TARANIS_AI
https://taranis.ai/

AUSTRIAN INSTITUTE OF TECHNOLOGY
TOMORROW TODAY

TARANIS AI – ADVANCED OSINT ANALYSIS WITH NLP AND AI

IT-SECX

Florian Skopik, Benjamin Akhras, Peter Leitmann, and Lukas Linauer
Center for Digital Safety and Security
AIT Austrian Institute of Technology

St. Pölten, Oct. 3rd, 2025

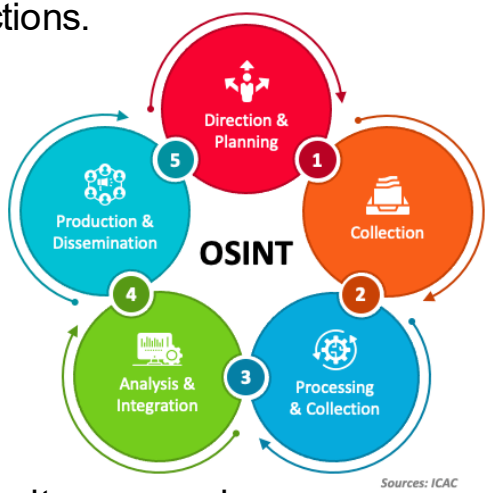
WHAT IS CYBER SECURITY OSINT?

- Open-source intelligence (OSINT) is the collection and analysis of data gathered from open sources to produce actionable intelligence.
- Technical Cyber Threat Intelligence (CTI) to configure detection systems:
  - Indicators to put into SIEMs
  - Domains to block in name-servers or proxies
  - Execution patterns to block in EDRs
- But also “soft” CTI:
  - News about threat actors
  - New (features of) security products
  - News about breaches, incidents, campaigns
  - News about vulnerabilities, patches, mitigations, counter-measures, exploitation, post-exploitation, …
  - Policy news: political/diplomatic initiatives, new EU policy documents, GDPR-related lawsuits
  - Updates on security standards (ISO, BSI, ANSI, CIS, OWASP, …)
  - Mergers, acquisitions, failures, … or other company news

Baseline today of one of our national stakeholders: approx 250 sources, up to 500 articles per day, more after longer weekends or large-scale events

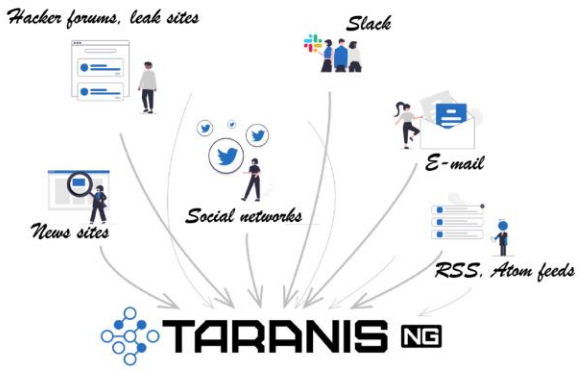
WHAT IS OSINT GOOD FOR?

- Gather public information on potential security threats, vulnerabilities, trends, attacker TTPs, new risks etc. to maintain situational awareness and take early counter actions.
- Input for products
  - Advisories
  - Summaries (daily, weekly)
  - Situational reports, white papers, fact-sheets
- Awareness / Preparedness
  - Consulting / Answering calls for help
  - Media inquiries
  - “Boss/CEO/Politician asking questions”
  - Trigger for proactive activities
- Challenge: Number of OSINT sources is high and the number of news items massive
  - Grasp quickly what’s relevant and omit the rest
  - Filter repetitive content
  - The workflow is actually pretty similar to a journalists work



TARANIS AI

- Based on *taranis3** and *taranis-ng***
  - Great tools to ingest raw unstructured data from various sources
  - Use human knowledge to identify relevant information
- Preserves the “taranis workflow” many CERTs are used to
  - Gather -> Assess -> Analyze -> Publish
- Introduces natural language processing (NLP) capabilities
  - Extraction of relevant **named entities**
  - **Clustering** of related **news items**
  - **Summaries** of “story clusters”
  - **Recommendations** of news items
  - Support for **creating OSINT products** (“reports”)



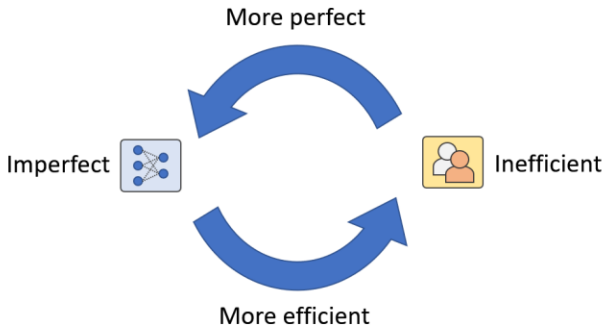
TARANIS_AI
https://taranis.ai/

* https://github.com/NCSC-NL/taranis3

** https://github.com/SK-CERT/Taranis-NG

HUMAN-IN-THE-LOOP APPROACH

- Combine automation with analyst oversight to produce publish-ready intelligence products.
  - “human in the loop”
- **AI** and automation for scaling
  - Simultaneous news item ingestion from hundreds of feeds
  - Deduplication of news items
  - Story clustering across heterogeneous feeds
- **Human** for maintaining trust
  - **Relevance**: How relevant is the data to your needs?
  - **Reliability**: Is the source trustworthy?
  - **Timeliness**: Is the information up-to-date?
  - **Impact**: What is the potential impact of the information?



Source: https://bair.berkeley.edu/blog/2022/05/03/human-in-the-loop/

APPLYING THE POWER OF NLP/AI ON OSINT

User Story 1: *What are the 'hot topics' of the last hours/days?*

User Story 2: *What do we know about a specific entity? (e.g., a vulnerability, malware, company, product, person, location etc.)*

User Story 3: *How can I find more related news items after reading this interesting article?*

User Story 4: *Which news items are pertinent to my organization/infrastructure?*

User Story 5: *How can I efficiently sum up my findings for my constituency/manager?*

TARANIS AI – IN ACTION

AUSTRIAN INSTITUTE OF TECHNOLOGY

TARANIS AI

Administration | Dashboard | Assess | Analyze | Publish | Assets

Dashboard

Edit Dashboard

Assess

There are 78936 total news items.
There are 82916 total story items.

Analyze

There are 4 completed analyses.
There are 13 pending analyses.

Publish

There are 9 products ready for publications.

Connectors

There are **0** conflicts detected.

Trending Tags (last 14 days)

CVE_VENDOR — Cluster size: 9

Apache	3
Apple	2
Dell	2
AMD	1
BlueZ	1

Cybersecurity — Cluster size: 17

Denial of service	15
SQL Injection	1
zero-day	1

Product — Cluster size: 296

iPhone	92
iPhone 17	58
Windows	51
ChatGPT	48
iOS	47

Organization — Cluster size: 753

Google	288
Apple	159
ZDNET	149
Microsoft	90
Samsung	67

Location — Cluster size: 297

US	70
USA	67
China	61
Deutschland	56
Europa	43

Person — Cluster size: 59

Donald Trump	26
Sam Altman	9
Mark Zuckerberg	9
Putin	8
Marco Rubio	7

TARANIS AI – IN ACTION

AUSTRIAN INSTITUTE OF TECHNOLOGY

TARANIS AI | total stories: 26 / displayed: 20 | search ctrl+k | Dashboard | Administration | Assess | Analyze | Publish | Assets

search

Items per page
20

Source

Source Group

Source

Filter | more details

First Day

Last Day

Tags

read

important

in reports

relevant

laptop switches? It's called the mikefive because it's 5 mm thick. That's okay, though, because now you're caught up and I can talk about his latest keyboard, the mikecinq. The inspiration for this one includes the aesthetics of Le Chiffre and the slimness of Le...

Published: 2025-05-19 00:00
Tags: CNCF
Article: redhat.com

Zero trust workload identity manager now available in tech preview

Non-human identities—also known as machine or workload identities—are becoming increasingly critical as organizations adopt cloud-native ecosystems and advanced AI workflows. For workloads spanning multiple cloud platforms, adhering to zero trust principles becomes challenging as they cross identity domains. A unified identity framework provides consistency in automating identity issuance and enforcing access control policies across diverse environments. SPIFFE/SPIRE, an open source identity...

Published: 2025-05-19 00:00
Tags: IBM, Red Hat, Kubernetes, HashiCorp, OpenShift
Vote: 0 0
Article: redhat.com
Author: Kirsten Newcomer; Nick Png

How HashiCorp Vault and Red Hat OpenShift can work together

In hybrid and multicloud environments, proper management of sensitive data-like secrets, credentials and certificates is critical to maintaining [illegible] across Kubernetes clusters. While Kubernetes provides a Kube-native way to manage [illegible] that Kubernetes secrets are not particularly secret: they are base64 encoded and are accessible to cluster admi[illegible]y, anyone with privileges to create a pod in a specific namespace can access the secrets for that namespace. [illegible]tion can be provided by encrypting sensitive data stored in etcd, even stronger protection is provided by integrating an external secret manager with Red Hat OpenShift. HashiCorp Vault delivers enhanced security capabilities to Red Hat OpenShift by providing centralized secret management across a hybrid and multicloud Kubernetes infrastructure.

As customers rapidly embrace cloud-native architectures and AI-infused workloads, complexity can invariably arise across increasingly hybrid environments, with respect to both management and security. The reality is that secrets–credentials that provide access to or within a system–are becoming more distributed and are at increased risk of secrets sprawl. Secrets such as usernames and passwords, TLS certificates, API tokens, database credentials and more are often stored insecurely, creating risk for data breaches, compliance issues and identity theft.

For hybrid and multicloud workloads, this challenge is exacerbated. Enterprises need consistent and scalable ways to manage their secrets across their lifecycle, and importantly, seek a solution that complements existing hybrid cloud application management. Red Hat OpenShift provides a comprehensive and consistent application platform trusted by thousands of customers for building, modernizing and managing traditional and cloud-native applications across any infrastructure. HashiCorp Vault with Red Hat OpenShift will enable more robust management of secrets at rest and in transit, comprehensive auditing, rich access control lists, with integrated support for multiple authentication methods across different cloud vendors, and dynamic secrets.

add to report | mark as read | mark as important | deselect | Stories selected: 1

Search

NER

Sharing Options

Source

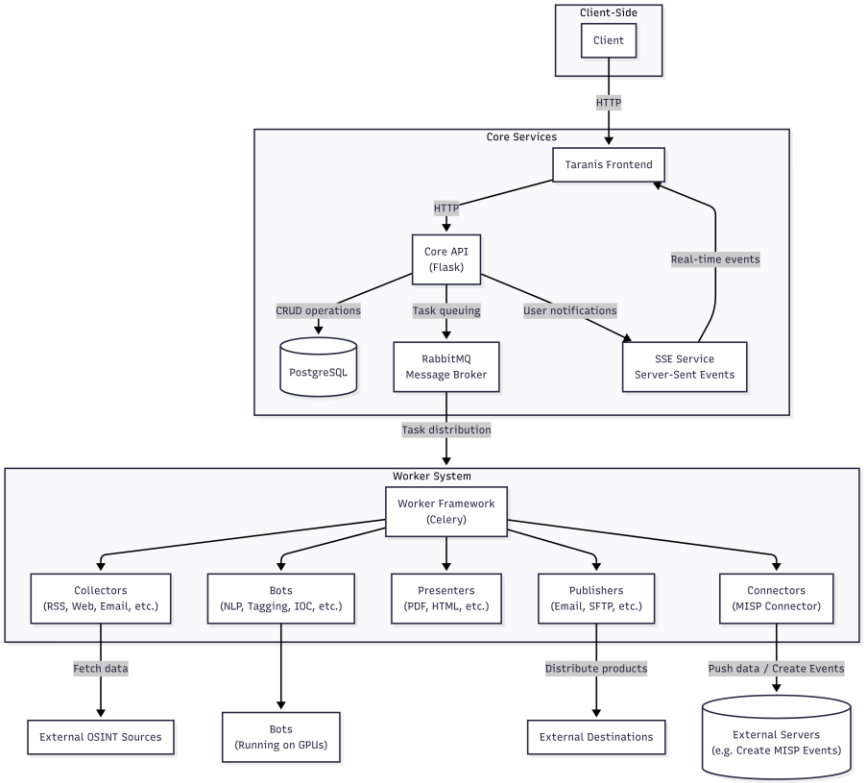
Advanced Filters

Summary

ARCHITECTURE

- **Event-driven processing** with RabbitMQ & Celery
- **Modular & extensible** design for OSINT workflows

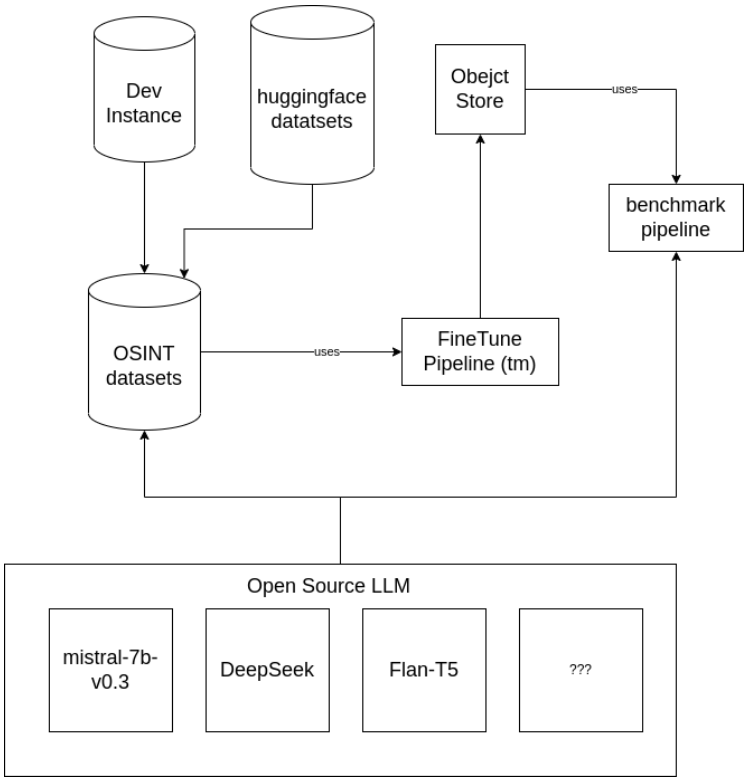
- **Collectors** fetch OSINT data
- **Bots** analyze & enrich (NLP, tagging, IOC)
- **Presenters, Publishers & Connectors** distribute results



FROM ARCHITECTURE TO MODELS

- Modular architecture allows adding new bots.
- Multiple models supported for each bot
  - Context Length
  - Hardware requirements
  - Language support
- Open models

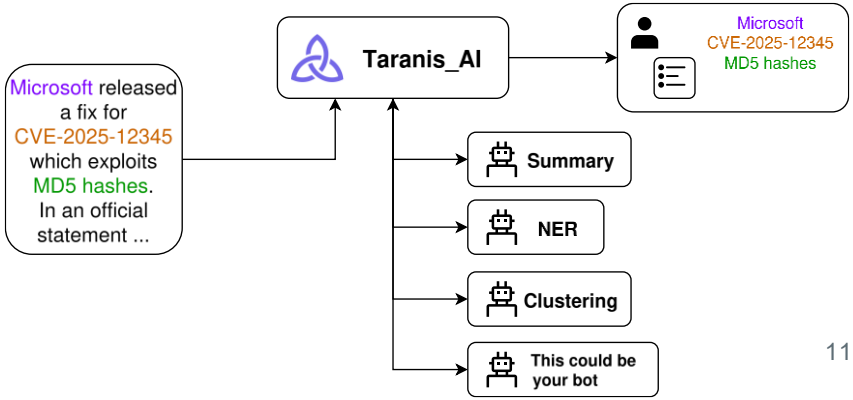
🤗 https://huggingface.co/llinauer/gliner_de_en_news



BOTS IN TARANIS AI

- **Named Entity Recognition (NER)**
- Summarization of lengthy news items/stories
- Classification (topic, relevance, sentiment)
- Story clustering (finding related articles)
- . . .

Natural Language Processing



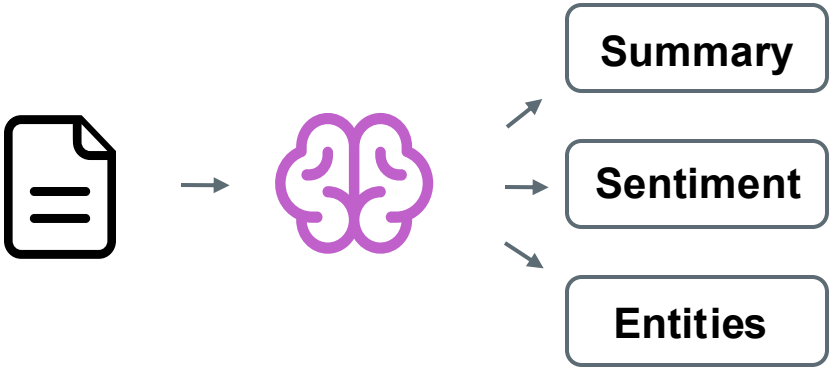
NATURAL LANGUAGE PROCESSING (NLP)

What is NLP?

- Automatic processing of natural language

Why is NLP hard?

- Unstructured data
- Ambiguity & context dependence
- Multilingual challenges
- Cybersecurity domain specifics



Concrete Example → NER

- Extracting people, orgs, places from text

NAMED ENTITY RECOGNITION (NER)

NER EXAMPLE (1/2)

"As a national CERT, one of our extremely important tasks is to proactively inform network operators about potential or confirmed security issues that could affect Austrian companies. […]. In Austria alone, shodan.io reports approximately 1.7 million devices accessible online [1]."

- Cert.at report from June 10th, 2024

Austria = **Austrian** ?

NER EXAMPLE (2/2)

"When President Donald Trump signed the so-called "Big, Beautiful Bill" on July 4, dedicating $45 billion to immigration detention with a goal to double or triple the population behind bars, it was a huge payoff. The victory was in the works for years. A private prison company handed consulting and lobbying gigs to Trump's allies, its political action committee was the first to max out its donation to Trump, and industry executives had already made plans to reopen shuttered prisons — laying the groundwork for what they promised investors would be an incarceration bonanza."

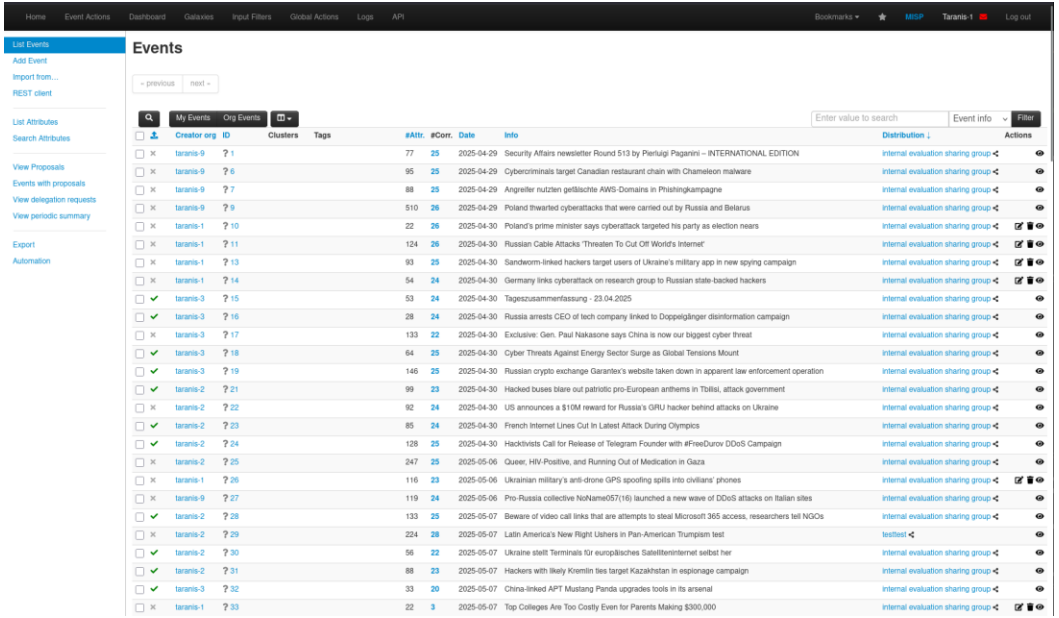
- theintercept.com from July 10th, 2025

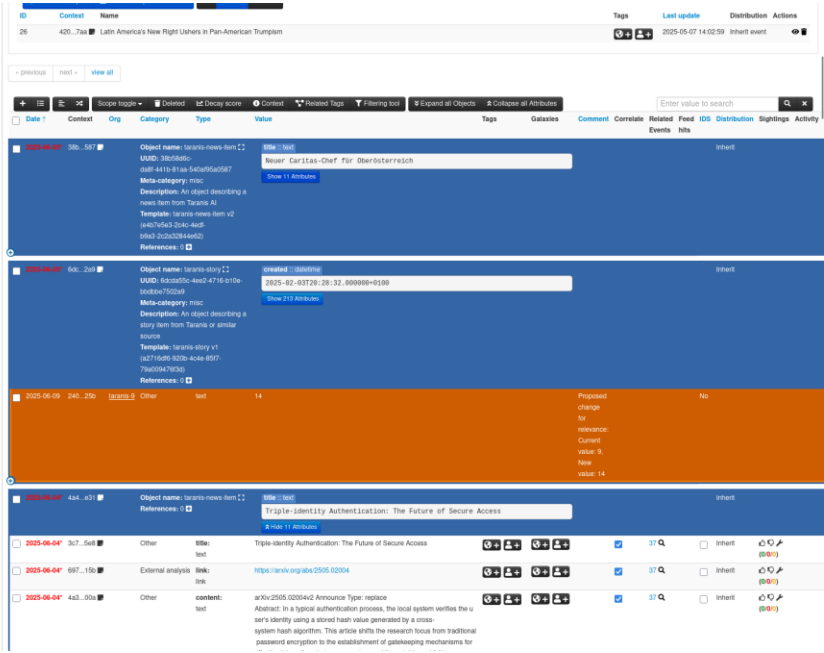
Donald Trump = **Trump** ?

SHARING & COLLABORATION OF STORIES

- Collaboration on the same Story over multiple Taranis AI instances
- Collaboration with MISP maintainers at CIRCL
- STIX-compatible data exchange enables integration with external tools and platforms for bi-directional intelligence sharing.
- Contributed to MISP – New MISP Objects
  - Taranis Story and News item objects
  - Allows collaboration using MISP only

COLLABORATION – MISP VIEW

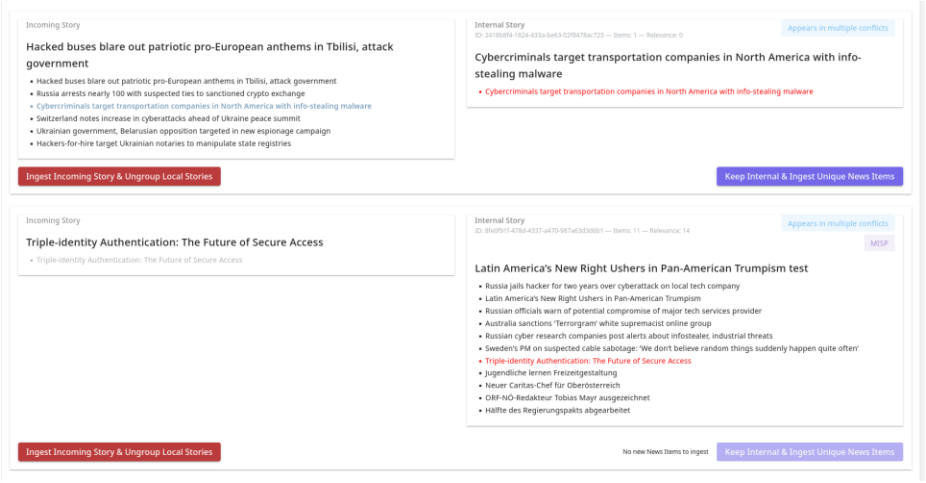




COLLABORATION CHALLENGES – CONFLICTS

- When conflicts arise?
  - Local changes to a shared story
  - Different sources
    - Ingestion of duplicated News Items
    - Partial overlaps of Stories

- Conflict resolution workflows for CTI analysts
  - Corrupting the story context vs. flexibility for CTI analysts
  - What is missing for CTI analysts?
    - Special view with filters (focused view)
    - Automated ignoring (aging problem)



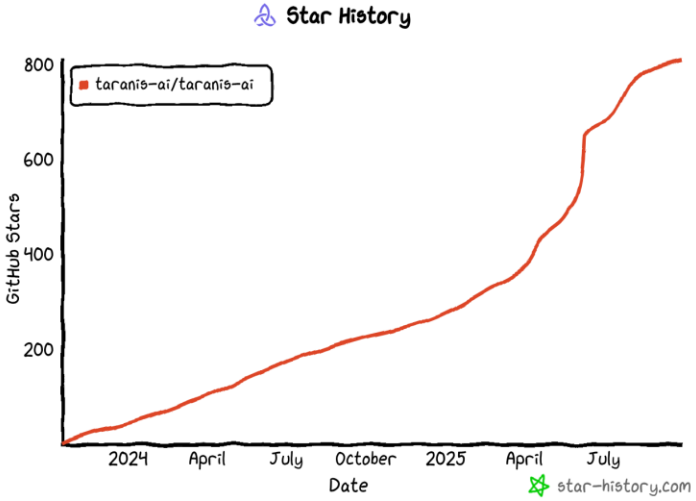
DEMO

Demonstration of a Taranis Workflow including collaborating with a partner via MISP

taranis.ai/2025_10_1_taranis-walkthrough-with-misp.webm

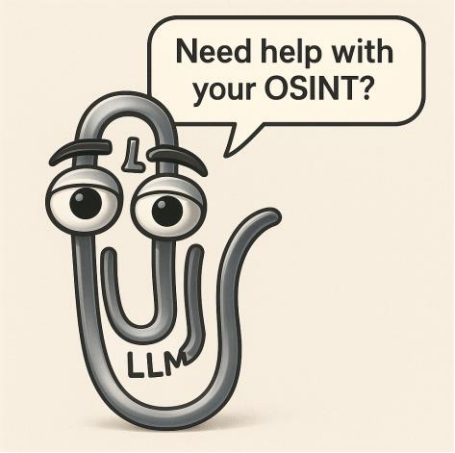
CURRENT STATUS

- Production-ready OSINT stack: REST interface, Celery workers
- Deployable reference architectures covering Docker Compose, Kubernetes
- Proven workflows and extensive documentation
- Fully customizable templates



FUTURE PLANS

- LLM based AI Assistants that understand analyst workflows and context, providing real-time support.
- Link OSINT with internal CTI and analyst feedback to build a living knowledge graph of threats.
- Recommender System suggesting stories to Analyst for sharing or interaction.



FINANCIAL SUPPORT AND COLLABORATIONS

- The work in this presentation received funding from the Connecting Europe Facility (CEF) program in course of the project **AWAKE** (2020- AT-IA-0254).
- The project has additionally received funding from the European Union - European Defence Fund under GA no. 101121418 (**EUCINF**), no. 101121403 (**NEWSROOM**), and no. 101168092 (**ECYSAP EYE**) Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union. The European Union cannot be held responsible for them.

- Active collaboration with CERT.at, Austrian Ministry of Interior and Austrian MoD
  - Early adoption of taranis-ai by analysts
- Exchange with SK CERT, the developers of taranis-ng
- Open Source project on Github
  - Opportunity to contribute: https://taranis.ai/

taranis-ai Public

Taranis AI is an advanced Open-Source Intelligence (OSINT) tool, leveraging Artificial Intelligence to revolutionize information gathering and situational analysis.

Python 809 101

TARANIS_AI

https://taranis.ai/

AUSTRIAN INSTITUTE OF TECHNOLOGY

TOMORROW TODAY

THANK YOU!

Please contact:

Florian Skopik

taranis@list.ait.ac.at

github.com/taranis-ai/taranis-ai

Oct. 3rd, 2025