Speakers (preliminary) - DeepSec IDSC 2026 Europe

Advanced Mobile Offensive Tradecraft

David Tancredi, Simone Scalco (Folposec)

Artificial Intelligence is rapidly changing the way offensive security engagements are conducted, but most practitioners are still using it as a glorified chatbot. This hands-on training integrate modern reasoning models and autonomous agents into real mobile penetration tests and Red Team operations.

This condensed session focuses on the synthetic attack techniques from the full course at Folposec, exploring how modern AI can be leveraged to execute realistic offensive opera live demonstrations and practical examples, attendees will see how autonomous social engineering, AI-assisted smishing, real-time deepfake and vishing attacks, persistent conve and AI-generated hardware payloads can be used to enhance Red Team engagements. The emphasis is on practical tradecraft, real-world tooling, and the operational challenges these techniques responsibly in modern offensive security assessments.

Topics covered

- One-Shot high fidelity Deepfakes: techniques for creating convincing synthetic identities for offensive security engagements including one-shot facial synthesis, high fidelity real-time synthetic video.
- Voice cloning, Vishing & Social Engineering: practical exploration of AI-generated and cloned voices for targeted social engineering, including voice impersonation, real time vishing scenarios and convincing pretext development.
- Agentic Smishing: implementation of autonomous AI agents for targeted smishing campaigns including automated target profiling, intel gathering, personalized message ge campaign orchestration.
- Autonomous Agentic conversational chatbots: practical implementation of autonomous conversational agents capable of conducting persistent social-engineering interactio contextual conversations, automated responses, conversational memory and long running engagement with targets.
- AI-Generated Hardware Payloads & Mobile persistence: practical intro to AI-assisted hardware payload generation against Android devices including automated HID scripts interaction, data manipulation and techniques for establishing persistence.
- Agentic Mobile SAST: introduction to autonomous AI agents for mobile application security analysis, including AI assisted static analysis, vulnerability identification, automat integration with mobile reverse engineering workflows.
- Legality & Scope of Mobile Red Teaming: introduction to the operational, legal and ethical considerations surrounding mobile-centric Red Team activities, including engagen authorization and responsible use of offensive AI techniques.

What students will receive
Students will receive all code and scripts used during the training, together with access to private repositories containing the material required to reproduce the techniques covered course. Participants will also receive access to a private Telegram channel for continuous updates on mobile offensive tradecraft, tooling and techniques.

In addition, students will receive exclusive free access to the upcoming complete course on Agentic SAST and Reverse Engineering for Mobile (online).

Entry requirements
Participants should have some experience with penetration testing and/or Red Teaming, basic Linux and CLI skills and some scripting or programming experience. The course is p to Red Teamers and pentesters interested in mobile specific targets and AI-generated synthetic attack vectors such as deepfakes, vishing and autonomous social engineering age relevant to Blue Teamers and defenders who want to understand modern TTPs used by advanced threat actors. Previous experience with AI and LLMs is useful but not required.

What students should bring
Students should bring a laptop capable of running standard security tooling, together with a suitable Linux environment or virtual machine where appropriate. For the hardware pay students may optionally bring their own compatible hardware plug or cable (HAK5), although this is not strictly required. Students should also bring an Anthropic API key, which wil AI driven practical exercises. Students are responsible for their own API usage costs. The trainers will provide a complete cloud-based environment and most of the APIs required demonstrations and exercises.

David Tancredi has spent more than a decade breaking things professionally. His background spans Red Teaming, Mobile Security, Web Applications, OT, and IIoT, with offensive engagements delivered across Europe and the APAC region. When he's not on an assessment, he's usually sharing research and tools with the community through his r3dcl1ff Yo and GitBook/GitHub. He holds a Master's Degree in Computer Science from the University of York (UK).

Simone Scalco is an Offensive Security Team Lead with a passion for Mobile Security, Web Application Testing, Malware Analysis, and Exploit Development. Outside of consulting AI/ML researcher with publications at leading international conferences, focusing on the intersection of artificial intelligence and offensive security. He holds Bachelor's and Master' Computer Science with a specialization in AI and Data Science.

Together, David and Simone enjoy exploring how AI is reshaping offensive security. Their training blends practical Red Team experience, mobile hacking, and plenty of real world l help attendees build workflows they can immediately take back to their own engagements.

AI Security Laboratory: Hands-On + Full-Stack (Lifetime Lab Access)

Dawid Czagan (Silesia Security Lab)

Overview

AI has introduced an entirely new layer of security risk — one that needs to be understood from both attacker and builder perspectives. This training is a hands-on, full-stack guide landscape, showing how modern AI systems are attacked, built, and used in real-world security.

You will work through the offensive side of AI security with prompt injection, jailbreaking, and fuzzing of LLM applications. You will also apply AI in security engineering and daily op AI building blocks into practical workflows and use AI security tools for everyday tasks.

Along the way, you will move into more advanced capabilities — building agentic AI for real-time security operations, applying AI to vulnerability research and PoC development, a smarter AI can assess other AI.

The training includes hands-on exercises, reusable Python scripts, and lifetime lab access — so you can continue practicing and applying what you learned long after the class en

Key Learning Objectives

- prompt injection: direct and indirect
- LLM jailbreaking
- fuzzing LLM applications
- AI-powered shell
- advanced prompting
- local LLMs / private AI
- AI programming
- AI attack detection
- OpenAI models and API
- embeddings
- quantization
- LLM Guard
- building agentic AI
- creating your own prediction model
- CVE research / PoC development with AI
- smarter AI assessing other AI
- specialized AI security tools
- and more …

Topics Covered

1) AI/LLM attack vectors, including various forms of prompt injection and LLM jailbreaking techniques

2) AI programming for security practitioners — working with local LLMs / private AI and cloud models (OpenAI / API), and building AI workflows for security use cases (e.g. fully priv

3) AI attack detection, including the use of local LLMs, anonymizing data before sending it to cloud LLM providers, and applying open-source defenses such as LLM Guard

4) Fuzzing LLM applications, which differs from traditional fuzzing due to the non-deterministic nature of modern LLMs

5) Using AI in security practitioners' daily operations — including AI-powered shell, advanced prompting, and AI security tools

6) Ready-to-use Python scripts, providing hands-on experience and reusable AI building blocks for daily security tasks

7) Smarter AI assessing other AI, along with interesting AI techniques and projects for security practitioners

8) CVE research and PoC development with AI

9) Building agentic AI for real-time security operations

10) and more …

What Students Will Receive

Students will be given a VMware image with a specially prepared lab environment to work on many topics and exercises in this training. When the training is over, students can tak environment home (after signing a non-disclosure agreement) to continue practicing at their own pace.

What Students Say About My Trainings

Recommendations are available on my LinkedIn profile (https://www.linkedin.com/in/dawid-czagan-85ba3666/) – training participants from companies such as Oracle, Adobe, ESE Trend Micro, Philips, the government sector, and others.

What Students Should Know

Students should have a general understanding of application security and some experience with web technologies and APIs. Basic familiarity with programming or scripting, secur practices, and working with the command line is recommended.

What Students Should Bring

Students will need a laptop with 64-bit operating system, at least 16 GB RAM, 120 GB free hard drive space, administrative access, ability to turn off AV/firewall and VMware Playe (64-bit version). Prior to the training, make sure there are no problems with running x86_64 VMs. You will need an OpenAI API key (required). A Lakera API key is optional.

As AI is an evolving field, additional requirements may be shared ahead of the training if needed.

Dawid Czagan is an internationally recognized security researcher and trainer. He is listed among top hackers at HackerOne. Dawid Czagan has found security bugs at Apple, Go Microsoft and many others. Due to the severity of his findings, he received numerous awards.

Dawid Czagan shares his offensive security experience in his hands-on training courses. He delivered training sessions at key industry conferences such as DEF CON (Las Vegas AppSec EU (Barcelona), Hack In The Box (Amsterdam), CanSecWest (Vancouver), 44CON (London), Hack In Paris (Paris), NorthSec (Montreal), HITB GSEC (Singapore) and fo clients. His students include security specialists from Oracle, Adobe, ESET, ING, Red Hat, Trend Micro, Philips and the government sector. Recommendations are available on Da LinkedIn profile (https://www.linkedin.com/in/dawid-czagan-85ba3666/).

Dawid Czagan is the founder and CEO at Silesia Security Lab. To find out about the latest in his work, you are invited to subscribe to his newsletter (https://silesiasecuritylab.com/ follow him on Twitter (@dawidczagan), YouTube (https://www.youtube.com/channel/UCG-sIlaM1xXmetFtEfqtOqg), and LinkedIn (https://www.linkedin.com/in/dawid-czagan-85ba3

Anti-Forensics (and Anti-Anti-Forensics) Techniques for Incident Responders

Stephan Berger (malmium.com)

Today, incident response is more challenging because attackers use anti-forensic tactics to block data collection and weaken forensic evidence. In this course, you will take a tech methods on Windows, Linux, and macOS. You will learn how attackers try to avoid logging, change filesystems, and slip past endpoint detection.

The training is built around the ideas of "Anti-Forensics" and "Anti-Anti-Forensics." For each evasion method, you will study how to detect it and how to recover evidence. The cou to find traces left behind by attackers, even when they try to hide their actions. You will work with real case studies and hands-on labs, including MFT reconstruction, SRUM analys and Linux runtime forensics.

Stephan Berger is the Head of Investigations for an Incident Response team at InfoGuard, a Swiss-based cybersecurity firm. With over a decade of experience investigating comp compromises, he specializes in the technical intersection of offensive tradecraft and defensive forensics. Stephan is the author of the DFIR.ch technical blog and is a regular speal international security conferences, including FIRST, Troopers, and hack.lu. He holds a Bachelor's degree in Computer Science and a Master's degree in Engineering and is the fo Malmium, a specialized technical training provider.

Automating LLM-powered Security Operations

Josh Pyorre (Pyosec (personal research), Cisco Talos (dayjob))

Security operations are full of repetitive, high-volume work: triaging alerts, processing threat intel, scraping vendor advisories, analyzing malware reports, generating detection rule summaries. Most of this work could be automated. This workshop will teach attendees to combine LLM reasoning with durable workflow orchestration to solve the specific problem security teams: automatically extracting IOCs from threat intel pages, generating detection rules from malware reports, classifying and triaging batches of alerts, and escalating tru automated response workflow, all without human intervention.

Over the course of the workshop, attendees will build and run:
- A threat intelligence scraper that fetches and parses public intel sources and uses an LLM to extract structured IOCs
- A malware report processor that reads email-based threat reports and generates Suricata/Snort detection rules via LLM
- An alert triage pipeline that classifies a batch of IDS alerts and escalates true positives to an automated response workflow
- A Temporal + Ollama environment pre-wired for security workflows that attendees take home and extend
- An LLM-based workflow to remove the tedium from security work, enabling security practitioners to get back to doing more creative tasks

The workshop does not require vendor products or paid subscriptions. Everything runs locally in Docker on attendee laptops. Where workflows integrate with third-party APIs (Viru etc.), free API tiers are sufficient. Attendees with existing production API keys are welcome to use them. Participants will leave with a working environment, all source code, and th it.

Josh has been in security for over 20 years, working with organization such as NASA, Mandiant, ZScaler, and Cisco.
He has presented at conferences and locations around the world, including DEFCON, Blackhat, Derbycon, InfoSecurity World, DeepSec, Qubit, and various companies and gover researching new and unique fields, often choosing to build tools that solve difficult problems.

From Research to CVE and Zero-Day Discovery: A Practical Guide to Discovering, Validating and Report

Ahmed Hassan (Cyber Security Expert)

Security vulnerabilities are discovered every day, yet many security professionals, developers, and researchers lack a structured methodology for identifying, validating, responsibl obtaining CVE assignments for their findings. As organizations increasingly depend on open-source software, the ability to discover and responsibly
report vulnerabilities has become a critical skill for improving the security of the global digital ecosystem.

This hands-on training provides a practical, end-to-end guide to vulnerability research, CVE acquisition, and responsible disclosure. Participants will learn how to identify security v open-source applications, analyze attack surfaces, perform source code reviews, conduct dynamic testing, and validate findings through practical proof-of-concept development. T common vulnerability classes including authentication
flaws, authorization bypasses, injection vulnerabilities, insecure configurations, business logic issues, and remote code execution vulnerabilities.

Through realistic case studies and hands-on exercises, attendees will explore vulnerability research methodologies used by professional security researchers and bug hunters. Th demonstrates how to document findings, assess security impact, create reproducible proof-of-concepts, assign CVSS scores, and communicate effectively with
vendors and maintainers.

Participants will also gain practical knowledge of the CVE ecosystem, including CNA processes, disclosure timelines, coordinated vulnerability disclosure practices, and the steps a CVE identifier. Real-world examples of publicly disclosed vulnerabilities and lessons learned from the disclosure process will be discussed throughout the training.

By the end of the course, attendees will possess a structured framework for discovering, validating, documenting, and reporting security vulnerabilities while contributing responsib of open-source software and the broader cybersecurity community.

Ahmed Hassan is an Austrian Cybersecurity Expert, Penetration Tester and Security Researcher with more than 8 years of experience in offensive and defensive information secu focuses on identifying, analyzing and mitigating complex security risks across enterprise environments, government organizations and critical infrastructure sectors.

Throughout his career, he has conducted numerous security assessments, penetration tests, red team engagements and comprehensive security evaluations for both national and organizations. His experience spans a wide range of industries, including government and public sector institutions, financial services, telecommunications, energy providers, heal aviation, technology companies, critical infrastructure operators and international organizations.

Ahmed Hassan has discovered and responsibly disclosed security vulnerabilities affecting renowned organizations and companies, including the United Nations (UN), Oracle, SAP numerous government institutions worldwide. His research has resulted in more than 53 assigned CVEs (Common Vulnerabilities and Exposures), contributing to the improvement cybersecurity standards and practices. In addition to his technical expertise, he holds several internationally recognized cybersecurity certifications, including OSCP (Offensive Sec Professional), OSWA (Offensive Security Web Assessor), CRTP (Certified Red Team Professional).

As a recognized speaker, Ahmed Hassan regularly presents at international cybersecurity conferences and shares his expertise with the global security community. His speaking e include prominent events such as Black Hat MEA (Saudi Arabia), Hack Red Con (USA), GISEC 2025 (UAE), Arab Security Conference (Egypt), Arab Internation Security Conferen BSIDES Vienna 2026 (Austria) and various other international cybersecurity conferences and industry forums and universities.

Furthermore, he has represented Austria at several high-profile cybersecurity and technology events across the world. These engagements were conducted in collaboration with a support of Austrian embassies, promoting Austrian cybersecurity expertise while fostering international cooperation, knowledge exchange and strategic partnerships.

Recommendations from previous labs and students/attendees are available on Ahmed Hassan's LinkedIn profile and post
(https://www.linkedin.com/feed/update/urn:li:activity:7464031469158477824/).
Ahmed Hassan is on every social media platform active and would be more than happy to be connected with all of you. To find out about the latest in his work, you are invited to c on LinkedIn, Facebook, Twitter and subscribe to his YouTube Channel where all his talks, presentations and event talks are uploaded. 🙂

YouTube: https://www.youtube.com/@titaniumcybersecuritysolut3431
Twitter https://x.com/hassanahmed8199/

LinkedIn: https://www.linkedin.com/in/ahmed-hassan-79559487/
Facebook: https://www.facebook.com/ahmedhamada81/

Hybrid Threats: OSINT for the Protection of Critical Infrastructure

Christina Lekati (Cyber Risk GmbH)

Today's threat landscape is fast evolving into a complex hybrid battlespace. Nation-states increasingly seek to gain a geopolitical advantage through strategic and multi-domain ad campaigns that are designed to operate in the grey zone: below the threshold of open conflict, yet designed to deliver significant strategic, economic, and societal harm. These hyb often combine social engineering, cyber intrusions, and physical damage or disruption (in any combination).

Whether the goal is damage and sabotage, the destabilization of societies, or the theft of critical information and intellectual property, public and private organizations have becom participants in this geopolitical chessboard.

In times of geopolitical tension, protecting critical infrastructure becomes a high priority, and securing the human and physical perimeters of this infrastructure necessarily starts by their level of exposure against adversarial tactics. Energy networks, transport systems, and other essential services are not only operational assets but also strategic pressure poir time, their growing interconnectivity and public visibility have expanded their exposure. Adversaries routinely leverage Open-Source Intelligence (OSINT) to identify vulnerabilities, dependencies, and plan operations from a safe distance and with precision both on the physical and human layers.

This 2-day training provides a practical descripton of how hybrid threats manifest into the real world with particular focus on Europe's evolving threat environment and the protectio infrastructure. OSINT will be a strong component of this class, both as an attack facilitator and as a vulnerability discovery tool for defenders. Hands-on exercises will be provided f class.

Day 1 will focus on providing a view into current hybrid threat landscape and some of the operations that have been observed within Europe. We will discuss tactics that involve hu hybrid campaigns focusing on infiltration, recruitment, and social engineering. We will move onto campaigns that involve the physical targeting of facilities through OSINT.

Working towards Day 2, we will continue discussing OSINT tools and techniques for the physical profiling of facilities. We will look into vulnerability exposure through a variety of p database-correlation. The second day will include a dedicated section on geolocation and image analysis. The course will also explore the growing role of artificial intelligence in m image analysis, and in accelerating OSINT-driven reconnaissance. Countermeasures and defense tactics will be discussed throughout the class.

By using real-world examples, participants will examine how human and physical vulnerabilities are systematically exploited by threat actors in the pursuit of geopolitical agendas v legal or military responses, and what to do to better defend against these threats.

Hands-on exercises are included throughout the training.

Key Training Topics:
• Understanding Hybrid Threats:
Overview, domains, operational characteristics & real-world incidents
• Adversaries using hybrid tactics:
State, non-state & state-affiliated actors. Goals & known modus operandi
• Human targets & recruitment operations:
Hybrid campaigns focusing on infiltration, recruitment, and social engineering
• OSINT as an attack facilitator and as a vulnerability identification tool
• OSINT techniques for critical personnel
• OSINT for the physical profiling of facilities:
Critical infrastructure databases, satellite and aerial imagery, mapping services, geolocation, image analysis, cross-database correlation, and more
• Defensive Hybrid Intelligence (DHI)
A defensive framework for identifying, assessing and mitigating hybrid risks

Christina is a psychologist working in cybersecurity. She specializes in social engineering, state-sponsored attacks, and in OSINT-based vulnerability assessments.

She brings over 10 years of experience working with Cyber Risk GmbH as a lead trainer and consultant in cyber-enabled hybrid threats and social engineering attacks. She has pa penetration tests, organizational trainings, high-value target & corporate vulnerability assessments, as well as in building security strategies against advanced persistent threats (A centric attacks. She also conducts OSINT vulnerability assessments to help organizations and high-value individuals identify and manage risks.

Her clients include Fortune 500 organizations as well as public- and private- sector companies of varying sizes. She has also delivered training to law enforcement personnel, milit similar institutional audiences on social engineering, open-source intelligence (OSINT), and human intelligence.

Christina is frequently invited as a speaker, keynote speaker or trainer at leading cyber security events around the globe including Black Hat, DEF CON, Blue Team Con, SANS Su more.

Surveilling Politicians, Unmasking Influencers, Fake Likes and Dying Together

Dries Depoorter

In this talk, Dries Depoorter takes you inside the technical and creative process behind projects that have made international headlines. "The Flemish Scrollers" runs real-time com government livestreams to catch politicians scrolling their phones and tags them publicly. "The Follower" combines open surveillance cameras with AI to match Instagram photos t footage of the moment they were taken. "Die With Me" is a chat app that only works when your phone battery drops below 5%, an experiment in digital scarcity that went viral.

The talk also includes a live demo: using a custom Python tool that controls social media accounts, Dries will manufacture fake likes on a post in real time, showing just how easy break the trust system that social media is built on.

This talk is for anyone interested in surveillance, computer vision, social media manipulation, and what happens when you approach security research from the art world instead o

Dries Depoorter is a Belgian artist and creative technologist who builds systems that exploit the gaps between surveillance infrastructure, AI, and public trust. His work turns openly streams, public cameras, social media APIs, parliamentary livestreams, into tools that expose how little privacy actually exists. His project "The Flemish Scrollers" uses real-time c government livestreams to detect and publicly tag politicians scrolling their phones during parliamentary sessions. "The Follower" combines open CCTV feeds with AI to match Ins the exact surveillance footage of the moment they were taken, proving how traceable our "curated" online lives really are. "Die With Me," a chat app that only works below 5% batt viral experiment in digital scarcity and human behavior under constraint. With a background in electronics, computer vision, and hardware prototyping, Dries works across Arduino ESP32, and custom PCB design to build installations that function as both art and proof-of-concept exploits. He has exhibited at the Barbican, Art Basel, Ars Electronica, and ZKM at SXSW, MoMA, and TEDx.

Dries Depoorter is a Belgian artist and creative technologist who builds systems that exploit the gaps between surveillance infrastructure, AI, and public trust. His work turns openly streams, public cameras, social media APIs, parliamentary livestreams, into tools that expose how little privacy actually exists. His project "The Flemish Scrollers" uses real-time c government livestreams to detect and publicly tag politicians scrolling their phones during parliamentary sessions. "The Follower" combines open CCTV feeds with AI to match Ins

the exact surveillance footage of the moment they were taken, proving how traceable our "curated" online lives really are. "Die With Me," a chat app that only works below 5% batt viral experiment in digital scarcity and human behavior under constraint. With a background in electronics, computer vision, and hardware prototyping, Dries works across Arduino ESP32, and custom PCB design to build installations that function as both art and proof-of-concept exploits. He has exhibited at the Barbican, Art Basel, Ars Electronica, and ZKM at Defcon,SXSW, MoMA, and TEDx.

Quantum Communication

Andreas Neuhold (XBC Digital GmbH)

Quantum computers will break the encryption we rely on today, and the clock is ticking.
But what does that actually mean, and what can you do about it right now?

This talk starts from the ground up, giving attendees an accessible introduction to the two quantum-safe building blocks: Post-Quantum Cryptography (PQC), a mathematical softw replacement for classical key exchange, and Quantum Key Distribution (QKD), a physics-based approach that uses quantum optical channels to distribute keys with information-th This talk cuts through the hype and delivers a technically grounded, practitioner-focused view of what quantum-safe communication actually means in 2026 and what you can dep

Whether you are a security engineer, network architect, or technical decision-maker, this session provides the concrete technical foundation needed to assess, plan, and implemer resistant communication today.

Andreas Neuhold has 16+ years of experience building and operating core and security solutions for cable, fibre and mobile internet access for leading international Tier-1 ISPs. Current areas of activity include projects and R&D to develop new technologies, standards and integrations in the areas of Quantum Communication like Quantum Secure VPN wi Key Distribution (QKD) and Post Quantum Crypto (PQC).

Hybrid Threats in an Interconnected World: From National Interests to Human-Targeted Cyber Operation

Christina Lekati (Cyber Risk GmbH)

The modern threat landscape has evolved into a hybrid battlespace in which national interests, state power, and private enterprise become increasingly intertwined. Nation-states pursue geopolitical objectives through multi-domain operations that combine social engineering, cyber intrusion, and physical disruption. These hybrid threats deliver real damage below the threshold of open conflict.
Geopolitical tensions have accelerated the use of these tactics. Through real-life case studies, this presentation shows how Advanced Persistent Threat (APT) actors launch hybri frequently begin with tailored phishing, impersonation, or psychological manipulation before escalating across digital and physical systems. The goals of these campaigns range fr gathering and financial gain, to sabotage operations and public perception influence campaigns.
The cases also show that private sector has become a frontline participant in this geopolitical conflict. Energy providers, defense suppliers, technology firms, and other types of bu being drawn into state-driven campaigns that employ hybrid tactics and exploit interconnectivity to gain geopolitical advantage.
Defending against these hybrid threats requires an integrated understanding of how people, networks, and physical systems interact in the pursuit of geopolitical advantage. Buildi demands coordination across sectors, threat awareness, and the recognition that cybersecurity, information integrity, and national security are inseparable in this age of hybrid thre

Christina is a psychologist and intelligence analyst working in cyber security. She specializes in the human element of cyber-attacks, open-source intelligence, and in vulnerability

She brings over 10 years of experience working with Cyber Risk GmbH as a lead trainer and consultant in cyber enabled hybrid threats and social engineering security. She has p penetration tests, organizational trainings, high-value target & corporate vulnerability assessments, as well as in building security strategies against advanced persistent threats (A centric attacks. She also conducts OSINT vulnerability assessments to help organizations and high-value individuals identify and manage risks.

Her clients include Fortune 500 organizations as well as public- and private- sector companies of varying sizes. She has also delivered training to law enforcement personnel, milit similar institutional audiences on social engineering, open-source intelligence (OSINT), and human intelligence.

PostGREShell: From One PostgreSQL Bug to RCE Across Seven Data Systems

Vladimir Tokarev (Cyera)

PostgreSQL's replication protocol lets a low-privilege user specify a plugin name that reaches dlopen() without the validation that exists on other code paths. On Windows, LoadLibrary() resolves the name over SMB and loads the attacker's DLL directly. That was the first RCE.

We broke the bug down to its parts and got: user-controlled path, OS native loader, missing validation. We looked for the same pattern across data infrastructure and found it in six more products: Firebird, Milvus, MonetDB, NVIDIA Triton, Apache Doris and Dgraph.

PostgreSQL powers millions of production databases worldwide and has a billion+ docker pulls. Milvus is the most widely deployed open-source vector database. NVIDIA Triton serves ML models across every major cloud provider. Apache Doris handles real-time analytics for large enterprises. Together they power tens of millions of deployments storing some of the most sensitive data on the planet.

We achieved pre-auth RCE on four of them. Firebird: one SQL statement to code execution, CVE-2026-40342, CVSS 10.0. MonetDB: JIT C compilation with hardcoded default credentials. Milvus: unauthenticated etcd poisoning into plugin.Open() on restart. Apache Doris: single HTTP request loads an attacker-controlled JAR. NVIDIA Triton: pre-auth model loading to code execution.

Dgraph broke the plugin-loading pattern, but the same audit approach produced two pre-auth DQL injection vulnerabilities leading to full database exfiltration (CVE-2026-41327, CVE-2026-41328, CVSS 9.1).

We present the exploit chains with demos, a repeatable methodology

for finding this bug class across any product, and a defense hierarchy built from the three databases that got runtime loading right.

Vladimir Tokarev is a vulnerability research tech lead at Cyera with a track record in cloud infrastructure, virtualization, and Linux/Windows kernel attack surfaces. His research has produced multiple CVEs across hypervisors, AI runtimes, and database engines. He has presented at Black Hat USA (2024, 2023), CODE BLUE (2025), DEF CON Recon Village (2025), and RSA (2024).

Behavioural Cyber Intelligence: The Missing Layer of Threat Intelligence

Liljana Pecova Ilieska (Association Impetus Skopje)

Cyber Threat Intelligence has traditionally focused on adversaries' infrastructure, malware, tactics, techniques, and procedures. Yet every cyber operation is ultimately shaped by h —how people learn, collaborate, adapt, and respond under pressure. This session introduces Behavioural Cyber Intelligence (BCI) as an emerging analytical layer that complemer threat intelligence by examining behavioural signals generated during cyber exercises, incident response, and operational environments. It explores how AI can help identify patter making, teamwork, resilience, and learning trajectories while raising important questions about ethics, privacy, and responsible use. Rather than replacing Cyber Threat Intelligenc Cyber Intelligence expands it, offering a new perspective on cyber resilience, workforce development, and the future of intelligence-driven cybersecurity.

https://shespeakscyber.org/speakers/liljana-pecova-ilieska

Towards Assumption-Homogeneous Post-Quantum Signature Systems

Maksim Iavich (Caucasus University / Scientific Cyber Security Association)

Post-quantum digital signatures are expected to become a foundational component of future secure infrastructures, including digital identities, certificate authorities, blockchain sys communication networks. However, many existing post-quantum signature architectures combine components based on different cryptographic assumptions, creating complex sec dependencies and potential points of weakness.

This talk presents a new lattice-based framework for post-quantum many-time signatures that unifies all major components under a single hardness assumption: the Short Integer problem. The proposed architecture combines functional commitments, Verkle trees, and HyperPlonk-based zero-knowledge proofs to create an assumption-homogeneous signat built-in privacy-preserving verification capabilities.

The presentation will introduce a novel functional-commitment-based one-time signature scheme, demonstrate how the same primitive can be reused to construct Verkle tree auth structures, and explain how zero-knowledge proofs can be employed to verify signature validity and authentication paths without revealing internal witness information. Particular a given to the security rationale behind assumption homogeneity and the benefits of eliminating the mismatch between hash-based and lattice-based cryptographic components.

Attendees will gain insight into emerging research directions at the intersection of post-quantum cryptography, authenticated data structures, and zero-knowledge proof systems, a practical challenges of designing scalable and privacy-preserving cryptographic infrastructures for the post-quantum era.

Prof. Dr. Maksim Iavich is a distinguished scholar and leading expert in computer science and cybersecurity. He holds a Ph.D. in Mathematics and serves as Professor and Head Science Department at Caucasus University. He is also the Director of the Cyber Security Center (CST, CU) and the CEO & President of the Scientific Cyber Security Association Iavich is a Corresponding Member of the Georgian National Academy of Sciences, a recognition that reflects his outstanding contributions to science and academia in Georgia and he has been working in Michigan University as a post-quantum cryptographer, focusing on the development of advanced cryptographic systems resilient to quantum computing thr his academic roles, he has served as a cybersecurity and computer science consultant for both Georgian and international organizations. Prof. Iavich is a sought-after speaker at conferences and has organized numerous cybersecurity events. Between 2018 and 2025, he delivered keynote talks at DefCamp, DeepSec, Hek.SI, ManuSec, QITW2025, and m prestigious forums, addressing topics at the intersection of cybersecurity and artificial intelligence. His achievements have been widely recognized: in 2018, he was awarded the tit Scientist in Computer Science by the Shota Rustaveli National Science Foundation, and in 2024, he was honored as Best Scientist at Caucasus University. He has also received i recognition as a DeepSec Scholar and a Weiser Scholar (2025). Prof. Iavich has authored numerous scientific papers covering cybersecurity, cryptography, artificial intelligence, m mathematical modeling, 5G security, and simulations. His work continues to shape the field of cybersecurity both in Georgia and internationally.

The Agentic SOC

Josh Pyorre (Pyosec (personal research), Cisco Talos (dayjob))

What does it look like to replace the humans in a Security Operations Center with agentic AI, and is it as effective as a traditional SOC?

This presentation covers the possibilities presented by LLMs to do the work typically required by humans in a SOC, and what it might mean for the future of security work, includin actors could adapt when the systems defending against them are running autonomously.

We will cover how a traditional SOC works, why the current industry trend of "AI-assisted analysis" is already behind where we are headed, and how the Agentic SOC was built us durable workflow orchestration engine), local and frontier LLMs, and open source security tooling. Seven agentic workflows that replace human roles will be demonstrated, includi automated incident response, threat hunting from public intel feeds, malware report processing, PCAP analysis, detection rule generation, and executive report writing.

The goal of this work is to bring affordable security operations and monitoring to small businesses and non-profits, organizations that are typically left behind by the cost and comp SOC models.
This is not a vendor presentation. Everything is open source, all code and instructions will be made available to attendees, and the results of this work will further the discussion of change what we do and what our purpose is.

Josh has been in security for over 20 years, working with organization such as NASA, Mandiant, ZScaler, and Cisco.
He has presented at conferences and locations around the world, including DEFCON, Blackhat, Derbycon, InfoSecurity World, DeepSec, Qubit, and various companies and gover researching new and unique fields, often choosing to build tools that solve difficult problems.

Playing with Quantum Keys

Lukas Mairhofer (Quantum Technology Laboratories)

Quantum Key Distribution (QKD) is a decisive step towards the future of cybersecurity. If implemented correctly, QKD allows unbreakable encryption that protects data even agains
unlimited computational resources or completely new technological platforms such as Quantum Computers. However, the concepts behind QKD contradict our everyday experienc
years after formulating the basic principles of quantum information, physicists are still discussing how to actually interpret them.
We will take an approach to quantum cryptography quite complementary to abstract mathematics. In a pen-and-paper game we create and share quantum keys hands-on, using s
entanglement of photons. We explore the impact of eavesdropping and experiment with different protocols.

Lukas Mairhofer has complemented his PhD in experimental quantum physics with a PhD in the philosophy of science and technology. He is an award-winning lecturer on quantur
currently leads the QURIOS Quantum Technologies Academy at qtlabs.

Keeping OT Insecure: A Masterclass on How To Make Hackers Happy

Melina Phillips (GCI)

Did it ever happen to you as a kid, you'd see a machine, a switch, or a big shiny button, and you just had to know… "What happens if I press this?"
Well, hackers never grow out of that phase. Except instead of breaking a toy, they poke at devices that control factories, power grids, HVAC infrastructure, alarm systems and reall
and sometimes, we don't even have to press a button. The system just hands us the keys like, "there you go, have fun".

To make matters worse, many of these environments were never designed keeping adversarial activity in mind. They were built to keep operations running, not to survive curious p
misuse, exposed interfaces, or attackers who understand how to abuse trust. In a lot of OT environments, convenience quietly becomes policy. Shared credentials becomes the nc
protocols stay enabled because "nothing has broken yet." Internal systems are assumed to be safe simply because they live behind a firewall somewhere nobody has touched sin

In this session, we will discuss how Operational Technology environments tend to follow a similar pattern: function prioritized over safety, convenience over design, and legacy ove
real world misconfigurations, unauthenticated interfaces, sensitive information exposure, inherited trust assumptions, and decades of accumulated technical debt, we'll walk throug
takes to keep hackers happy.

We will also explore how attackers do not always need sophisticated tools and binaries to create operational impact. Sometimes the most dangerous thing in the environment is its
functionality and the trust nobody thought to question.

==========

OUTLINE

Intro:
• About me.

• Wait, what are we doing here?

Baby HP.

• I typically like to start my presentations with a short story and then correlate it to the main topic.

Mexican writer Juan Jose Arreola describes a fictional device that straps to children to harvest kinetic energy. It's marketed as efficient, modern, safe, and guaranteed.
But it's absurd, unethical, and nobody asked the kids.

• It mirrors how we treat OT:
○ Deploying tech in the name of productivity
○ Minimal concern for risk or ethics
○ Blaming users when the device fails

Key learning: "Modern" doesn't mean "secure."

"Warnings about accidents, system overload, and 'attracting lightning' are dismissed as irresponsible rumors. After all, the device is factory guaranteed, modern, and reliable."…wh
go wrong?

What's OT anyway?

• We are not going to reinvent the wheel but cover our basics.
• Common types of OT devices

That's good and all, but why do we need that stuff?

IT vs OT: The breakup that nobody ever talks about.

IT is primarily concerned with managing and processing data to ensure its availability, confidentiality, and integrity. Its domain includes computers, servers, and similar devices that
storage, retrieval, and transmission tasks. OT, on the other hand, is tasked with controlling and monitoring physical processes and equipment, ensuring operational effectiveness a
industrial settings.

Getting our hands dirty

What makes OT environments dangerous isn't just one vulnerability, it's the incorrect assumption that internal deployments can be trusted by default.

Note: Each of these will be individually explained, what the device does, sample logs, each misconfiguration and how it can be abused.

Vulnerability Real Example
No GUI auth Vertiv Watchdog 15
Anonymous FTP Eltek Smartpack controllers
Default SNMP APC NetBotz (public/private)
Info leaks Rockwell/Siemens logging unprotected
Backdoors WEMS BEMS 21.3.1

Telnet open Industrial routers, UPS consoles
No RBAC Shared operator logins
No protocol auth Modbus, DNP3 in the clear
No integration Can't install agents, no logging support

The call is coming from inside the house: Weaponizing existing functionalities

Unlike typical IT attacks, which commonly target applications, OT attacks can have direct physical consequences. Attackers can manipulate OT systems to cause equipment malfu
critical infrastructure, or even create dangerous situations.

Accuenergy - Accuvim axm-web2
Potential abuse vectors:

• Denial of service
• Command injection
• Out of band callbacks
• Information disclosure

Hazard in numbers.

The following misconfigs were observed in an Upconverter vendor web platform.

- Unauthenticated application logic disclosure via verbose HTTP responses
- IDOR
- Lack of auth enforcement.
- Unauthenticated role elevation via parameter manipulation.

But…but, the system was working as designed: Oldsmar Water Treatment facility incident.

· This was a direct threat to Operational Technology (OT)/Industrial Control Systems (ICS) that manage physical processes.
· This is a perfect example on how cyber attacks could potentially have real world consequences and manipulate outcomes such as water chemistry, becoming a really severe hea
· This incident highlights how infrastructure in a smaller scale (in this case municipal) has even more limited resources in comparison to large energy companies

Hot girl summer - critical infrastructure edition: When operational trust becomes the attack surface.

Why would anyone care about securing the infrastructure of beauty retailers? At first glance, it might seem like cybercriminals have bigger targets, who would launch a sophisticate
snag a cart full of blush, lip plumpers, and highlighter palettes?

Sounds harmless, right?

WRONG

Beauty is power. Data is currency. And attackers are cashing in.

· An attacker discovers an environmental monitoring appliance via OSINT.
· Keep in mind that in many cases, these appliances do not offer an authentication option.
· Weak SNMP community.
· Management interface accessible internally as well.
· Shared operator credentials,
· Lack of segmentation between facilities monitoring and broader corporate systems,
· Attacker abuses existing functionality instead of deploying sophisticated malware.
· SOC visibility gaps delay detection.

Countermeasures

Let's talk about things that actually work, and don't involve shutting down the factory.

Foundational practices:
• Network segmentation
○ Separate OT from IT and from other OT systems (e.g., HVAC isn't the same as SCADA)
○ Use firewalls and VLANs to isolate devices
• Logging + central collection
○ Forward logs to a SIEM or OT-aware platform
○ At a minimum, catch authentication attempts and config changes

Authentication and access control:
• Enforce strong passwords and rotate default creds
• Enable RBAC where supported
• Use jump servers and secure remote access (e.g., hardened bastion hosts)

Visibility and monitoring:
• Use passive monitoring tools (e.g., Nozomi, Dragos, Claroty)
• Identify what's even on the network (many orgs don't know!)
• Detect new devices or unauthorized changes

Protocol hardening:
• Disable unused services (FTP, Telnet, SNMPv1/v2c)
• Move to SNMPv3 or secure protocols when possible
• Disable Modbus write access unless explicitly needed

Reduce Attack Surface:
• Remove unused services and default pages from web interfaces
• Disable debug endpoints
• Physically secure ports (USB, console, Ethernet)

Optional but kind of cool:
• Deploy OT-aware honeypots to trap internal threat movement
• Use DNS logging to catch malware or C2 callbacks on hybrid systems

Applying counter measures Guidelines

AgileTAC: Operational Proactive security. AgileTAC is an iterative framework for applying continuous security improvements to infrastructure, rooted in agile methodology and grou
Trust principles by default

Instead of waiting for the next compliance cycle or incident driven security update, you use sprints to:
· Continuously identify security gaps in OT networks
· Apply Zero Trust principles in a structured, iterative way
· Prioritize fixes based on risk impact instead of waiting for a full project overhaul

• Why Traditional OT Security Fails
• Slow, compliance-driven security updates
• Legacy systems + resistance to change
• Reactive security mindset instead of proactive

• AgileTAC for OT: Security as a Continuous Process
AgileTAC inherently implements Zero Trust by design:
• Sprint Planning - Identify security gaps (e.g., Telnet still enabled? Backdoor accounts still exist?)
• Sprint Execution - Small, actionable fixes every sprint instead of massive delayed projects
• Retrospectives - Assess what security controls worked and adjust the next sprint

• AgileTAC = Operationalizing Zero Trust
Every asset and user is treated as untrusted by default
Least privilege, microsegmentation, and continuous validation are applied incrementally
OT and engineering teams collaborate in short, focused cycles to harden systems progressively

Reality bites, so let's fix it: APC NetBotz Rack Monitor 250

APC's NetBotz devices are used to monitor environmental conditions (temperature, humidity, door status, airflow, etc.) in server rooms, data centers, and telecom environments. T
edge devices that often fly under the radar of traditional security efforts.
These devices:
• Have web interfaces (HTTP/HTTPS)
• Often expose SNMP with default community strings
• Sometimes leave Telnet enabled by default
• Rarely support modern endpoint protection
• Are deployed in critical environments, but not always segmented or monitored

Sprint Activities Why It Matters
Sprint 1 Audit and remove default credentials, change SNMP public/private, disable Telnet NetBotz devices are often discoverable with default SNMP strings and have weak or ha
Sprint 2 Place devices into a segmented VLAN (e.g., "OT-monitoring"), restrict access via firewall Prevents lateral movement from IT to OT, or pivoting from compromised endpoin
Sprint 3 Enable SNMPv3 or syslog forwarding to central SIEM, configure alert thresholds Most NetBotz devices can forward events, but it's rarely set up; logging brings visibility
Sprint 4 Test controls: verify no open management interfaces, no weak creds, confirm alerts fire Ensures the previous sprints success, documents residual risk.

• The Takeaway: Making OT Security Manageable
• Small, frequent security improvements = less risk accumulation
• Makes security part of the operational process instead of an afterthought
• Execs love it because it delivers measurable security progress without huge downtime

Final Thoughts

• OT isn't just legacy, it's critical infrastructure.
• Security by obscurity is a great strategy…to set yourself for failure.
• Segmentation is your best friend (for both IT and OT networks).
• Monitor your environment like you monitor IT.
• Assume everything is using default credentials and exposed configurations everywhere.
Patch when you can, compensate when you can't.
• Disable unused services.
• Collaborate with engineers (AgileTAC).
• Visibility is everything, if you can't see it, you can't secure it.
• The bad guys already understand your environment, do you?

Melina Phillips is an Offensive Security Engineer with a background in Security Operations and Incident Detection. She has over ten years of IT experience and six years working
cybersecurity, blending hands on blue team work with her current focus on adversary simulation and endpoint compromise.

Her recent talks have been featured at Bsides Cambridge, Security Fest, BruCon, LeHack, HackLu, BlackAlps, Bsides Luxembourg and Insomni'hack. She's known for making co
concepts accessible without watering them down, and for delivering practical insights grounded in real world attack and defense experience. She strongly believes that information
have to be presented in a boring way, and that technical depth and creativity can (and should) coexist.

Outside of breaking into infrastructure and chasing down Linux threats, she's usually at CrossFit or playing with makeup, ideally not at the same time.

VibeShell: How Trusting Your AI IDE Costs You Your Machine

Etizaz Mohsin (N/A)

Forget prompt injection. While the industry obsesses over manipulating model inputs to bypass guardrails, it is overlooking a far more dangerous threat requiring no user interactio
the application embedded beneath the AI itself: the underlying platform architecture.

To deliver on the promise of full autonomy, AI-powered coding environments wire Large Language Models directly into the developer workflow—handing them local filesystem acce
execution, and cloud credentials. Developers accept this tradeoff for the massive productivity gains, but the security cost is severe.

In these environments, privileged OS access is not a misconfiguration but a product requirement. For a chat interface to truly become an autonomous "agent," it must be equipped
Equipping the AI with these tools creates a structural conflict with traditional application isolation, like Electron's security model. To make the AI function, developers are forced to b
and expose highly permissive IPC (Inter-Process Communication) bridges between the web renderer and the local operating system.

This presentation provides a technical deep dive into how chained IDOR vulnerabilities can be escalated into zero-click RCE via persistent LLM conversation injection and unsafe designs. To prove the real-world impact, we will debut novel research into Orchids, a leading local, Electron-based AI coding IDE with over a million users, reported to be used by t including Google, Amazon, and Uber.

By weaponizing the IDE's automated context ingestion, we turn the platform's own features into a silent backdoor, achieving full remote system takeover with no user interaction be application.

Attendees will leave this session with a practical understanding of this emerging attack surface, a live demonstration of the zero-click exploit chain, and actionable defensive desig safely architecting the next generation of AI-enabled desktop applications.

Etizaz is a Pakistani cybersecurity researcher focused on advanced threat research and offensive security. His work has explored real-world attack techniques used by sophisticat including research on NSO Group's Pegasus spyware, which he presented at Black Hat MEA. Earlier in his career, he investigated the DarkHotel espionage campaign targeting ex diplomats through compromised hotel networks.

His research has been featured by outlets such as Forbes, BBC, Wired, TechCrunch and Al Jazeera. He has presented at international security conferences including Black Hat M Africa, HITCON, 44CON, CONFidence, DeepSec, SECTOR, Hacktivity, GreHack, HackFest, and Wild West Hackin' Fest.

Over the Air, under the Radar

Jonas Stein & Matthias Weitz (Institute for Occupational Safety and Health of the German Social Accident Insurance DGUV e.V.)

Whilst all eyes are on office IT and there is debate over whether RSA or elliptic curves are the method of choice, certain niche areas have been completely overlooked. In the indus are radio-controlled devices with safety functions whose connections are completely unprotected. Security researchers investigated the situation in the US market and published a the subject in 2018. In our laboratory, we carry out testing and certification, conduct accident investigations and conduct research into prevention. Over the last few years, we have various products in our laboratory and have come across quite a few surprises. It is not a single manufacturer that has made a mistake, but practically every other product with rad share the hardware and software tools we use in the lab to carry out our investigations, the typical faults we've identified, and how these could easily be avoided. Using a model se demonstrate live how such a system works and explain the problems. We'll examine the toolchain surrounding software-defined radio and antennas, and finally discuss whether ar can be detected.
The insights gained from this are being incorporated into new European regulations and international standards to make machinery and control systems safer. Jonas Stein is a me expert panel for the European Commission's Machinery Directive guidelines and is also active in international standardisation. The new findings are shared with manufacturers to products become increasingly safer.

Jonas Stein specializes in the safety and security of machinery. He is head of the Industrial Security Section at the German Social Accident Insurance (DGUV). He also leads the manufacturing project group of the ISSA Machine and System Safety Section, as well as the DGUV AK Security Working Group. Within the CENELEC/TC44x/WG02 standardizati he contributes to the EN 50742 security standard for machinery and functional safety devices. He is also a member of the Expert Group SG4 of the European Commission, contrib Machinery Regulation guide. As certifier for industrial security, he provides various laboratory services to help make workplaces safer.

Unintended Computations

Nicolas Seriot (Swissquote Bank)

How can a system be repurposed? When does it become a computer?

We treat everyday systems as static tools with fixed purposes, yet beneath the surface lies latent computational potential. Misuse reveals what they are by ignoring what they were

I'll show how I turned a laser printer into a chess engine, built an audio modem in five lines of AWK, built a Turing-complete machine from Unix text tools, and shaped ten wooden adder.

We'll look at these creative projects through the formal lens of models of computation, and learn to see everyday systems as malleable primitives for unintended computations.

Along the way, we'll see how constraints and minimalism can sharpen our thinking more than unlimited capability ever could.

Nicolas Seriot is a security engineer. His earlier work focused on iPhone privacy, the Twitter API, and JSON parsing. He now writes about unusual corners of computing, most nota programming. His articles appear in Paged Out! magazine, his code lives at github.com/nst, and he is reliably drawn to making everyday tools do things their designers never inter

Trust No Signal - Exploiting Digital Energy at Level 0

Paul Coggin (nou Systems, Inc)

The convergence of the digital and physical worlds has opened a physics-based attack surface that traditional cybersecurity does not address, particularly at the foundational Purd define this new vulnerability through digital energy: the physical manifestation of computation. Our core argument is that manipulating this energy—through electromagnetic interfe mechanical force—allows attackers to side-step software defenses and compromise operational technology. Because advanced threats may exploit the physical environment to di and actuators, security must undergo a fundamental shift. The way forward is the urgent integration of physical layer security monitoring to protect critical infrastructure at its deepe

Paul Coggin is a recognized cybersecurity expert at nou Systems, Inc., specializing in solving complex problems at the convergence of space, digital energy, telecommunications, and cybersecurity. His expertise spans space systems, service provider networks, ICS/SCADA infrastructures, and digital energy. With extensive experience in network architectur analysis, and penetration testing, Paul has led security assessments for critical infrastructure across various sectors, including aerospace, energy, service provider, financial and ta

A frequent speaker and instructor at international conferences, he shares insights on advanced cybersecurity topics. Paul holds multiple advanced degrees: BS in Mathematics/Co MS in Space Systems, MS in Systems Management, MS in Information Assurance and Security, and MS in Computer Information Systems, along with numerous industry-recogni

Agent Memory Guard: Securing Long-Term Context in Autonomous AI Systems

Vaishnavi Gudur (Microsoft)

As autonomous agents evolve from stateless chatbots to persistent systems with long-term memory, they inherit a massive new attack surface. "Agent Memory Guard" explores th vulnerabilities introduced when agents can read, write, and act upon persistent context across sessions. This talk dissects real-world failure modes—such as memory poisoning, c leakage, and persistent prompt injection—and introduces a robust architectural framework for securing agent memory. Attendees will learn how to implement cryptographic memor context-aware access controls, and adversarial memory sanitization to build agents that remember safely.

Vaishnavi Gudur is a Senior Software Engineer at Microsoft, where she architects secure and scalable systems that safeguard more than 145 million Microsoft Teams users worldv focuses on phishing protection, compliance, and AI-driven security, blending deep technical expertise with a commitment to responsible innovation. Beyond her engineering role, V active researcher and thought leader in AI governance and ethical AI. She has published on topics including privacy-sensitive smart contracts, zero-day cyber threat detection, and architectures.

As a recognized voice in the field, Vaishnavi has delivered invited talks at leading industry and academic forums, including NDC Sydney, SecurityWeek's AI Risk Summit, IEEE Clo Symposium, and CodeSecCon, where she highlights the real-world risks of deploying AI in enterprise environments and offers practical frameworks for building resilient, ethical sy serves as a peer reviewer for various IEEE, ACM journals and papers and as a lead editor with Springer Nature, helping shape the next generation of scholarship in AI ethics and

Passionate about bridging research, engineering, and governance, Vaishnavi has built an open source tool, Agent Memory Guard, adopted by OWASP foundation. It is a runtime c screens every read and write to your AI agent's memory, blocking prompt injection, secret leakage, and integrity tampering before they corrupt agent behavior across sessions. It is by startups and has many contributions on Github.

Connected, Tracked, Exposed: Threat Modeling a Production GPS Fleet Platform

Maria Afonina (Arcoda)

A GPS fleet platform is a fleet of embedded devices in the field, a link over somebody else's network, and a cloud backend holding the location history of real vehicles and real peo customers in transport and logistics are increasingly subject to NIS2 obligations, which turns the vendor's security posture into a contractual matter, so a weakness in the tracker is technical bug, it is a compliance and procurement problem for everyone downstream.

This talk walks through a structured threat modeling exercise built around a production GPS fleet management platform, conducted as part of an MSc thesis collaboration between Institute of Technology, the University of Trento and an Italian IoT company. We decompose the attack surface across three layers: the tracking device, the device-to-backend chan backend. For each layer we identify realistic threat scenarios, from device capture and firmware extraction through channel manipulation and replay to abuse of backend APIs, and feasibility and business impact rather than theoretical severity.

The focus is method, not a single finding. How to model a production system as it actually operates, how to keep the analysis grounded in realistic attacker capabilities and operati which threat catalogues earn their keep and which just generate noise, and how to turn the resulting model into something that survives an auditor or an enterprise customer's sec questionnaire. Findings are anonymised. Attendees leave with a reusable, layer-by-layer workflow and templates applicable to any device-to-cloud product.

Maria Afonina is an information security and compliance consultant with 12+ years in GRC, security audits and risk management, certified CISM and CISA, and an ISO/IEC 27001 She has led 25+ audit and consulting engagements for enterprise clients, designing ISMS, threat and risk models and full policy frameworks under ISO 27001, NIST, COBIT and G completing an MSc in Cloud and Network Infrastructures at KTH Royal Institute of Technology and the University of Trento, where her thesis develops threat modeling for IoT fleet systems in collaboration with an Italian IoT company.

AI Hype Cycle - Give me a break

Stephan Berger (InfoGuard AG)

Walk the vendor hall at any major infosec conference this year, and you're getting pitched the same apocalypse: autonomous AI threat actors are coming for your network. Natural selling an expensive, next-gen silver bullet to stop them.

But from the trenches of actual incident response, the view is a lot less sci-fi. We aren't fighting Skynet. We're fighting an affiliate who bought valid VPN credentials on a Telegram bucks and just logged in.

This talk breaks down the massive disconnect between boardroom AI panic and the reality of 2026 intrusions. Pulling from hundreds of recent IR engagements, we'll look at how th actually operating right now. They aren't spinning up zero-days with custom LLMs; they're mostly just using GenAI to fix the grammar in their phishing lures.

We'll walk through the anatomy of these breaches and highlight the classic "crunchy outside, chewy inside" problem: decent perimeter security masking total internal chaos. We're networks, stale admin accounts, misconfigured cloud permissions, and easily bypassed MFA. The takeaway is simple: you can't buy a shiny AI appliance to patch over bad IT hygi

Stephan Berger is the Head of Investigations for an Incident Response team at InfoGuard, a Swiss-based cybersecurity firm. With over a decade of experience investigating comp compromises, he specializes in the technical intersection of offensive tradecraft and defensive forensics. Stephan is the author of the DFIR.ch technical blog and is a regular speak international security conferences, including FIRST, Troopers, and hack.lu. He holds a Bachelor's degree in Computer Science and a Master's degree in Engineering and is the fou Malmium, a specialized technical training provider.

Damage Control: the gap between assumed and actual control coverage across enterprise estates

Lucy Caiger (Panaseer)

Every security team keeps score with control coverage numbers; how much of the estate has endpoint protection? Or how many assets are patched inside SLA? As an operator, y numbers to be right: they decide where remediation effort goes, and they're the evidence behind every assurance you pass upward. But each number rests on an assumption you that reported coverage is actual coverage. When it isn't, you're exposed without knowing where.

In this session, I'll start with your assumed coverage, comparing it against real-world enterprise data. The gap is usually wider than expected, but what turns it into real exposure is

Because a standard coverage metric counts every gap the same and assumes they are unrelated. But that's not how an attacker experiences an estate, and neither does frontier A concentrates on the few assets carrying several weaknesses at once. They offer a ready-made way in, and a siloed toolset has nothing watching for the combination.

You'll leave the session understanding where assumed and actual coverage diverge, why that gap is your true exposure, and how to prioritise and reduce the exposures that put th risk.

With a strong customer focus at the core, Lucy has a wealth of experience delivering complex data solutions to industries including Financial Services, Telco, and Government. He areas of innovation, primarily Cyber Advanced Threat Detection and Cyber Controls Effectiveness.

Defending Identity Infrastructure of the Active Directory with Deception Technologies

Ahmed Hassan (Cyber Security Expert)

Active Directory serves as the backbone of identity and access management across government agencies, critical infrastructure operators, and enterprise environments. As attack focus on identity systems to achieve privilege escalation, lateral movement, and domain dominance, traditional defensive controls often struggle to detect early-stage reconnaissar compromise activities before significant damage occurs.
This presentation explores how deception technologies can be strategically deployed within Active Directory environments to detect, delay, and disrupt adversaries before they rea Rather than relying solely on preventive controls, deception provides defenders with an opportunity to identify attacker activity at an early stage by introducing realistic but controlle throughout the identity infrastructure.
Through practical demonstrations and real-world attack scenarios, attendees will learn how deception techniques such as honey users, honey credentials, honey groups, deceptiv accounts, canary tokens, and decoy systems can be integrated into Active Directory environments to expose malicious reconnaissance, credential harvesting, privilege escalation lateral movement activities. The session will demonstrate how attackers using tools such as BloodHound, PowerView, SharpHound, and other Active Directory enumeration frame with deceptive objects and inadvertently reveal their presence.
The presentation also examines the operational and investigative value of deception technologies, including alert generation, threat hunting opportunities, forensic evidence collec response support. Participants will gain insight into how deception can improve visibility into attacker behavior while reducing false positives commonly associated with traditional r solutions.
Attendees will leave with practical guidance on designing, deploying, and maintaining deception-based defenses within Active Directory environments, enabling organizations to st resilience, improve early threat detection capabilities, and better protect critical identity infrastructures against modern cyber threats.

Ahmed Hassan is an Austrian Cybersecurity Expert, Penetration Tester and Security Researcher with more than 8 years of experience in offensive and defensive information secu focuses on identifying, analyzing and mitigating complex security risks across enterprise environments, government organizations and critical infrastructure sectors.

Throughout his career, he has conducted numerous security assessments, penetration tests, red team engagements and comprehensive security evaluations for both national and organizations. His experience spans a wide range of industries, including government and public sector institutions, financial services, telecommunications, energy providers, heal aviation, technology companies, critical infrastructure operators and international organizations.

Ahmed Hassan has discovered and responsibly disclosed security vulnerabilities affecting renowned organizations and companies, including the United Nations (UN), Oracle, SAP numerous government institutions worldwide. His research has resulted in more than 53 assigned CVEs (Common Vulnerabilities and Exposures), contributing to the improvement cybersecurity standards and practices. In addition to his technical expertise, he holds several internationally recognized cybersecurity certifications, including OSCP (Offensive Sec Professional), OSWA (Offensive Security Web Assessor), CRTP (Certified Red Team Professional).

As a recognized speaker, Ahmed Hassan regularly presents at international cybersecurity conferences and shares his expertise with the global security community. His speaking e include prominent events such as Black Hat MEA (Saudi Arabia), Hack Red Con (USA), GISEC 2025 (UAE), Arab Security Conference (Egypt), Arab Internation Security Conferer BSIDES Vienna 2026 (Austria) and various other international cybersecurity conferences and industry forums and universities.

Furthermore, he has represented Austria at several high-profile cybersecurity and technology events across the world. These engagements were conducted in collaboration with a support of Austrian embassies, promoting Austrian cybersecurity expertise while fostering international cooperation, knowledge exchange and strategic partnerships.

Recommendations from previous labs and students/attendees are available on Ahmed Hassan's LinkedIn profile and post (https://www.linkedin.com/feed/update/urn:li:activity:7464031469158477824/).
Ahmed Hassan is on every social media platform active and would be more than happy to be connected with all of you. To find out about the latest in his work, you are invited to co on LinkedIn, Facebook, Twitter and subscribe to his YouTube Channel where all his talks, presentations and event talks are uploaded. 🙂

YouTube: https://www.youtube.com/@titaniumcybersecuritysolut3431
Twitter https://x.com/hassanahmed8199/
LinkedIn: https://www.linkedin.com/in/ahmed-hassan-79559487/
Facebook: https://www.facebook.com/ahmedhamada81/

Vishing At Scale: Humans vs. Voice Agents

Indy Mellink (Doppel)

Abstract:
Over the past several years, I've run more than a thousand human-to-human vishing calls. Now I'm conducting the same operations with fully automated voice agents, at scale, ac government targets.

This presentation is the before/after: humans versus voice agents in real operations. Not a controlled academic study, that's not how internal red teams work, but a breakdown of w changes when voice becomes software.

I'll demonstrate the operational shifts that voice agents introduce: persistence, parallelization, and the ability to chain SMS or email mid-call to push targets through password rese and sensitive data disclosure.

But the picture isn't one-sided. I'll define where agents outperform humans today, and where humans still win.

Defenders will leave with actionable guidance for tuning voice-based security exercises and hardening the workflows that make this threat possible at scale.

Full outline:
1) Open: the current threat landscape (5 min)
- The escalation of voice phishing techniques by attackers over time
- Voice agent attacks in the wild by threat actors

2) Human vishing at scale (10 min)
- What human-led vishing looks like in enterprise reality: variability, fatigue, skill ceiling
- The moments that drive outcomes and the moments that save the day
- High-level outcome stats: compromise, reporting, shutdown

- What actually increases reporting/shutdown in employees (and what doesn't)

3) Agentic voice at scale (12 min)
- What's changed operationally when the caller is software
- Capabilities that matter to internal teams:
- Persistence (callbacks/retries, timing optimization)
- Parallelization (no labor bottleneck)
- Consistency (script adherence, no drift/fatigue)
- Multi-vector orchestration (SMS/email/messaging while on-call)

4) Comparison: Humans vs. Voice Agents (10 min)
- Where agents outperform humans today (scale + persistence + orchestration)
- Where humans still win today:
- interruptions/barge-in and conversational chaos
- emotional pivoting when challenged
- deeper elicitation/influence
- company-specific nuance + edge cases
- The capability scoreboard you can take back and apply to your own program

5) Takeaways Notes (3 min)
- A post-human vishing threat model grounded in real operations
- A practical capability gap matrix (humans vs agents)
- Concrete hardening priorities (password resets, MFA resets, data disclosure)

With a background in neuroscience and forensic psychology Indy has a good understanding of the criminal mind. Applying her knowledge as a social engineer, she now runs large simulations for enterprise and government organizations including (but not limited to) vishing, phishing, deepfake scenarios and physical pentesting.

From Scanner to Full Account Takeover: Chaining Subdomain Takeover with OAuth Wildcard Redirects

Shira Ziv (Cyera)

Infrastructure drift in SaaS environments creates attack paths that manual review consistently misses. This talk presents a case study in which AI-assisted continuous reconnaissa dangling DNS CNAME record pointing to a decommissioned hosting provider project and demonstrates how that infrastructure finding chained into a full OAuth account takeover.

The research is built on four purpose-built scanners running weekly via GitHub Actions, using an LLM (Claude via AWS Bedrock) as a reasoning engine for week-over-week infras analysis. The scanners cover subdomain takeover detection, external attack surface mapping, static code analysis, and AI/MCP configuration audit. When the subdomain takeove an abandoned hostname, I claimed it through the hosting provider's domain assignment flow and discovered it also matched a wildcard OAuth redirect_uri pattern on the applicati server.

By combining the subdomain takeover with an attacker-initiated OAuth 2.0 authorization code flow, I demonstrate a complete account takeover path. The victim authenticates throu identity provider with MFA. The authorization code is silently delivered to attacker-controlled infrastructure. The attacker completes the PKCE token exchange using a verifier they operates as designed in this scenario but cannot compensate for an attacker-controlled redirect endpoint: the attacker initiates the flow, generates the verifier and challenge, and r authorization code at infrastructure they control.

The result: access tokens, ID tokens, and refresh tokens — persistent access including roles, tenant context, and admin privileges — without phishing credentials, exploiting zero- malware.

Outline (45 + 5 min):

Introduction & Problem Statement (5 min)
Scanner Architecture (8 min) — four modules, GitHub Actions orchestration, Claude/Bedrock as reasoning engine, week-over-week diffing, deterministic probes first then LLM prio
The Finding & Subdomain Takeover (7 min) — dangling CNAME, provider verification gap, SSL provisioning, disclosure context
The OAuth Chain (12 min) — wildcard redirect_uri, public client with PKCE, attacker-initiated flow, why PKCE cannot compensate, token harvest, victim experience
Live Demo (5 min) — end-to-end on controlled infrastructure
Defensive Recommendations (5 min) — DNS cleanup, exact redirect_uri allowlists, cross-team chain modeling
Q&A (5 min)

Shira Ziv is a Security Researcher at Cyera, specializing in offensive security, red teaming, and cloud security research. Previously, she conducted security research on Azure clou Microsoft.

Damage Report: What Breaks When LLM Agents Run Your SOC

Aysel Rahimli (SOCAR Tech (Caspian Innovation Center))

We gave LLM agents real access to a working SOC (the systems teams actually use every day for detection, endpoint response, threat intel, and email security) and let them inve in context, and take action, always with a human approving the final step. This talk is my honest report of what went wrong.
Some of it was uncomfortable. Agents told us an action had succeeded when it never ran. They claimed to have analyzed "all events" after seeing only the first page of results. Wh didn't exist, instead of saying so, they quietly picked a similar-looking one and carried on. Sometimes the framework itself failed and printed a tool call as plain text instead of execu single "proceed" from a user could send the whole thing into a loop.
For each one, I'll show the fix that actually held, none of it tied to a particular product or model: a gateway between the agents and the tools, rules the agents can't talk their way ar whether something exists outside the model instead of trusting what it says, and reusing data we already fetched instead of hammering the same API again.
This isn't a product pitch and it names no vendors. It's the architecture patterns, the real failure data, a few short recorded demos, and an honest look at the problems we still have

Aysel Rahimli is a Junior AI Security Researcher at SOCAR's Caspian Innovation Center and a final-year Process Automation Engineering student at Baku Higher Oil School. Her secure multi-agent AI orchestration for security operations, including agent coordination, MCP-based tool integration, human approval mechanisms, and the reliability of AI-assiste She has contributed to research on artificial intelligence, deepfake detection, and distributed multi-agent systems, and has presented her work at international technology events.

Reverse Engineering Industrial Control System Malware

Diyar Saadi (Independent Researcher)

This presentation explores how malware targeting Industrial Control Systems (ICS) and operational technology environments can evolve from a digital intrusion into a real-world op
The session begins with an overview of how ICS environments function, why they are critical to daily life, and the security challenges created by legacy systems, increasing conne
convergence.

The talk then transitions into advanced malware tradecraft used against industrial targets, including stealth execution, process manipulation, memory-resident activity, and staged i
techniques. Attendees will follow a practical reverse engineering and investigation workflow focused on understanding attacker behavior, identifying malicious execution chains, an
activity, and correlating host and network artifacts.

Designed for defenders, reverse engineers, and incident responders, the session provides practical insight into how industrial malware operates and how security teams can impro
investigation, and response capabilities within critical infrastructure environments.

Diyar Saadi Ali is a cybersecurity expert specializing in cybercrime investigations, SOC operations, and malware analysis. A certified MITRE ATT&CK Contributor and CVE discove
CVE-2024-25400 and CVE-2024-25399 and CVE-2026-20841), Diyar helps organizations defend against evolving digital threats. They have spoken at major international events i
BlackHat MEA in Saudi Arabia, DeepSec in Vienna, COSAC in Ireland, VulnCon in India, and Arab Cyber Security Conference in Egypt. Diyar's expertise and global experience co
and lead in the cybersecurity community.

Like the back of your hand…

Mattias Jacobsson (Kryptogram AB)

How well do you think you know your environment? Are you prepared for, and know what to do when an incident comes along? If not, a good idea would be to get ready by practic

This presentation gives you a glimpse into how you can take advantage of CDXE when building yourself a realistic exercise environment for your company's cyber defence exercis
is an all-in-one internet-like environment that helps you setup the internet part of your air-gapped exercise environment with low resource requirements. We do also get into other u
how teachers could use CDXE to help students learn about the inner working of the internet.

Expect a wild ride into the unknown of virtualizing the internet and how that can be useful for a wide range of needs. For more information of the open source project CDXE, see
https://kryptogram.se/cdxe

Mattias has been working with computers from an early age. Today, he is running his own cyber security company, helping clients with their security needs and pushing for better C
training opportunities for security teams.

Blind Spots in the Software Supply Chain: What Security Scanners Miss

Ofri Ouzan (JFrog)

Modern software supply chain (SSC) security relies heavily on automated security controls. Organizations trust SBOM generators and Software Composition Analysis (SCA), secr
malware scanners, reachability analysis tools, and recently also AI security scanners, to identify risks before software reaches production.

But what happens when the assumptions behind those tools are wrong?

Over the past several years, our research focused on identifying fundamental blind spots across multiple application security categories. Rather than targeting a specific vendor, w
entire classes of security solutions make assumptions about where software, vulnerabilities, secrets, malware, and risks exist. We will present how these blind spots manifest in ev
popular software components, in addition to attacker-crafted components.

In this talk, I will present real-world examples of blind spots we discovered, including software that completely disappears from SBOMs because it was installed outside traditional
managers, secrets hidden in locations ignored by many secret detection solutions, malicious content embedded in ways that evade modern malware scanners, and vulnerability pa
bypass common static analysis assumptions.

More importantly, I will demonstrate a methodology for identifying blind spots in security products, explain why these gaps continue to exist across the industry, discuss emerging a
agents and the new assets they bring with them, including next-generation SSC security tools where similar weaknesses are likely to appear.

Attendees will leave with a practical framework for evaluating the security controls they rely on today and understanding where attackers are most likely to operate tomorrow.

Ofri Ouzan is a Security Researcher and speaker at JFrog with over five years of experience in cybersecurity. Her research focuses on identifying emerging threats and security is
software supply chain, as well as developing open-source tools and automation solutions that help organizations address these risks.

Your Traffic Is Lying to You - How Layer 7 Attacks and Bot Campaigns Hide in Plain Sight

Lisa Froehlich (Link11 GmbH)

Bots generate more than half of all global HTTP traffic. Most of these bots are designed to appear legitimate. This talk examines how modern Layer 7 attacks and bot campaigns c
every threshold your monitoring system is calibrated for and explores what actually stops them.
The core problem is not technical sophistication. Rather, it is the absence of a behavioral baseline. Organizations that have never profiled their legitimate traffic cannot distinguish
from a demand spike. Traffic volume, unique IP counts, and request rates have become unreliable indicators, and attackers know it.
This talk presents four real attack scenarios from production network operations, traces the available but missed detection signals, and explains what behavioral detection must do
short. All cases were mitigated, making the defense decisions and detection timelines analytically useful.

Lisa Fröhlich is Head of Corporate Communications at Link11, a European provider that specializes in DDoS mitigation and application-layer protection. With a background in scie
communication and over a decade of experience in technology public relations (PR), she translates complex threat data into actionable insights. She works closely with the Link11

on observed incident data from the company's production network operations. Since 2023, she has hosted the IT security podcast Follow the White Rabbit. In 2025, she completed cybersecurity program at the Eurobits Women Academy.

Genericpocalypse: Fixing macOS Malware Identification with ML

Kseniia Yamburh & Nazar Grycshuk (Moonlock by MacPaw)

The macOS threat landscape is growing fast, yet identification lags behind: roughly 70 % of macOS samples on VirusTotal carry labels like “Generic Trojan” or “Generic Agent”, giv almost nothing actionable to work with. Manual triage can drag on for days, and sandbox evasion by modern macOS malware makes automated analysis even harder.
In this talk we present a practical, layered approach that turns “Generic” into something useful. First, we introduce The Eye - an in-house ML-powered tool that clusters macOS bin combining byte-level patterns, entropy, Symhash, Ssdeep, Bundle IDs, Team IDs, and a custom YARA-like ruleset. The clustering engine groups similar samples automatically and by potential danger, prevalence, so analysts immediately know what to investigate first.
Second, we unveil SWARM - a Mach-O feature-based detection ruleset mapped to macOS MITRE ATT&CK techniques. SWARM fires on behavioural indicators before AV vendors label, enabling proactive threat hunting rather than reactive response. We will show real-world cases where SWARM flagged active threats weeks before VirusTotal detections app We close with the guidance for integrating ML clustering into existing threat-hunting workflows, and an honest discussion of current limitations and the road ahead.

Kseniia is a macOS security researcher at Moonlock by MacPaw, where she focuses on malware detection, threat hunting, and threat intelligence. She has spent several years an specific malware families, working on attribution and raising awareness about macOS threats.

I Own Your "Internal" Domain… Weaponizing Name Collisions at Scale

Philippe Caturegli (Seralys)

What happens when your internal domain name is actually registered and operated by someone else? You start leaking credentials, email gets misrouted, deployment pipelines pu untrusted places, and most of the time, you don't even notice for years.

With over 1,200 new TLDs introduced over the past decade, many organizations are unknowingly using internal domain names that now exist as valid, publicly registerable FQDN own and never thought to claim.

Name collision itself is not new. It has been documented as a theoretical risk since 2013. But no one has studied it at scale, or answered simpler questions: how could threat actor potential collision and what would be the impact if they claimed a colliding domain and operated it ?

We analyzed nearly 39 million SSL certificates and over 9.5 million services to surface internal names leaking onto the public internet, found more than 92,000 colliding domains th registered. We then decided to claim 200 of them. Over two years, those domains passively recorded more than 11 billion DNS requests. And when we started answering some of web requests, email, credentials, auth tokens, build artifacts and much more started flowing in.

The affected organizations ranged from police departments and airports to Fortune 500 companies, enterprise software vendors, and critical infrastructure. We'll walk through how were discovered, what happened after claiming the domains, why organizations kept leaking data for months/years, and what happened when we tried to report it. Between radio s threats, the disclosure process turned out to be almost as revealing as the technical findings themselves.

Philippe started his hacking career in the 1990s with a multi-node hacking/phreaking BBS in his parents' basement. Over the past 30 years, he has worked across all sides of secu large enterprise networks, advising financial institutions, and ultimately following his passion for breaking things as a penetration tester and vulnerability researcher.

Ghost Approvals: Decision-Provenance Integrity Failure in Agentic AI Frameworks

Yuvraj Pradhan (MIT ADT University)

A LangGraph workflow blocked a $100,000 wire transfer. We ran one SQL statement. The agent executed it anyway. The audit log records J. Smith's approval at 09:14:32Z. J. Sm consulted.

Threat model: attacker has write access to checkpoint storage, isolated from application code and model execution, representing a lower-privilege persistence boundary.

We examined eighteen production agentic AI frameworks against a single architectural question: does the system have any mechanism to distinguish a human approval event fror Fourteen do not. Four do, because they never serialize the approval decision to persistent storage at all. We call this Decision-Provenance Integrity failure.

The attack lives entirely below the model layer. No prompt, no jailbreak, no API call. A conditional branches on attacker-controlled state and the human oversight guarantee dies th time. In twelve of the fourteen vulnerable frameworks, the entire payload is plain JSON. In two of them, Haystack and PydanticAI, the attack does not forge an approval flag. It forg arguments themselves. The attacker does not change who approved the action. They change what the model called the tool with.

The four frameworks that are not vulnerable made no deliberate security decision. They simply never serialized the approval event to storage. You cannot forge what was never wr design choice costs nothing. Every framework that failed made the opposite choice without ever examining it as a security boundary.

Microsoft Security Response Center confirmed the finding accurate and classified checkpoint storage integrity as outside their servicing boundary. Coordinated disclosures are file affected vendors. No vendor has disputed the technical characterisation.

EU AI Act Article 14 enforcement is now active. Current audit methodology asks one question: is there a log entry showing human approval? The attack answers yes, definitively, w and a named approver, while the human event never occurred.

The forged trail is not merely undetected. It is certifiably compliant.

Yuvraj Pradhan is an undergraduate student at MIT ADT University specialising in AI systems security. First author of a Springer Nature Q2 publication, he discovered Decision-Pr failure, a vulnerability class enabling audit-clean forgery of human approvals across fourteen production agentic AI frameworks. confirmed the technical characterisation is accurat disclosures have been filed with all affected vendors including OpenAI, Microsoft, and Google. No vendor that has responded has disputed the finding.

Security, Trust & Privacy by Design for Smartphone Solutions

Kartik Lalan (PIC (Independent))

Abstract: Lacking security standardization in terms of usage policies, smartphones solutions are being poorly designed compromising end user's data. Also lack of thorough risk m
developing such solutions leads to uncalculated risks emerging out of the platform. In this talk we will be discussing on drawing risk boundaries, so that making any Smartphone ba
manufacturers and end users can achieve real data security and transparency. For instance, Very often we see solutions using additional fingerprint reader capturing raw images o
despite beyond the promise of OS security e.g. user national identity program, using location service for security & safety solutions but those can be manipulated even w/o jailbrea
device e.g. Covid'19 apps or banking or logistics etc., Payment POS without kiosk hardening which allows anyone use it fraudulently to phish the card user, apps made to abuse p
to steal data officially, etc. With loosely coupled OEM security policies, it reduces the overall data security & privacy posture e.g. internet, permission model, privacy. Not being clas
vulnerabilities these never catch attention and are openly misused.

Product Security Architect @ Security Centre of Excellence – Philips Innovation Campus. He is M.Tech. in CS with Specialization in Information & Network Security. He conducts f
workshops on Info Sec topics @ several places including OWASP AppSecBLR, Nullcon, C0C0N, Defcon BHV, Bsides & Null Chapters, DroidCon-IN. Kartik loves to write technica
leisure time.

Contact General Terms & Conditions Anti-Harassment Policy Privacy Policy