Discussion Group Request

(1) DG name

Injection Attack Discussion Group (IA DG)

(2) Discussion topic, purpose, and motivation

In biometric data injection attacks, an attacker bypasses the camera and feeds manipulated data, such as a deepfake or a replayed video, directly into a remote identity verification system. The response to injection attacks is recent and is moving on several fronts at once.

Layer	Work	Status
Definitions and taxonomy	CEN/TS 18099	Published 2024
International standard	ISO/IEC 25456 (SC 27 and SC 37)	In development
National evaluation scheme	BSI evaluation guidance under TR-03147	Draft, in public comment
Commercial evaluation	Independent laboratory programs	Operating and issuing results
Assurance frameworks	eIDAS LoA, NIST SP 800-63-4, ETSI TS 119 461, Kantara IAL	Established, not yet connected to the layers above

Each of these efforts is sound within its own scope, and the technical questions of test methodology, coverage and metrics belong with the standards and scheme bodies and will improve through their own processes. What remains unassigned is the work of connecting them.

- There is no shared way of talking about injection attacks and injection attack detection. A relying party, and often anyone outside the standards community, cannot say what CEN/TS 18099 covers, how an injection attack differs from a deepfake, or where and how injection happens in a verification flow. Within the technical community the terms exist, but they sit across several documents, and there is a lack of a single framework to work from.
- A relying party, an assessor or a government program receiving evaluation reports from two or three of these schemes has no way to tell what each result proves, what it leaves out, or how long it remains meaningful. NIST SP 800-63-4, for example, requires injection attack detection at IAL2 without naming any evidence that would satisfy it.
- Injection tooling and generation techniques change within months, and every vendor sees only the attacks aimed at its own deployments. There is no shared picture of the threat and no defined way for the industry to learn from attacks collectively.

What the DG will work on

The group would bring the standards bodies, laboratories, vendors and public and private sector relying parties together around four priorities, beginning with the first two, which the others build on.

- **Define injection attacks** and establish a common framework for discussing them. What they are, how they work, where in a verification flow they occur, and why and when they are used. This aggregates what CEN, ISO, BSI and others have established into one picture, useful to a Kantara audience and to specialists who currently have to assemble it from several documents. It builds on the earlier deepfake work in the IDV discussion group.
- **Define injection attack detection** and produce a condensed reference to the standards and test and evaluation frameworks, for example how CEN/TS 18099 and ISO/IEC 25456 relate to one another. What each one covers, what it requires, and where they differ. Not a judgement of any of them, a description.
- Moving from the frameworks to the results they produce, **understand evaluations and create a framework for the interpretation of results**. What a given result demonstrates, what scope it covers, how long it stays valid, and what someone outside a scheme needs in order to make a decision from several results at once.
- **Create a framework for group-wide and industry-wide communication**. How might data sharing work, how do we balance security and transparency, and how do we join forces while respecting IP. The group would come out with a position on each of these. Similar problems have been solved before in other fields, including anti-malware testing standards, maintained attack taxonomies, mutual recognition of security certificates, and governed threat intelligence sharing. The DG will look at which of these models carry over.

Outcome and path to a Working Group

The first two priorities are the group's concrete deliverables, a Kantara document on what injection attacks are and how detection is defined, together with the condensed reference to the standards and evaluation frameworks. The third and fourth can run in parallel once that foundation exists, and the group would deliver a position on each.

The DG would close, within six to nine months, with a recommendation to the Leadership Council on whether a Working Group should follow and what it should own. One possibility worth testing is a Working Group charged with the continuous identification and communication of injection attack threats, protections and standards, given how quickly these threats move with generative AI and how much they resemble cybersecurity problems.

This work relates directly to Kantara's purpose of developing and operating trust frameworks for digital identity.

(3) Leadership

The proposed chair is Neda Eskandari, Paravision, subject to confirmation by a vote of DG participants. A co-chair from a relying party or an evaluation laboratory is encouraged.

(4) Related work and liaisons

Within Kantara, the Biometric Data DG, which covers biometric data broadly, and the Digital Identity Risk Universe DG, where injection attacks are one risk among many.

(5) Audience

Relying parties in financial services, government identity programs and healthcare; biometric evaluation laboratories; identity verification and biometrics vendors; standards and scheme owners; privacy and legal practitioners; researchers.

(6) Copyright policy

Copyright: Creative Commons Attribution-Share Alike (CC A-SA) 3.0, per Appendix D of the Kantara Initiative IPR Policies (v2.0).

(7) Proposers

Paravision.