

AI Forensics Annual Report

2025

Table of Contents

Welcome from the Director	03
<hr/>	
About us	06
<hr/>	
Who We Are & What We Do	07
They speak for us	09
Our achievements	11
<hr/>	
Agile Investigations: Evidence That Triggers Action	11
Collaborations: Leveraging Collective Expertise	15
Capturing Emerging Risks and Shaping AI Governance	21
Fundraising Overview	27
Thank you!	28
<hr/>	

Welcome from the Director

When a global scandal unfolded on X over the holiday season, with Grok used to generate thousands of non-consensual intimate images, [AI Forensics collected and analyzed the evidence before regulators had even returned to their desks.](#)

Within days, regulators from all around the world had contacted us. X subsequently banned the functionality, and the case triggered the European Commission to open a new DSA investigation.

The Grok crisis was a powerful illustration of how we operate: detect emerging harms early, collect rigorous evidence at scale, and deliver it to those with the power to act on it.

But this was also just one of several major investigations this year. In 2025, our investigations spanned a growing range of platforms and issues, including documenting the surge of AI-generated content and agentic forms of manipulations on social media, monitoring the integrity

of key elections, scrutinizing the deployment of age verification tools, holding platforms accountable for pornographic content, or exposing large networks of scam advertisements and illegal targeting practices.

[We focus on harms that are of widespread public concern and driven by the AI systems that shape how information flows.](#) We then act where regulatory frameworks provide a pathway for our evidence to drive change.

[This year crystalized our position as a leading technical partner in platform oversight, particularly in the EU, and increasingly beyond.](#) Our first large long-term contracts with the European Commission under the DSA formalized our role as evidence provider that our investigations had long established. Our collaborations with civil society follow the same logic. We bring data collection capacity, while our partners bring contextual expertise and advocacy reach, transforming evidence into impactful accountability.

Welcome from the Director

Meanwhile, AI is opening up entirely new avenues for platform accountability research. [We are pioneering novel AI-assisted approaches that unlock unprecedented scale and granularity in our content analysis.](#) This builds on our technical DNA and on our ambition to scale our work and impact.

While these developments are promising, [it's important to acknowledge the complex and challenging nature of the environment in which we operate.](#) Organizations working on platform accountability face growing pressures including lawsuits aimed at silencing oversight, harassment and intimidation, with real consequences at both the personal and organizational levels.

We are proactive about strengthening our resilience, but doing so also diverts resources from our core mission.

Governments must urgently step up to ensure that civil society has the legal protection and resources necessary to protect the public interest.

At the same time, the philanthropic landscape is also under growing strain. We have sustained our growth in large part thanks to institutional contracts. Although the growing relevance and demand for our work provides us with confidence, a critical funding gap remains, [which we can only fill with new and renewed philanthropic support.](#)

Looking ahead, 2026 will be a decisive year. The challenges facing digital rights, information integrity, and democratic governance today are global. They are compounded by the growing concentration of power over information infrastructure in the hands of a few actors, as well as by mounting pressure from political interests. Europe will be a key testing ground for whether regulatory ambition can deliver meaningful protection.

[The first major DSA enforcement decisions](#), many of which we are directly involved with, will set the enforcement landscape for years to come.

Welcome from the Director

Beyond regulation, there is an urgent need to invest in alternative digital ecosystems that entrench digital rights and sovereignty.

This also includes advancing new legislations such as a right to algorithmic pluralism, which would put citizens, not platforms, at the center of how information flows.

In this landscape, the role of an independent, technically rigorous organization that can collect evidence at speed and scale to hold the most powerful platforms to account is not a luxury, but a necessity.

It is what makes information integrity, free expression, and digital sovereignty possible.

That is the unique role that AI Forensics plays, and we will continue to drive this work with the urgency and conviction it demands.



Marc Faddoul
Director,
AI Forensics

Marc Faddoul at Politico AI Tech Summit 2025

About Us

Who We Are & What We Do

AI Forensics is a European non-profit that audits influential and opaque algorithms.

It was established to address a critical gap in civil society: [providing technical expertise and evidence-gathering capabilities for algorithmic auditing](#). In just four years, we have emerged as the leading European actor in this field.

In other words, we hold major technology platforms accountable [by conducting independent, high-impact data-driven investigations](#) into the AI and algorithmic systems which shape how information flows.

We develop unique evidence-collection tools and methodologies to expose digital rights violations through technical reports, which we combine with our multidisciplinary expertise to inform regulators and policymakers.



Who We Are & What We Do

01. Data Collection

Develop Tools for Algorithmic Auditing

Collect Evidence

Our audits are always data-driven, but the data we need is typically not made available by the platforms. Our tools, therefore, allow us to mimic user behaviors, and collect behavioral data on the algorithm, using adversarial methods such as sock-puppeting, scraping, or user data donation. We also leverage the researcher data access mechanisms that the DSA requires of platforms.

02. Research

Conduct Investigations

Analyse & Interpret

Our research and data analysis methodologies are informed by the socio-technical expertise of our trans-disciplinary team, spanning computer and data science, social psychology, tech policy and ethics, and media studies, as well as by partners with complementary domain expertise.

03. Dissemination

Expose Violations & Inform Regulation

Testify

Our investigations often receive high-profile press coverage, which raise public pressure and awareness. We also provide insights and actionable recommendations to inform and support policymakers, regulation enforcement, and advocacy efforts. We regularly provide expertise through congressional testimonies, academic conferences, and other public forums.

They speak for us

Our impact is best seen through the eyes of those who support and collaborate with us:



AI Forensics make a critical contribution to Europe's AI accountability community. Their **unique combination of technical expertise and smart, timely policy recommendations has been uniquely effective** in making sure that Europe's tech regulation can be used to hold the biggest tech companies responsible for the impacts of their products on people and society



Catherine Miller

Co-Director of the EU AI & Society Fund, on AI Forensics' role in Europe's AI accountability ecosystem

European
**Artificial Intelligence
& Society Fund**

They speak for us

Our impact is best seen through the eyes of those who support and collaborate with us:



I consider AI Forensics as an essential actor in the tech accountability field. *Their technical expertise, coupled with a smart approach to advocacy and communication is what we need to counter the propaganda and deceitful claims Big Tech is spreading.* Moreover, I particularly appreciate how they keep a European and global perspective on tech issues, but at the same time are able to zoom in on specific geographies and issues, from election integrity to gender-based violence. Beyond their work in France, that Civitates directly supports, our cohort of national-level organizations can count on a generous and collaborative partner to better understand how social media and AI systems are having an impact on democracies.G



Claudio Cesarano

Senior Programme Manager for Tech and Democracy at Civitates, reflects on AI Forensics' impact



Agile Investigations: Evidence That Triggers Action

AI Forensics provides the independent evidence and technical expertise that make accountability possible. We respond rapidly to emerging harms, documenting what platforms do, not what they claim, while delivering court-admissible evidence to regulators. Our work informs policymakers, supports legal cases, and reaches the public through an extensive media network.



Exposing Grok AI's Unsafe Content Generation

Between Christmas and New Year's Eve 2025, a global scandal unfolded on X. The platform's integrated AI chatbot, Grok, was being used to digitally undress thousands of women without their consent, including some minors. By [January 5th, 2026](#), as regulators around the world returned from holiday, [AI Forensics had already documented and analyzed the evidence.](#)

Our investigation produced the first [quantitative analysis](#) of the phenomenon, examining over [20,000 Grok-generated images](#) and [50,000 user requests](#). Key findings included:

- [53%](#) of images depicted individuals in minimal attire, [81%](#) of whom were women
- [2%](#) appeared to show minors
- [6%](#) depicted public figures, about one-third political figures
- [Generation of Nazi and ISIS propaganda material](#)
- Ineffective [take-down mechanisms](#)

Exposing Grok AI's Unsafe Content Generation

IMPACT

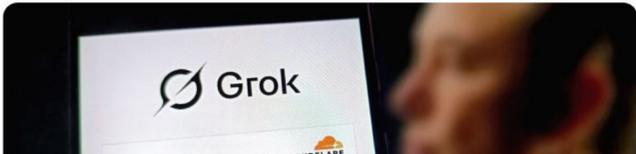
Within days, regulators from the EU, UK, France, California, and Australia contacted us, requesting our data to support their enforcement actions. Under mounting pressure, X first restricted, then banned the functionality.

Our findings have informed oversight from the UK's Ofcom to California's Attorney General, including a presentation to the European Commission which later launched a new investigation under the Digital Services Act to assess whether X had properly mitigated risks associated with Grok. This complemented the Commission's ongoing review of X's recommender systems and systemic risk management, highlighting the critical role of AI Forensics' evidence in shaping regulatory scrutiny. Investigations remain ongoing.

The Guardian

Grok AI still being used to digitally undress women and children despite suspension pledge

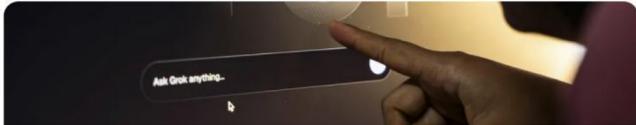
Degrading pictures being posted on Elon Musk's site despite the platform pledging to suspend people who generate them



DER SPIEGEL

Nutzer missbrauchen Grok für sexualisierte Bilder von Frauen und Kindern

Seit Wochen wird die Plattform X mit sexualisierten KI-Bildern geflutet, das Unternehmen greift kaum ein. Vor allem Frauen sind betroffen, auch Minderjährige. Eine neue Studie zeigt das Ausmaß.



Le Monde

IA Grok : la moitié des images générées pendant les fêtes représentaient des personnes partiellement dénudées

L'ONG AI Forensics a collecté les images générées par l'outil d'intelligence artificielle d'Elon Musk entre le 25 décembre et le 1^{er} janvier, période pendant laquelle des milliers d'internautes s'en sont servis pour « déshabiller » des femmes, des adolescentes et des enfants.



Bloomberg

Illegal Images Allegedly Made by Musk's Grok, Watchdog Says

L'ONG AI Forensics a collecté les images générées par l'outil d'intelligence artificielle d'Elon Musk entre le 25 décembre et le 1^{er} janvier, période pendant laquelle des milliers d'internautes s'en sont servis pour « déshabiller » des femmes, des adolescentes et des enfants.



Exposing Grok AI's Unsafe Content Generation

IMPACT

The Grok investigation generated [unprecedented global attention](#), with coverage in over [70 articles, features, and interview requests](#) across top-tier media, including [The Guardian](#), [Der Spiegel](#), [Le Monde](#), [Bloomberg](#), [Wired](#), and a live appearance on [CNN](#), where [Buse Raziye Çetin](#), Head of Policy at AI Forensics, discussed the lawsuit against X for its Grok-generated images.



Collaborations: Leveraging Collective Expertise

AI Forensics thrives on a collaborative spirit, partnering with organizations that complement our technical investigations through qualitative research, local knowledge, or policy advocacy and regulatory expertise. By combining our evidence collection and technical capacity with our partners' strengths, we produce robust data that effectively informs advocacy, policymaking, and public awareness.



Truth Be Dammed: A Year After Valencia Floods, Disinformation Persists

In line with this collaborative approach, AI Forensics partnered with [Fundación Maldita.es](#) and [ClientEarth](#) to [investigate](#) disinformation surrounding the 2024 Valencia floods (DANA). While we provided the [technical investigation and data analysis](#), our partners contributed local expertise and policy advocacy capacities, enabling a comprehensive study of platform practices and their societal impact.

The study highlights ongoing platform failures: YouTube still lacks a dedicated policy for climate disinformation, while TikTok fails to enforce its policy. This work also underscores the urgent need for [researcher access to algorithmic data](#) to study amplification and monetization practices.

This work received coverage in major Spanish media, including [La Vanguardia](#) and [El Periódico](#), reaching a broad audience and fueling public and regulatory discussion.

Truth Be Dammed: A Year After Valencia Floods, Disinformation Persists

LAVANGUARDIA

YouTube y TikTok mantienen vídeos con negacionismo climático un año después de la dana

Suman 21 millones de visualizaciones y siguen sin advertencias ni etiquetas, según Maldita.es y AI Forensics

elPeriódico

Los bulos de la DANA acumulan 21 millones de reproducciones en Youtube y TikTok sin que se tomen medidas

Una investigación de la Fundación Maldita destaca que se publicaron más

Our analysis examined all YouTube and TikTok videos related to the floods posted in the month following the event, including targeted sub-samples, and revealed the following:

- Climate dis/misinformation videos reached at least **13 million views on YouTube and 8.3 million on TikTok**
- On YouTube, dis/misinformation content averaged **21,250 views** (nearly four times the platform average), while on TikTok videos averaged **32,394 views**
- Engagement with misleading content was significantly higher than platform averages: YouTube videos were **48% more likely to be liked** and **123% more likely to be commented on**; TikTok videos were **85% more likely to be shared**
- One year later, fewer than **one in four YouTube videos** with dis/misinformation carried warnings, and TikTok videos denying climate change remained unlabelled

Truth Be Dammed: A Year After Valencia Floods, Disinformation Persists



AI Forensics is an immensely valued partner in the big tech accountability space. We have learned so much from AI Forensics' research, producing evidence that has been critical in supporting our legal analysis of platform's roles in aggravating issues like climate disinformation, and potential breach of legal obligations. **AI Forensics' technical skills and data-gathering techniques are second to none. Their deep understanding of social media platforms and how they work is unrivalled, meaning their insights are always strategic.** We are so pleased to have been connected to AI Forensics and to benefit from the research they contribute in this space.

Rosa Pritchard

Lawyer, Accountable Corporations (ClientEarth)



ClientEarth 

OBSERVATORY PARTNERS

POST-X SOCIETY

AI FORENSICS

TROLLRENSICS

UNIVERSITY OF
AMSTERDAM

JUSTICE FOR
PROSPERITY

Dutch Parliamentary Elections 2025: Monitoring Digital Threats

During the October 2025 Dutch parliamentary elections, AI Forensics [collaborated](#) with the [Hybrid Election Integrity Observatory \(HEIO\)](#), a consortium of Post-X Society, Trollrensics, University of Amsterdam, Justice for Prosperity, and other partners.

In this collaboration, AI Forensics contributed [technical investigations and systematic data collection](#), while consortium partners provided field expertise, platform-specific knowledge, and advocacy capacities. Our analysis identified several concerning patterns across social platforms:

- [Coordinated foreign engagement](#) targeting Dutch political pages
- [Amplification of far-right content](#) by cross-border networks on X
- [AI-generated anti-immigration content](#), including a song reaching the Dutch Top 50 and spreading on TikTok
- [Violent AI-generated imagery](#) targeting politicians and minorities
- Livestreams containing [death threats](#), as well as [racist and antisemitic hate speech](#)
- A spike in [election-fraud narratives](#)

Dutch Parliamentary Elections 2025: Monitoring Digital Threats

Platform safeguards fell short of addressing these threats: political ad transparency rules were applied inconsistently across platforms, and reported content violating rules was often [not removed](#). This work also exposed regulatory weaknesses at both Dutch and EU levels, highlighting the [limits of DSA enforcement powers](#) and delays in responding to data requests.



AI Forensics was a vital partner in the Hybrid Election Integrity Observatory, a consortium that monitored digital threats during the Dutch parliamentary elections of October 2025. For example, their systematic data collection on YouTube showed how algorithms distributed and amplified fraud narratives. [It provided the kind of rigorous, structured evidence that election integrity monitoring demands.](#) It directly strengthened our advocacy towards the European Commission, Dutch regulators, and Parliament. AI Forensics' experience from election monitoring across multiple European countries brought a level of methodological maturity that elevated the entire project.



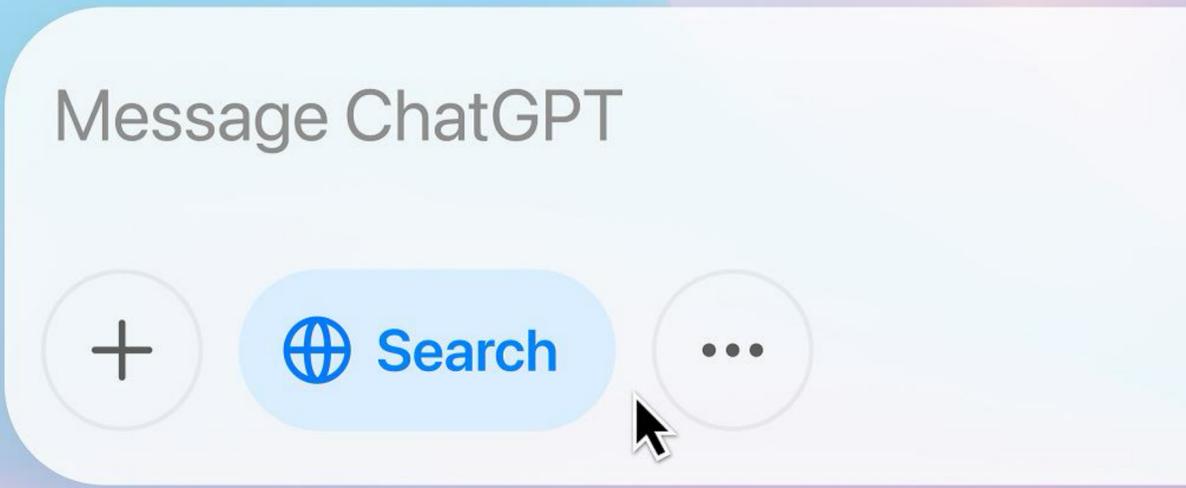
Pieter van Boheemen

Director,
Post-X Society



Capturing Emerging Risks and Shaping AI Governance

AI Forensics combines technical expertise with qualitative and policy research, allowing us to capture emerging AI-driven harms while maintaining rapid, actionable insight. Our multidisciplinary team is comprised of computation researchers, software developers, post-doctoral researchers, as well as AI ethics, policy, and governance specialists. Technical specialists map and analyze AI systems, while qualitative and policy researchers contextualize behaviors, uncover societal impacts, and support advocacy. This integrated approach ensures our investigations are both technically rigorous and socially meaningful.



Message ChatGPT

+ Search ...

From “Googling” to “Asking ChatGPT”:

Governing AI Search

CHECK OUR REPORT ! 📄

AI FORENSICS

Policy Report on AI as Search: From ‘Googling’ to ‘Asking ChatGPT’

AI search, whether carried out through chatbots or other AI-enhanced functionalities, amplifies existing risks associated with traditional search engines, including by deepening asymmetries in informational imbalances.

[Our report](#) therefore highlights how the current European regulatory landscape struggles to keep pace: the [Digital Services Act](#) focuses on ongoing operational oversight of user-generated content, while the [AI Act](#) emphasizes pre-deployment product safety. This creates a [regulatory gap](#), with AI search sitting uneasily between operational oversight and product safety requirements, leaving key risks under-addressed.

To address these gaps, we propose [anticipatory governance frameworks](#) and a set of [policy propositions](#) grounded in three case studies: [Copilot in Bing, Gemini, and ChatGPT](#). These recommendations aim to better navigate regulatory fault lines and mitigate emergent AI risks, supporting innovation that is safe and accountable

Prompt, Upload, Repeat: Agentic AI Accounts Flood TikTok

[Our investigation](#) into [Agentic AI Accounts \(AAAs\)](#) revealed how automated actors flood social media with repetitive AI-generated content to “game the algorithm,” influencing visibility and user engagement at scale. Across 354 accounts monitored over one month in 20 languages, we found that:

- Over [43,000 AI-generated](#) posts reached 4.5 billion views
- [TikTok](#) labeled less than [1.38%](#) of content as AI-generated
- Sexualized content, including depictions of minors, was prevalent in nearly one-third of accounts and half of the top posts
- Anti-immigrant narratives, false news, and other harmful trends were widespread

This work demonstrates the [technical sophistication of our team](#), combining platform-specific data collection, snowball sampling, and AI content detection, while our qualitative researchers contextualize the societal harms and policy implications. Findings from this report were also covered in major media outlets, including [The Guardian](#) and [Der Spiegel](#).

The Guardian

Anti-immigrant material among AI-generated content getting billions of views on TikTok

Researchers uncovered 354 AI-focused accounts that had accumulated 4.5bn views in a month



DER SPIEGEL

KI-Konten fluten TikTok mit Fakebildern von kindlich aussehenden Frauen

Welche Posts sind echt auf TikTok? Was ist KI? Laut einer neuen Studie verbreiten sich auf der Plattform fragwürdige Videos. Die Beiträge mit lasziv posierenden Mädchen sammelten Millionen Aufrufe.



Technical Insight Meets Social Impact

Martin Degeling

Senior Research at AI Forensics, and a post-doctoral researcher who has studied opaque algorithms for over a decade in academia and civil society, reflects:



Photo credit: Dr. Thorben Mämecke / theorie-apparate.de – Episode IoT, Kontrolle & Gewalt, featuring Martin Degeling



Being a technical researcher at AI Forensics is uniquely rewarding because we work within an interdisciplinary team. Here, technical rigor is met with deep expertise from policy, advocacy, and algorithmic auditing. This collaboration ensures our research isn't just academically interesting, it's strategically positioned for real-world impact. Whether deconstructing algorithmic bias or investigating content proliferation, being able to iterate in real-time with colleagues who understand the broader implications is invaluable. I'm proud that our technical findings serve as a foundation for meaningful interventions, actively shaping technology for the better.



Technical Insight Meets Social Impact

Silvia Semenzin

Senior Researcher at AI Forensics, and a post-doctoral researcher focusing on technology, gender, and power, shares:



Silvia Semenzin , AI Forensics



My work at AI Forensics is grounded in a shared commitment to policy-oriented research addressing urgent digital harms, including technology-facilitated gender-based violence and online misogyny.

What makes this work particularly impactful **is the close collaboration with the technical team:** their analyses of platforms and infrastructures allow us to uncover patterns that qualitative methods alone cannot fully capture.

This interdisciplinary approach strengthens our ability to understand how online harms emerge and scale, while ensuring that our research remains directly relevant for policy and advocacy. By combining qualitative insight with strong technical evidence, we are able to contribute to more effective governance responses and support those working towards safer and more equitable digital environments.



Capturing Emerging Risks and Shaping AI Governance

This unique blend of technical precision and qualitative insight extends beyond the pages of our reports **by directly informing regulatory action.** Our evidence-gathering and analyses feed into the European Commission’s DSA enforcement work, ensuring that platform oversight is grounded in robust, multidimensional research.

Building on our interdisciplinary approach, AI Forensics also contributes directly to regulatory oversight through two European Commission DSA tenders.

Lot 1 focuses on building technical capacity for evidence gathering and providing technical assistance for compliance monitoring on Very Large Online Platforms (VLOPs) and Very Large Online Search Engines (VLOSEs), including the identification of systemic risks and emerging digital threats.

Lot 2 addresses prominent and recurring risks across Member States, including harmful content, threats to civic discourse, gender-based violence, public security, electoral integrity, and the protection of minors.

Prabhat Agarwal

Head of Unit, Digital Services and Platforms, DG CONNECT, European Commission, notes:



AI Forensics supports our DSA investigations through rigorous data analysis, providing targeted quantitative evidence. Their assessment of digital platform features generates robust insights that contribute to effective DSA implementation. **This strengthens our supervisory capacity and helps ensure our interventions are grounded in solid empirical findings.**



Fundraising Overview

In 2025, our annual budget [increased by more than 50%, consistent with our growth rate in previous years](#). These expenses are primarily driven by the growing size and seniority of our team, now comprising 18 colleagues. This has allowed us to significantly expand our investigative capacity and the scale of our impact, and to better respond to the surging demand for our work.

This growth has been primarily enabled by philanthropic funding, thanks to both renewed commitments from our existing core funders as well as to partnerships with new supporters. In parallel, we have consolidated a vital and growing stream of institutional funding through contracts with the European Commission, further strengthening the regulatory impact of our research.

At the same time, [we must highlight new challenges in the philanthropic ecosystem](#). Highly visible and comparatively well-funded organizations like ours are sometimes perceived as being in less critical need of support, making it more difficult to secure additional resources.

Additionally, because of the clear impact and public value of our work, it is often assumed that it would be supported by governments. However, the type of work we conduct is novel, and very few public funding mechanisms are currently in place. Finally, as our budget grows, existing commitments cover a smaller share each year, requiring us to secure new sources of funding to fill this gap.

To address this, [we are exploring new strategies to expand our network of partners and secure long-term sustainability](#). By combining the trust of our core funders, the interest of new supporters, and the growing relevance and tangible impact of our work, we remain confident in our ability to overcome these challenges and continue scaling our work and impact.

We would like to express our [heartfelt gratitude to our funders for their unwavering support](#). Your generosity has been instrumental in advancing our mission and making a tangible difference. Thank you for partnering with us to create positive change.

OPEN SOCIETY
FOUNDATIONS

Luminate

LIMELIGHT
FOUNDATION

ON
OMIDYAR NETWORK™

SIDNfunds

civitates



European
Artificial Intelligence
& Society Fund

Thank you!

We are deeply grateful to our donors, partners, and peer organizations for your unwavering support and collaboration. Over the past year, your contributions, whether through funding, shared expertise, collaborative projects, or resources have been essential in helping us expand our reach and tackle pressing challenges. Together, we have advanced our mission, achieved meaningful milestones, and created impact that none of us could accomplish alone. Your trust and commitment continue to inspire us, and we look forward to building on this momentum in the year ahead.



Thank you!

Credits

Katsiaryna Viadziorchyk:
Report writing

Denis Constant – Ittai Studio :
Graphic & Brand Design
[Linkedin](#)

Sophia Crabbe-Field:
Editing

All other content (c) AI Forensics 2026
Email : info@aiforensics.org
Social Media : [Linkedin](#) | [Bluesky](#)

AI Forensics Annual Report

2025