

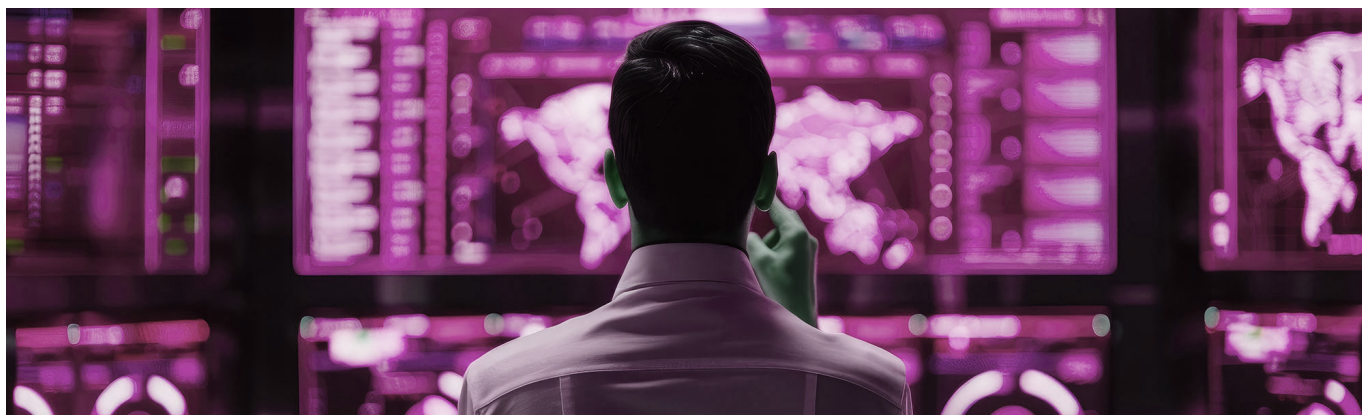


GANNDALF

NextGen Collaborative Cybercrime Investigation Framework

3rd Newsletter

Building the EU's NextGen investigation framework against advanced forms of cyber threats and cyber-dependent crimes.



GANNDALF is an EU-funded Horizon Europe R&D project designed to empower Law Enforcement Agencies (LEAs) and the wider judicial ecosystem in detecting, investigating, and preventing advanced cybercrime. The project develops a secure, modular collaboration framework with privacy-aware data sharing, AI-powered decision support tools, and a customizable sandbox for collaborative investigation.

Beyond technology, GANNDALF promotes policy innovation, professional training, and citizen engagement through its Cyber Hygiene 2.0 approach, helping society recognise, report, and resist cyber threats more effectively.

[Learn more](#)

Use Cases

Cybercrime Investigation & Prevention

AI-supported tools for detecting and investigating cybercrime, financial and investment fraud, and related threats.

Multi-Stakeholder Collaboration

Secure data sharing and coordinated investigations between LEAs, prosecutors, and other relevant stakeholders.

Cyber Hygiene 2.0

Advanced training and awareness tools to strengthen cybersecurity practices among professionals and citizens.



GANNDALF Technology Highlights

GANNDALF Portal, DSS and Sandbox

Unified environment for case management, tool deployment, data analysis, and collaborative investigations.

Integrated SIEM

Centralised monitoring and alerting capabilities to support cyberattack detection and incident response.

AI Scenario Builder

Uses LLMs and predictive analytics to generate realistic training and policy scenarios.

Crime Psychology Tools

Behavioural analysis tools supporting threat assessment, profiling, and investigative decision-making.

Human-Swarm Teaming

Multi-agent collaboration approach designed to support and accelerate complex investigations.

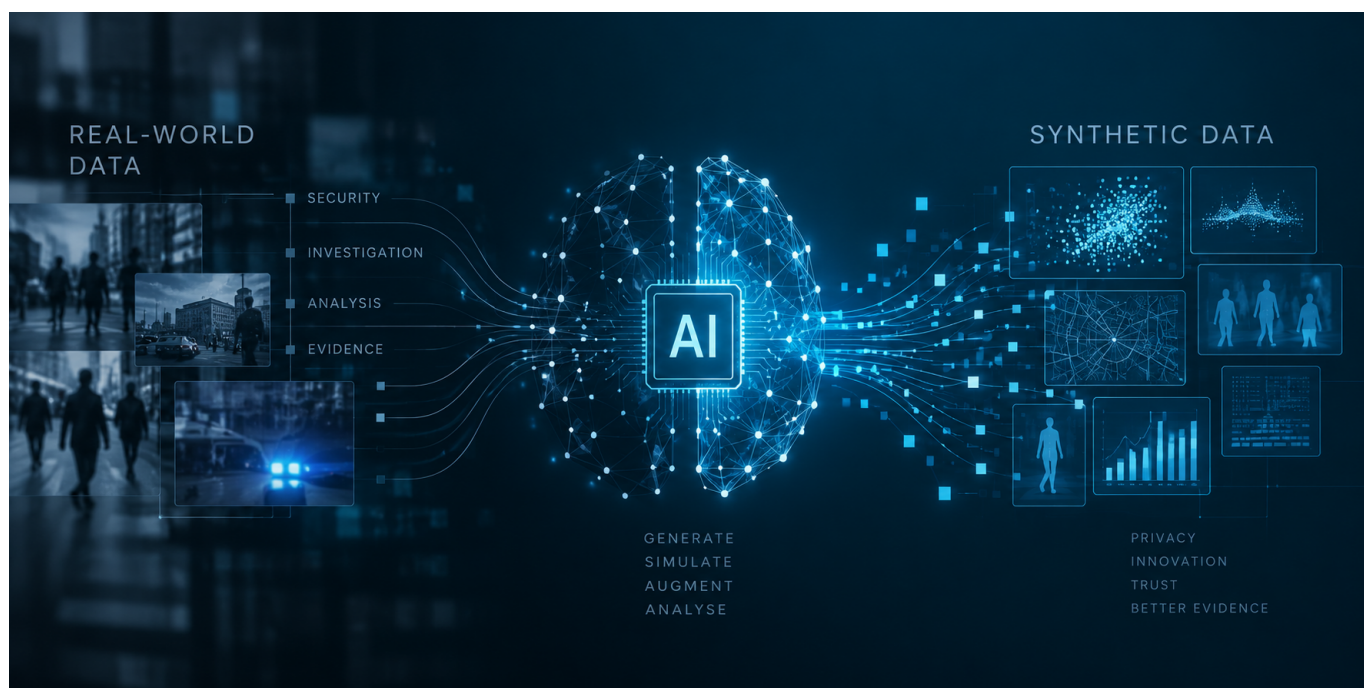
Privacy-Enhancing Technologies (PETs)

Privacy and security solutions supporting data protection and secure information sharing.

Research & Insight

SYNTHETIC DATA IN RESEARCH ON INVESTIGATIVE TECHNOLOGIES: WHY PROJECTS GENERATE THEIR OWN DATA

By Pablo Trigo Kramcsák — Vrije Universiteit Brussel (VUB)



Research projects developing investigative technologies face a persistent problem: the data that would make evaluation meaningful is the data least available for it. Law enforcement-related data sit with law enforcement agencies and stay there. Published performance figures therefore rest on datasets external parties cannot inspect, and competing claims cannot be assessed against a common benchmark.

Synthetic data are presented as the response. A workable rule for research design in this field states an order of preference rather than a general judgment about the merits of generated material. Where a research objective can be met with synthetic data, synthetic data should be used, and recourse to real personal data should be justified. It does not assert that generated material is adequate for every purpose, and it leaves the question of adequacy to be answered case by case.

The EU digital rulebook supplies grounds for the allocation. Necessity under Article 6 GDPR requires that the objective pursued cannot reasonably be achieved as effectively through means that are less intrusive of the rights concerned (see also Recital 39). Article 4a(1)(a) of the AI Act permits providers of high-risk systems to process special categories of personal data for bias detection and correction only where that purpose cannot be effectively fulfilled through the processing of other data, including synthetic or anonymised data. On that reading, the availability of an adequate synthetic alternative forms part of the assessment of whether processing real personal data is necessary at all.

A project that processes operational records without having asked whether generated material would have served will struggle to defend its necessity assessment after the fact. Accountability expects that assessment on the record.

What counts as synthetic

Synthetic data can be understood as artificially generated data with analytical value that resembles real-world data. Statisticians proposed releasing artificial records for disclosure control decades ago. Rubin suggested replacing confidential microdata altogether with values drawn from a model fitted to the original file, and Little worked out the partial version of the same idea, retaining genuine values while imputing only the sensitive ones. Both treated artificial records as a way of reconciling access to useful statistics with the protection of confidentiality (Little added that masking alters the properties of the data, so an analyst who ignores the masking mechanism may draw inferences the released file does not support). What has changed since is the scale of the operation, the range of purposes for which synthetic data can be generated, and the ambition attached to it.

Gal and Lynskey describe a data-generation revolution that unsettles a governance regime designed with collected data in mind. They offer a typology of synthetic data, based on generation methods: methods that transform collected data, methods that reduce the need for it, and methods that dispense with any direct use of it.

In Cofone, Strandburg and Tilmes's reading, that criterion, the degree of dependence on collected data, captures only part of what matters. They keep the three categories but sort them by how heavily each leans on assumptions about the ground truth the data purport to represent. Transformation trades fidelity for confidentiality, moving the file away from the world rather than closer to it. Augmentation aims at the opposite, though whether it succeeds turns on the validity of the assumptions guiding the gap-filling. Simulation rests almost wholly on background knowledge, and it either outperforms anything realistically collectable or merely encodes the modeller's assumptions with unwarranted confidence.

The differences carry legal weight in two directions. Exposure to data protection law thins across the categories, since transformation arguably carries a file beyond instruments that reach only identifiable information, and augmented and simulated data, derived from no particular person's record, are less likely still to fall within scope. The ranking inverts for group harms. What is safest against reidentification remains most exposed to unreliable assumptions, and so most apt to support inferences that are at once intrusive and wrong.

EU law uses the term without defining it. The AI Act makes synthetic data an alternative that must be exhausted before special categories of personal data may be processed for bias detection, and imposes a comparable condition on further processing within regulatory sandboxes. The Data Governance Act lists it in a recital among the techniques that permit more privacy-friendly analysis, and the EHDS invokes it in the architecture for secondary use. No instrument says what the term covers. They treat it as a single thing, which the taxonomy suggests it is not.

Why research reaches for generated data

Research on law enforcement technologies reaches for generated data more readily than other fields do, for a range of reasons.

First, some categories of investigative material cannot lawfully reach researchers at all, and generation supplies a substitute where no lawful route exists.

Where a route does exist, the applicable regime complicates it. Investigative records held by competent authorities fall under the Law Enforcement Directive, which categorises data subjects according to their position in an investigation and separates factual data from assessments. The GDPR knows neither requirement, and a research consortium is not a competent authority.

There is also the question of verification. The Directive permits restrictions on data subject rights and channels their exercise through the supervisory authority, so individuals rarely test these systems themselves and scrutiny

has to come from outside. A synthetic benchmark can be published, cited and re-run, which lets someone other than the deploying agency check claims about detection performance.

Class imbalance matters too. Serious incidents are rare, so the classes that matter most have the least training material. A classifier that performs well on a balanced evaluation set may still produce false positives in large absolute numbers at operational prevalence. Generation covers scenarios that operational records scarcely contain, but it cannot supply prevalence.

Generation has accordingly become a starting assumption rather than a fallback. The order of preference holds only if generated material really falls outside the regime it avoids. Part two asks whether it does.

This is the first of two posts on synthetic data in research on investigative technologies. The second, on the data protection status of generated material and its limits, follows in the next issue.

New GANNDALF Scientific Publication

BALANCING EXPLAINABILITY STRENGTH AND PRIVACY IN COUNTERFACTUAL EXPLANATIONS FOR INSIDER THREAT DETECTION

By Naila Azam, George Loukas and Manos Panaousis — Centre for Sustainable Cyber Security, University of Greenwich

The paper explores the trade-off between **AI explainability and privacy preservation** in the context of insider threat detection. It investigates how counterfactual explanations can support the transparency of AI-based decision-making while reducing the risk of disclosing sensitive information, contributing to the development of trustworthy and privacy-aware cybersecurity solutions. These topics are closely aligned with GANNDALF's objective of developing explainable and trustworthy AI technologies for cybercrime prevention and investigation.

[Read more](#)

Dissemination highlights

INCYBER Forum Europe 2026

April 2026



GANNDALF was presented at **INCYBER Forum Europe 2026** by **PLV**, showcasing how AI-driven tools can help Law Enforcement Agencies detect complex cyber threats, connect fragmented data, and protect critical infrastructures. The project was also promoted at the GANNDALF stand, where consortium members engaged with security professionals, policymakers, researchers and industry representatives, strengthening collaboration opportunities and expanding the project's European cybersecurity ecosystem.

[Read more](#)

Joint Stakeholder Workshop

April 2026

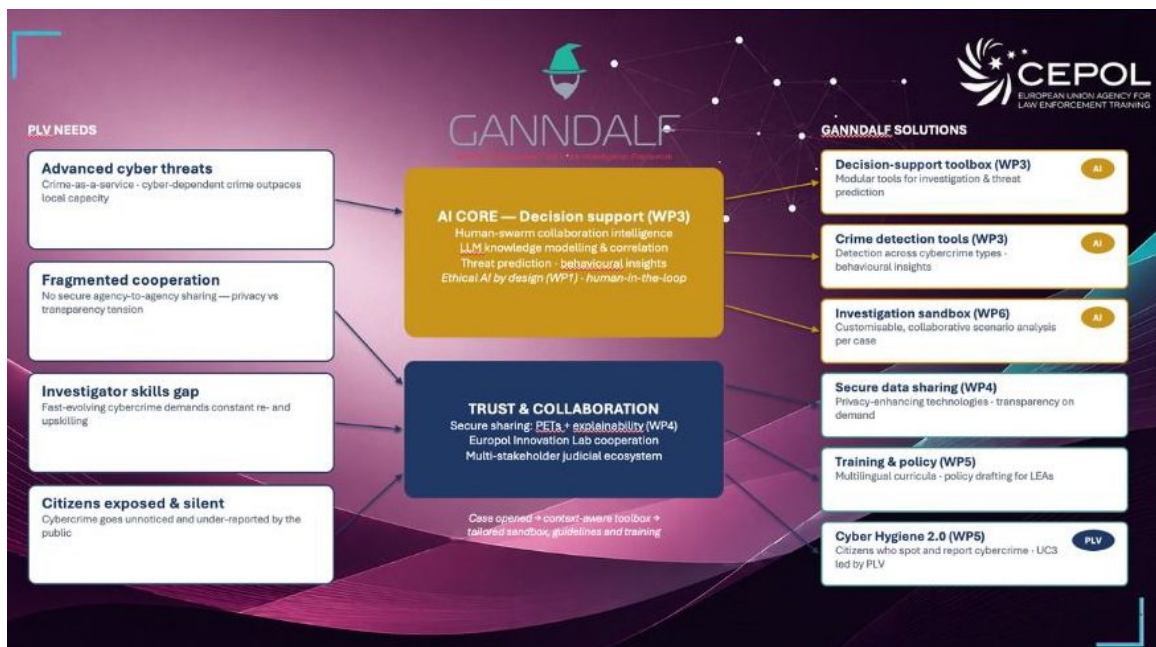
CONV, represented by Marina Natsia, participated in the **“Beyond Tools: Governing AI in Law Enforcement Investigations”** workshop at the University of Bari, co-organised by **PRESERVE, ENSEMBLE, SafeHorizon and the LEA Projects Cluster**. The workshop focused on **AI governance, ethics, operational continuity and organisational readiness in law enforcement**, contributing to a joint position paper and strengthening collaboration across European security projects.

GANNDALF at CEPOL Workshop

July 2026



PLV presented GANNDALF at the **CEPOL workshop “Artificial Intelligence, Online Threats and European Innovation Projects”**, hosted by the **German Police University (DHPol)** in Münster. The presentation showcased GANNDALF's vision for an integrated framework supporting law enforcement in preventing, investigating and cooperating across borders, highlighting the project's contribution to the future of AI-driven policing and cybersecurity.



[Read more](#)

GANNDALF Training Opportunities

GANNDALF Training in Lisbon

The GANNDALF training “**Social Engineering and Phishing Detection**” will take place on **6 October 2026 in Lisbon, Portugal**. The one-day training will bring together **Law Enforcement Authorities, prosecutors and judicial actors** to strengthen their capabilities in detecting and investigating social engineering and phishing attacks. The programme will combine practical, scenario-based exercises with a demonstration of **GANNDALF tools**, covering phishing detection, incident investigation and response, and prevention-oriented public awareness. The training is part of GANNDALF's efforts to support the **re-skilling and up-skilling of law enforcement and justice professionals** in the fight against cybercrime.



[Read more](#)

[LinkedIn](#) | [X](#)



This project has received funding from the European Union's Horizon Europe research and innovation programme under the Grant Agreement No. 101167951. The website reflects only the view of the author(s) and the Commission is not responsible for any use that may be made of the information it contains.



Copyright (C) 2026 GANNDALF Project. All rights reserved.