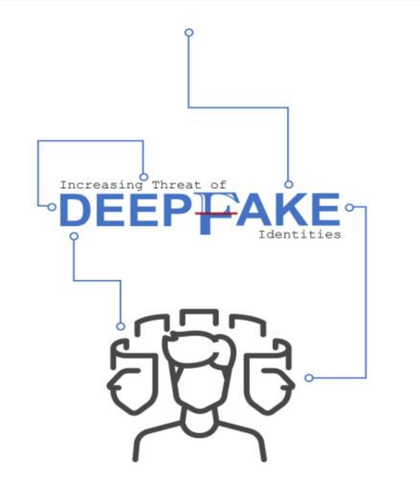
IDENTIFYING EMERGING CYBER SECURITY THREATS AND CHALLENGES FOR 2030

MARCH 2023

*"In **2030**, non-state actors like criminal groups, hackers-for-hire as well as government actors will likely have the **technological capabilities** (e.g., deepfakes) to expand their disinformation efforts in the EU to **manipulate communities**."*

ENISA, 2023

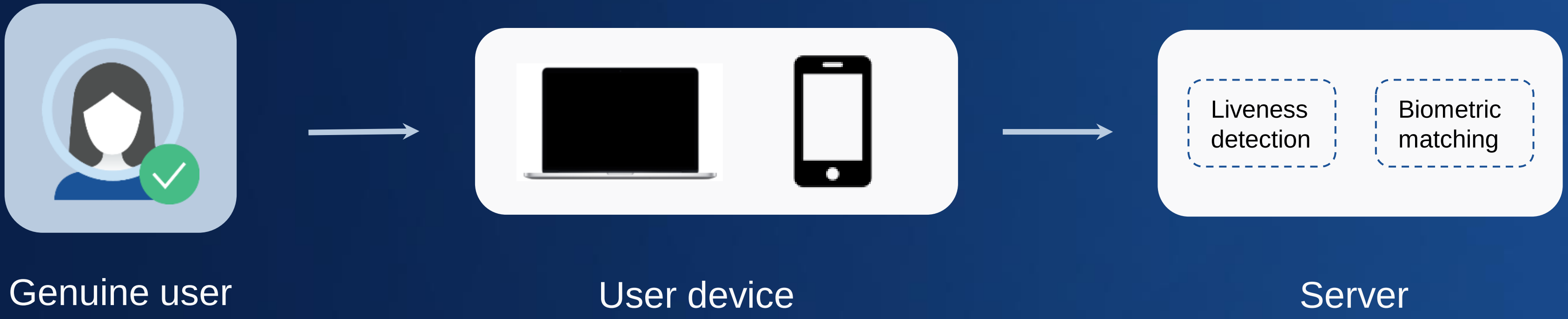
The threat from synthetic imagery is increasingly well understood by policymakers



"Deepfakes and the misuse of synthetic content pose a clear, present, and evolving threat to the public across national security, law enforcement, financial, and societal domains."

US Dept of Homeland Security, 2022

Genuine User Verification



Genuine user

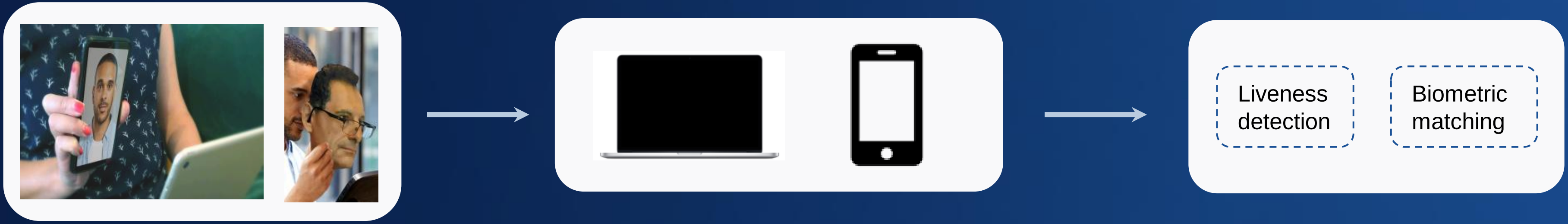
User device

Server

Liveness detection

Biometric matching

Presentation Attack



Liveness detection

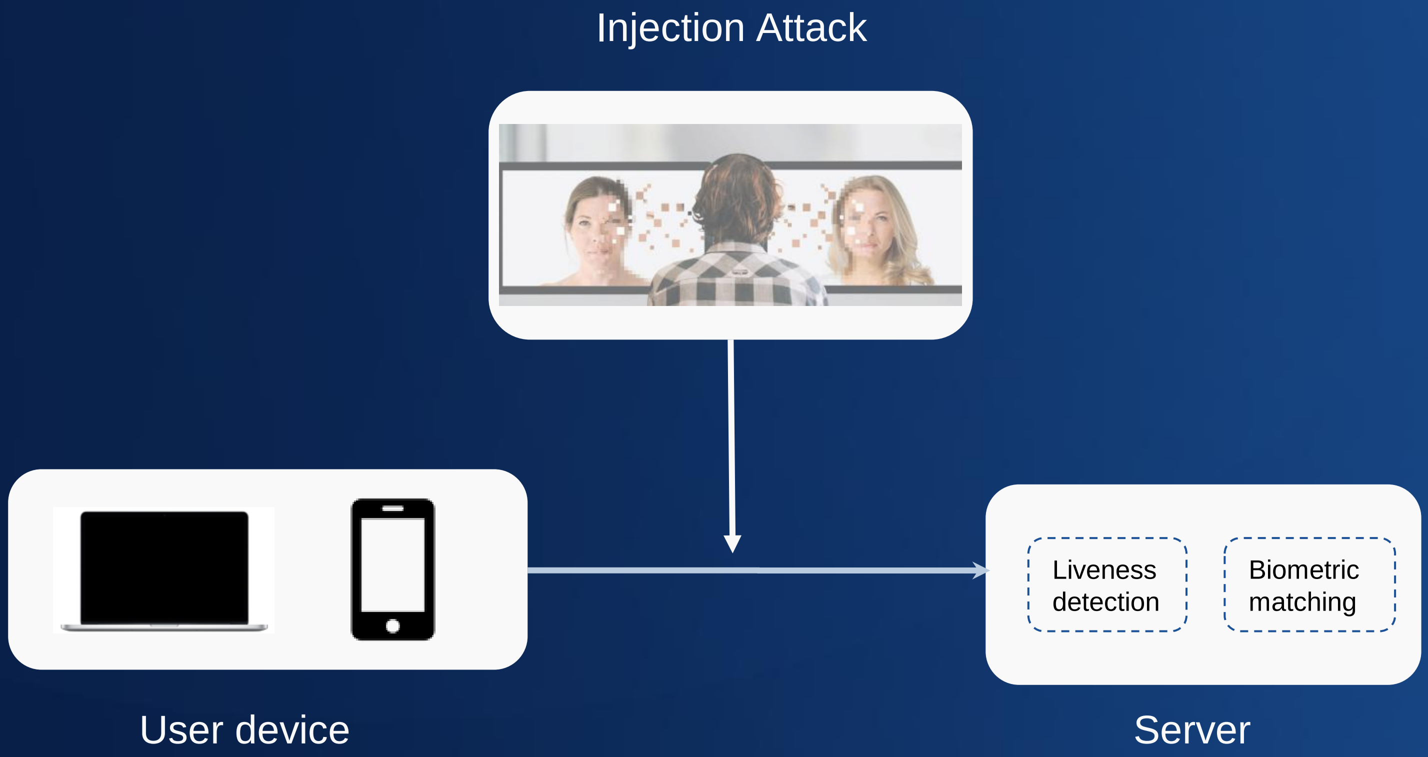
Biometric matching

Presentation Attack

User device

Server

An Injection Attack



Injection Attack

Liveness detection

Biometric matching

User device

Server

Presentation Attacks vs Digital Injection Attacks

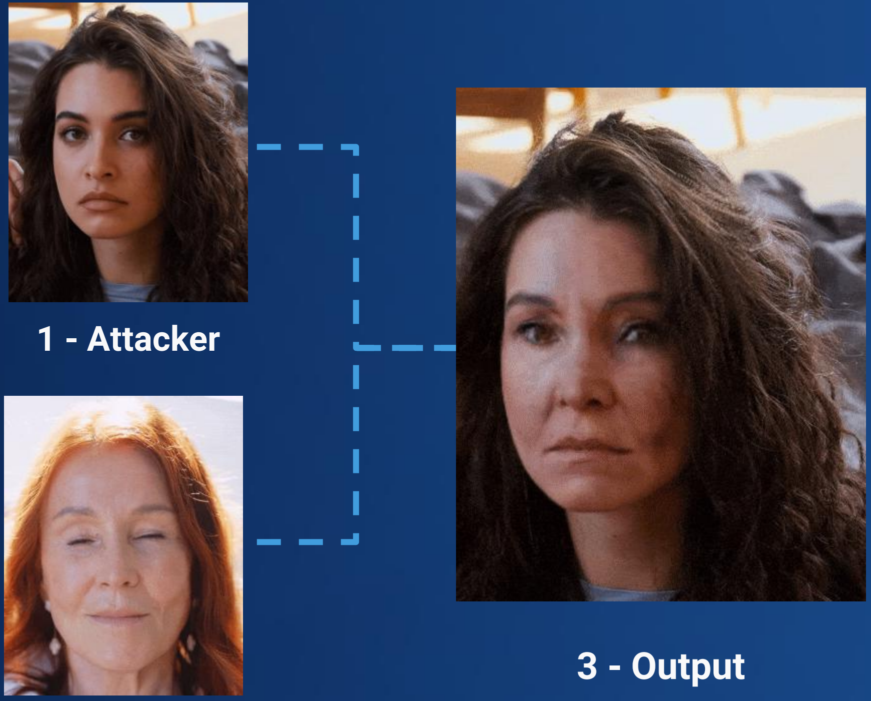
	Presentation Attacks
Deployment:	An artifact is presented to the camera
Detection:	Detected via clues in the imagery
Scalability:	Limited in scale. Threat actors can deploy one attack at a time
Testing:	Existing globally recognized standards for Presentation Attack Detection (ISO/IEC 30107)

Presentation Attacks vs Digital Injection Attacks

	Presentation Attacks	Digital Injection Attacks
Deployment:	An artifact is presented to the camera	Imagery is injected directly into the video stream
Detection:	Detected via clues in the imagery	Detected either via analyzing metadata or imagery-based testing
Scalability:	Limited in scale. Threat actors can deploy one attack at a time	Unlimited in scale. Threat actors can create highly automated attack machines
Testing:	Existing globally recognized standards for Presentation Attack Detection (ISO/IEC 30107)	No existing globally recognized standards for Digital Injection Attack Detection

Face Swap Attacks

Face swaps are a type of deepfake, created from two inputs. A new identity is superimposed over an existing video or live stream in real-time



1 - Attacker

2 - Target

3 - Output

Injection Attacks: Why They Matter

Prevalence

- Injection attacks are a present threat (5x PA rate on web)
- They now present a threat to all platforms (149% increase H1->H2 2022 on mobile web, Android and iOS)
- Injection attacks the primary route for persistent threat actors

Evolution

- Rapid of evolution of synthetic imagery methods (currently tracking >80 tools for faceswaps alone)
- Increased availability of injection and combined tools
- Example (295% increase in faceswap injection attacks H1->H2 2022)

Scalability

- Injection attacks can be launched by attack machines which can be fully automated
- Enables threat actors to explore areas of the threat landscape with minimal marginal cost per identity
- Current observation of bursts of IAs (00s or 000s) over short periods

Video injection attacks present a current threat which is highly scalable and evolving rapidly

Injection Attack Mitigations: High Level Approaches

Meta-data based

- Detect whether an injection has occurred
- Reliant on information that comes from the device
  - relies on obfuscation of the device code
- can be perfectly forged

Injection Attack Mitigations: High Level Approaches

Meta-data based

- Detect whether an injection has occurred
- Reliant on information that comes from the device
    - relies on obfuscation of the device code
- can be perfectly forged

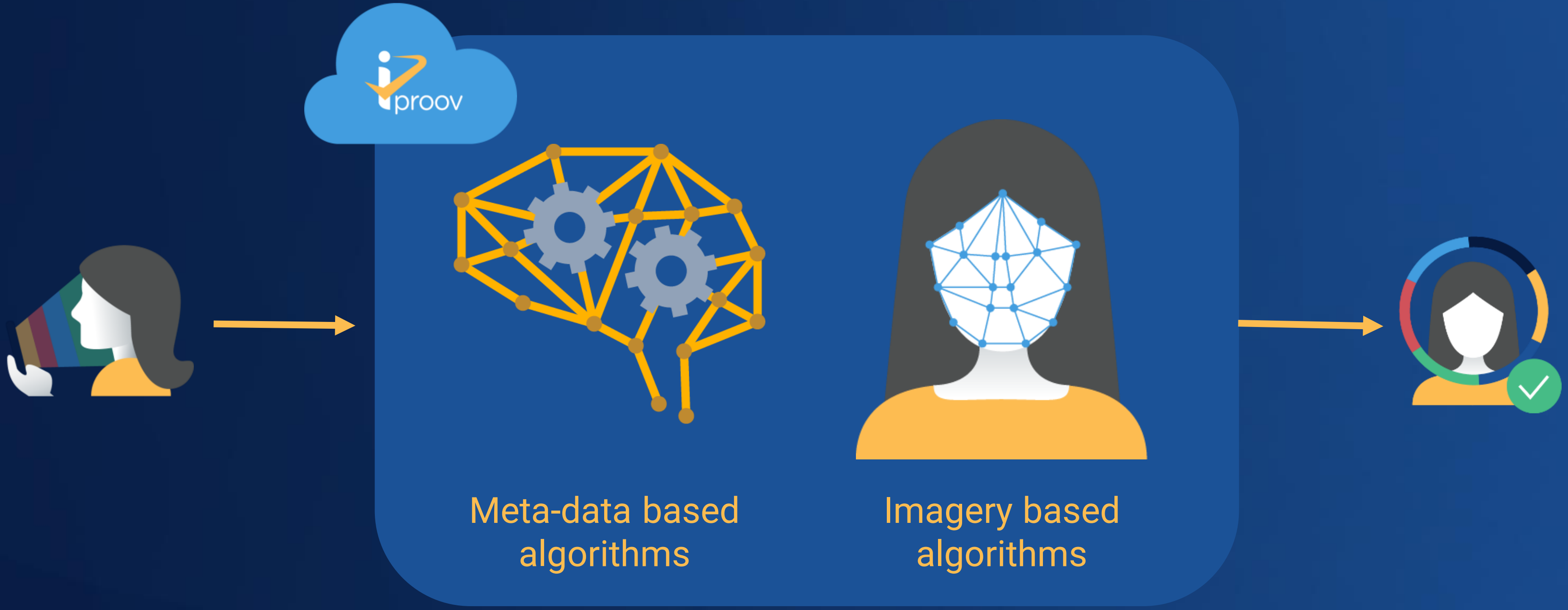
Imagery-based

- Determine whether the imagery comes from a bonafide user
- Detection of synthetic imagery
    - hard to synthesise
    - not repeatable
    - high usability
- cannot be perfectly forged

Applicability of Approaches

	No user action required	User action required
Repeated biometric signal	E.g. single frame, passive video **Meta data approaches only**	E.g. user blinking, head turning **Meta data approaches only (replay attack)**
One-time biometric signal	E.g. controlled illumination **Meta data approaches and Imagery-based approaches**	E.g. user reading words, numbers, sequences of actions **Meta data approaches and Imagery-based approaches**

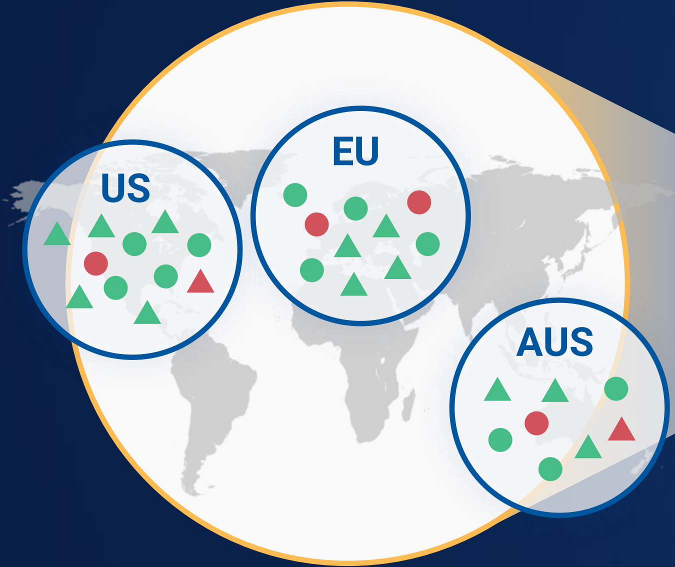
iProov Approach to Attack Mitigation



Meta-data based algorithms

Imagery based algorithms

iSOC: Sourcing Biometric Threat Intelligence

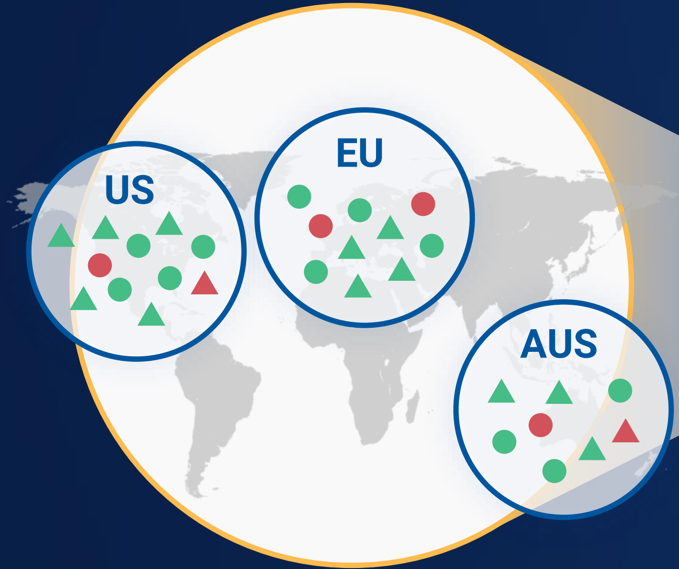


Multiple platforms across multiple geographies

iProov's global real-time threat intelligence system - iSOC

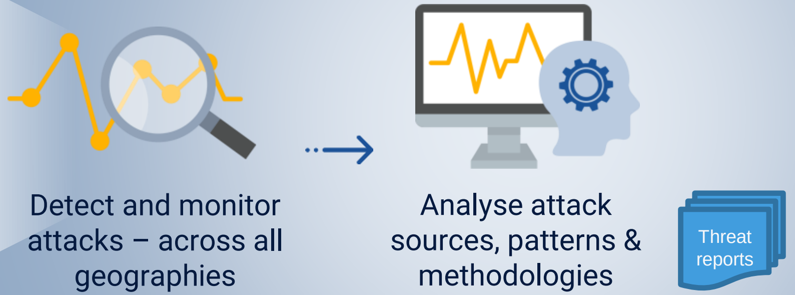
Detect and monitor attacks – across all geographies

iSOC: Sourcing Biometric Threat Intelligence



Multiple platforms across multiple geographies

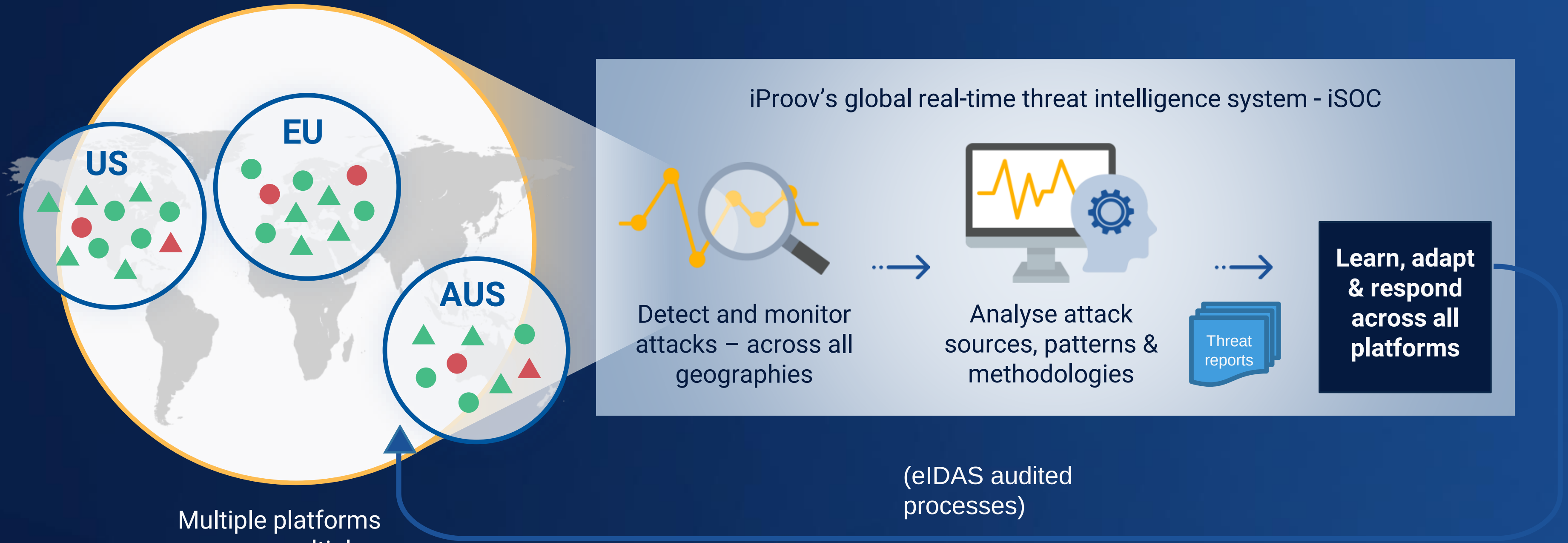
iProov's global real-time threat intelligence system - iSOC



Detect and monitor attacks – across all geographies

Analyse attack sources, patterns & methodologies

iSOC: Sourcing Biometric Threat Intelligence



iProov's global real-time threat intelligence system - iSOC

Detect and monitor attacks – across all geographies

Analyse attack sources, patterns & methodologies

Learn, adapt & respond across all platforms

(eIDAS audited processes)

Multiple platforms across multiple geographies

Thank you

Genuine Presence Assurance

Right person, Real person, Right now

contact@iproov.com