Autonomous LLM agents for cybersecurity and more

Roberto González (NEC Labs Europe)
Roberto.Gonzalez@neclab.eu

NEC Corporate Profile

Leading social value innovator

Company Name: NEC Corporation

Established: July 17 1899

Provides telecommunications, AI solutions, Public solutions, Enterprise business solutions…

- Headquarter: **Tokyo, Japan**
- **125 years** of brand success
- **105,246** employees worldwide
- **$22 billion** net sales
- **254** consolidated subsidiaries
- Footprint in **169** countries and territories

THOMSON REUTERS

FORTUNE 500

- **Headquarter**: **Tokyo, Japan**
- World's **top 100** most innovative organisations
- Fortune **500** global company

NEC Laboratories Europe (Heidelberg, Germany)

Advancing Information & Communications Through Research Excellence

KEY R&D METRICS

1,350+ Patents

50+ Peer-reviewed publications per year

140+ European projects

40+ University Cooperation's

OPERATIONAL AREAS

RESEARCH
Leading scientific discovery in Europe

TECHNOLOGY TRANSFER
Commercializing R&D results in existing and new company business segments

STANDARDS
Defining European technology standards and best practices

RESEARCH AREAS

DATA SCIENCE

- Biomedical AI for infectious diseases and immuno-oncology
- Explainable AI & human-centered AI
- Large language models & natural language processing
- Machine learning for computational science
- Multi-modal & relational data (graph)

SECURITY

- Decentralized trust
- Blockchain privacy, security & scalability
- Secure confidential computing

5G & 6G NETWORKS

- Smart surfaces
- Network resource optimization
- Vertical solution enablers
- Virtual radio networks optimized with AI

IoT & AI PLATFORMS

- Distributed AI
- Cyber threat intelligence analysis
- Edge-cloud programming models for IoT
- High-performance AI platform for scalable neural networks

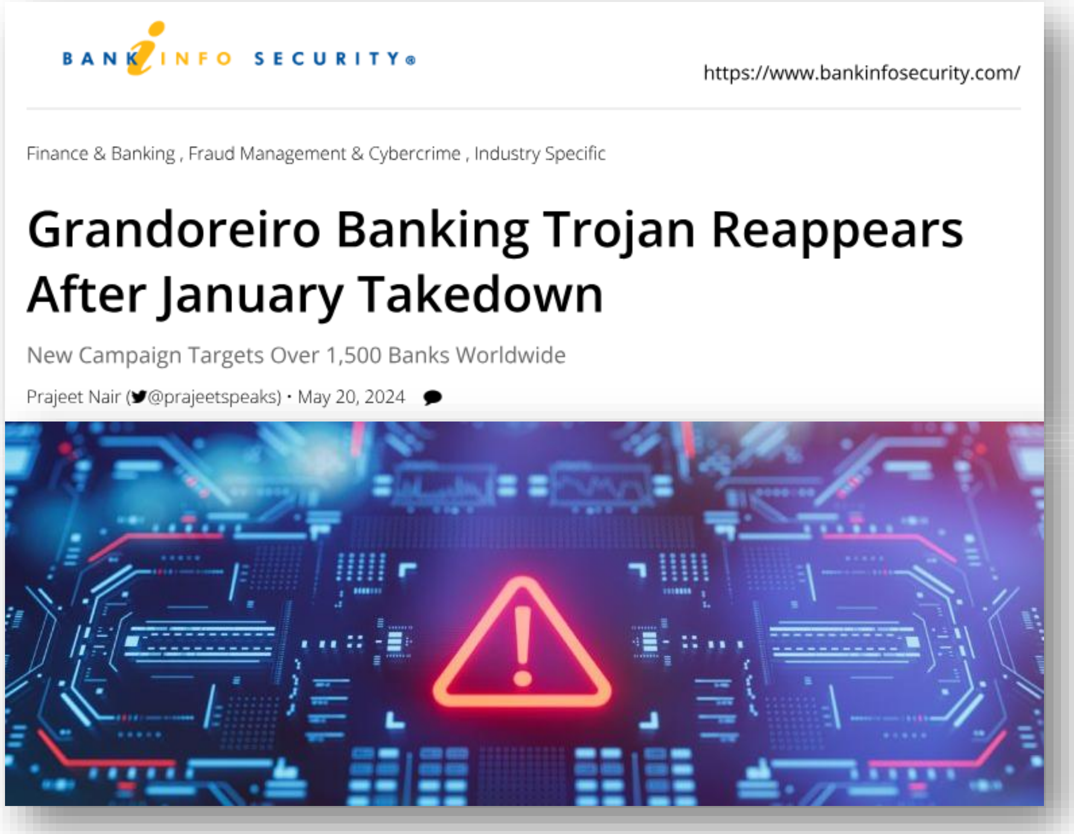
STANDARDS

- ETSI
- FIWARE
- IEEE
- IETF
- 3GPP

What is Cyber Threat Intelligence (CTI)?

- **Cyber Threat Intelligence** (**CTI**) refers to the information that organizations use to understand the cyber threats they are currently facing or might face in the future. It's the organized effort to gather, analyze, and disseminate information about these threats, offering a deeper insight into potential attacks, the tactics, techniques, and procedures (TTPs) of adversaries, their motivations, and the vulnerabilities they may exploit.

An example of CTI



BANK INFO SECURITY

https://www.bankinfosecurity.com/

Finance & Banking , Fraud Management & Cybercrime , Industry Specific

Grandoreiro Banking Trojan Reappears After January Takedown

New Campaign Targets Over 1,500 Banks Worldwide

Prajeet Nair (🐦@prajeetspeaks) • May 20, 2024

Researchers identified new phishing campaigns spreading the Grandoreiro banking Trojan

The latest malware variant also specifically targets over 1500 global banks [...] in over 60 countries.

Researchers observed that the campaigns are now expanding beyond Latin America, targeting countries such as Spain, Japan, the Netherlands and Italy.

Knowing who carries the attack (threat actor) enables to learn likely used attack tools and procedures. This information allows security experts to prepare defense and control risks. This is what **CTI** provides.

Problem: collecting and retrieving CTI is difficult

CYBER THREAT ALLIANCE

Shared information about cyber threats
(NEC is a member!)

11`000`000
new reports per month

Per company!!

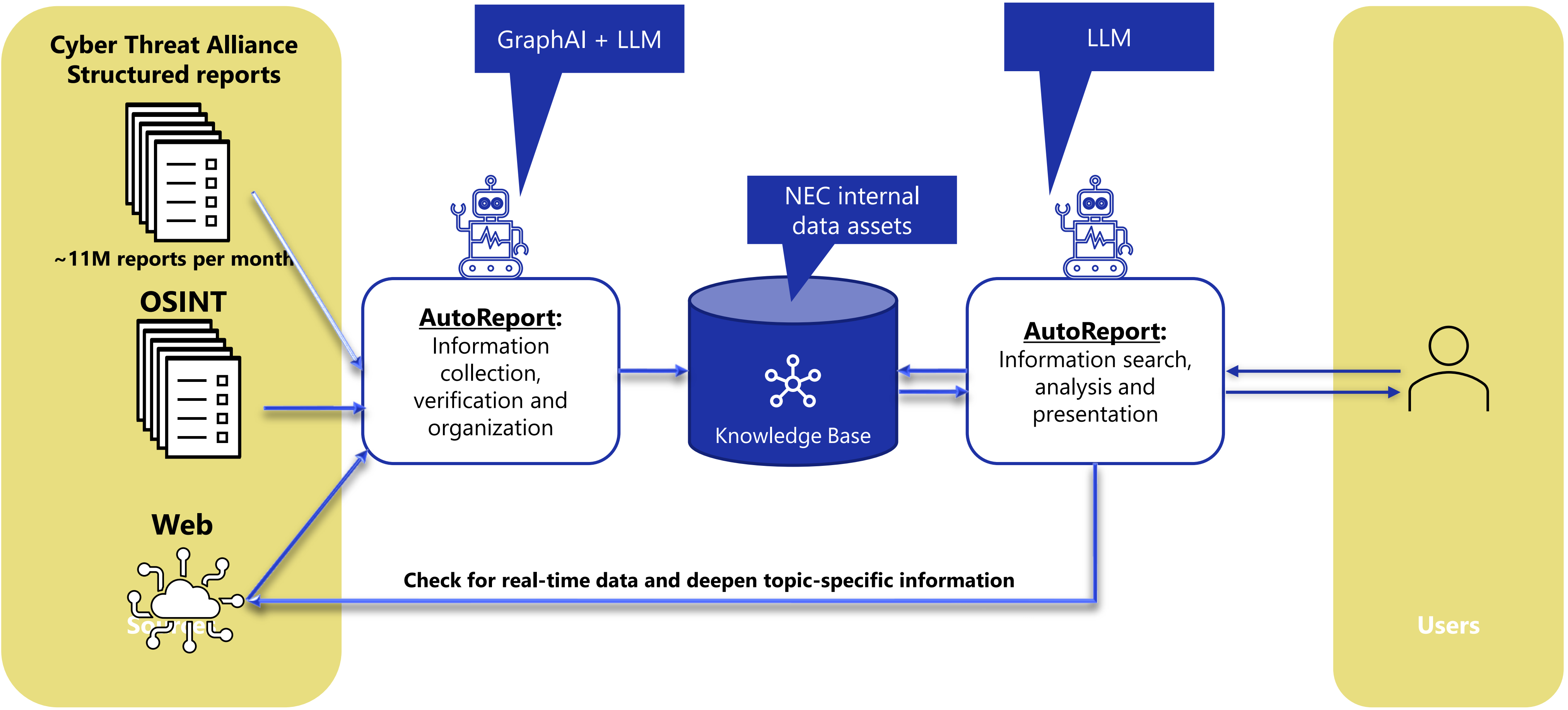
20`000 full-time
Security professionals would be needed to analyze all the reports

Check reports and relate them to the own company (several hours per report)

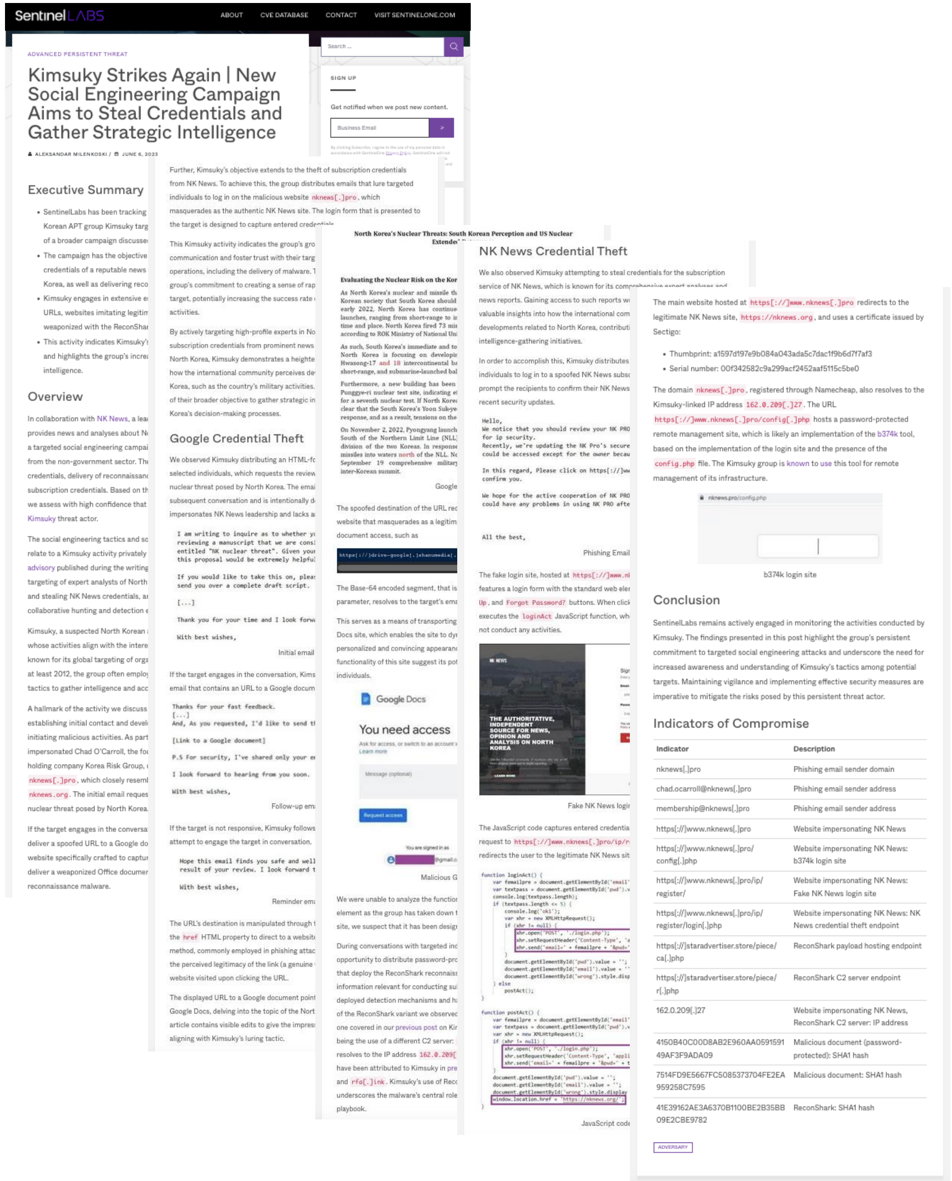
Even when subscribed to industrial alliances to simplify collection and analysis of CTI, there is still too much information that needs analysis by human experts

AutoReport

- An automated Cybersecurity agent that can browse the web, read databases, relate information and generate natural text reports for humans and structured reports for computer systems.
- Started development in 2022; in production since Jan. 2023


Cyber Threat Alliance Structured reports
~11M reports per month
OSINT
Web
Sources
GraphAI + LLM
AutoReport: Information collection, verification and organization
NEC internal data assets
Knowledge Base
LLM
AutoReport: Information search, analysis and presentation
Check for real-time data and deepen topic-specific information
Users

Example of automated analysis output



SentinelLABS — ABOUT — CVE DATABASE — CONTACT — VISIT SENTINELONE.COM

ADVANCED PERSISTENT THREAT

Kimsuky Strikes Again | New Social Engineering Campaign Aims to Steal Credentials and Gather Strategic Intelligence

ALEKSANDAR MILENKOSKI / JUNE 6, 2023

Executive Summary

Further, Kimsuky's objective extends to the theft of subscription credentials from NK News. To achieve this, the group distributes emails that lure targeted individuals to log in on the malicious website nknews[.]pro, which masquerades as the authentic NK News site. The login form that is presented to the target is designed to capture entered credentials.

Overview

Google Credential Theft

NK News Credential Theft

The main website hosted at https[://]www.nknews[.]pro redirects to the legitimate NK News site, https://nknews.org, and uses a certificate issued by Sectigo:

- Thumbprint: a1597d197e9b084a043ada5c7dac1f9b6d7f7af3
- Serial number: 00f342582c9a299acf2452aaf5115c5be0

The domain nknews[.]pro, registered through Namecheap, also resolves to the Kimsuky-linked IP address 162.0.209[.]27. The URL https[://]www.nknews[.]pro/config[.]php hosts a password-protected remote management site, which is likely an implementation of the b374k tool, based on the implementation of the login site and the presence of the config.php file. The Kimsuky group is known to use this tool for remote management of its infrastructure.

b374k login site

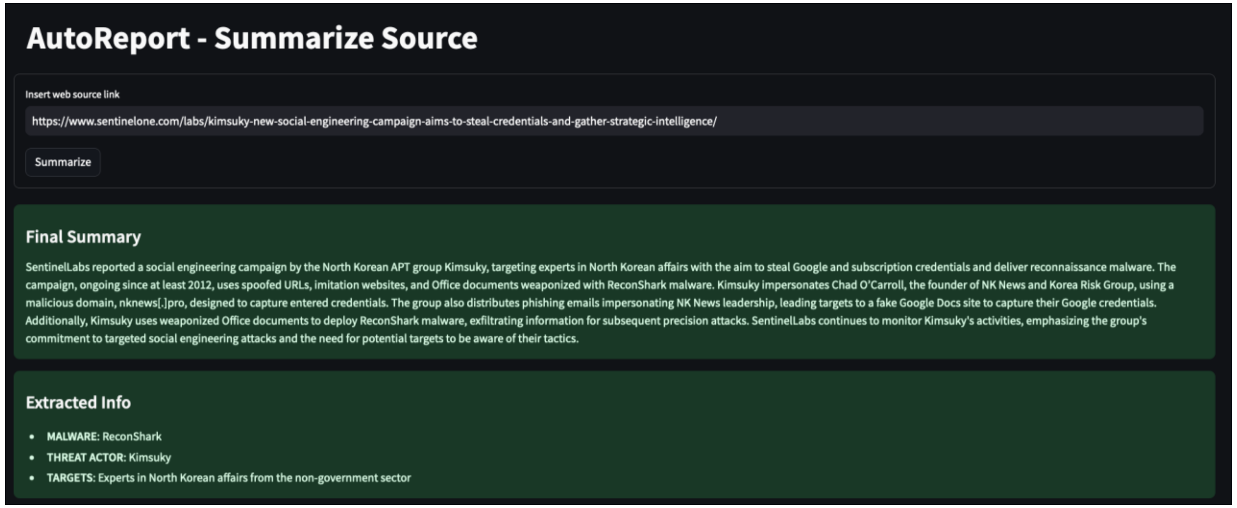
Conclusion

SentinelLabs remains actively engaged in monitoring the activities conducted by Kimsuky. The findings presented in this post highlight the group's persistent commitment to targeted social engineering attacks and underscore the need for increased awareness and understanding of Kimsuky's tactics among potential targets. Maintaining vigilance and implementing effective security measures are imperative to mitigate the risks posed by this persistent threat actor.

Indicators of Compromise

Indicator	Description
nknews[.]pro	Phishing email sender domain
chad.ocarroll@nknews[.]pro	Phishing email sender address
membership@nknews[.]pro	Phishing email sender address
https[://]www.nknews[.]pro	Website impersonating NK News
https[://]www.nknews[.]pro/config[.]php	Website impersonating NK News: b374k login site
https[://]www.nknews[.]pro/ip/register/	Website impersonating NK News: Fake NK News login site
https[://]www.nknews[.]pro/ip/register/login[.]php	Website impersonating NK News: NK News credential theft endpoint
https[://]staradvertiser.store/piece/ca[.]php	ReconShark payload hosting endpoint
https[://]staradvertiser.store/piece/r[.]php	ReconShark C2 server endpoint
162.0.209[.]27	Website impersonating NK News, ReconShark C2 server: IP address
4150B40C00D8AB2E960AA059159149AF3F9ADA09	Malicious document (password-protected): SHA1 hash
7514FD9E5667FC5085373704FE2EA959258C7595	Malicious document: SHA1 hash
41E39162AE3A6370B1100BE2B35BB09E2CBE9782	ReconShark: SHA1 hash

ADVERSARY



AutoReport - Summarize Source

Insert web source link

https://www.sentinelone.com/labs/kimsuky-new-social-engineering-campaign-aims-to-steal-credentials-and-gather-strategic-intelligence/

Summarize

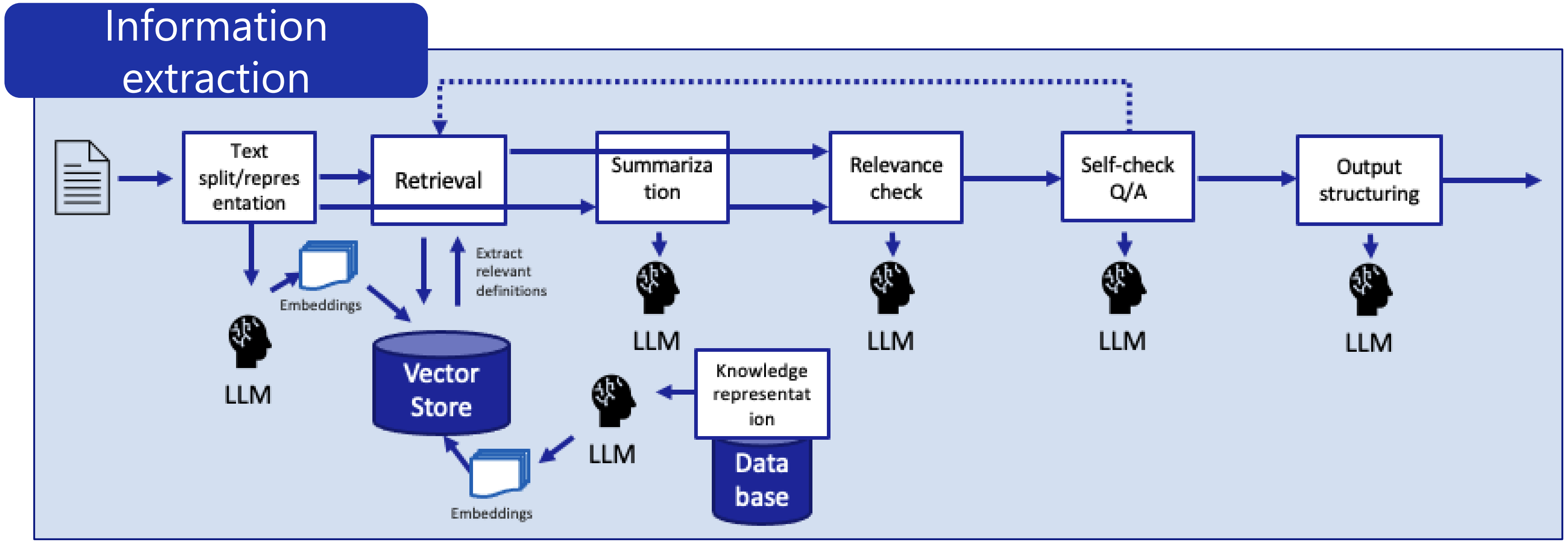
Final Summary

SentinelLabs reported a social engineering campaign by the North Korean APT group Kimsuky, targeting experts in North Korean affairs with the aim to steal Google and subscription credentials and deliver reconnaissance malware. The campaign, ongoing since at least 2012, uses spoofed URLs, imitation websites, and Office documents weaponized with ReconShark malware. Kimsuky impersonates Chad O'Carroll, the founder of NK News and Korea Risk Group, using a malicious domain, nknews[.]pro, designed to capture entered credentials. The group also distributes phishing emails impersonating NK News leadership, leading targets to a fake Google Docs site to capture their Google credentials. Additionally, Kimsuky uses weaponized Office documents to deploy ReconShark malware, exfiltrating information for subsequent precision attacks. SentinelLabs continues to monitor Kimsuky's activities, emphasizing the group's commitment to targeted social engineering attacks and the need for potential targets to be aware of their tactics.

Extracted Info

- **MALWARE:** ReconShark
- **THREAT ACTOR:** Kimsuky
- **TARGETS:** Experts in North Korean affairs from the non-government sector

AutoReport LLM pipeline


Information extraction

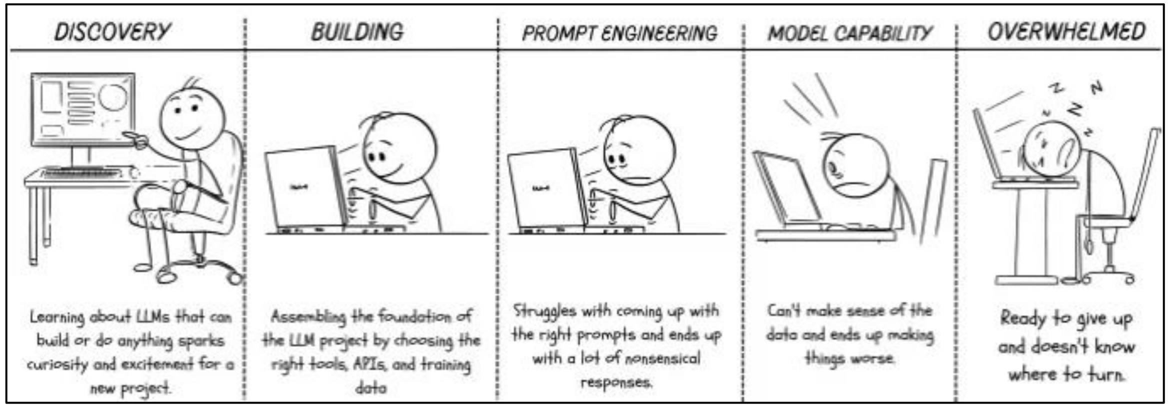
Split complex task into smaller tasks (LLM powered)

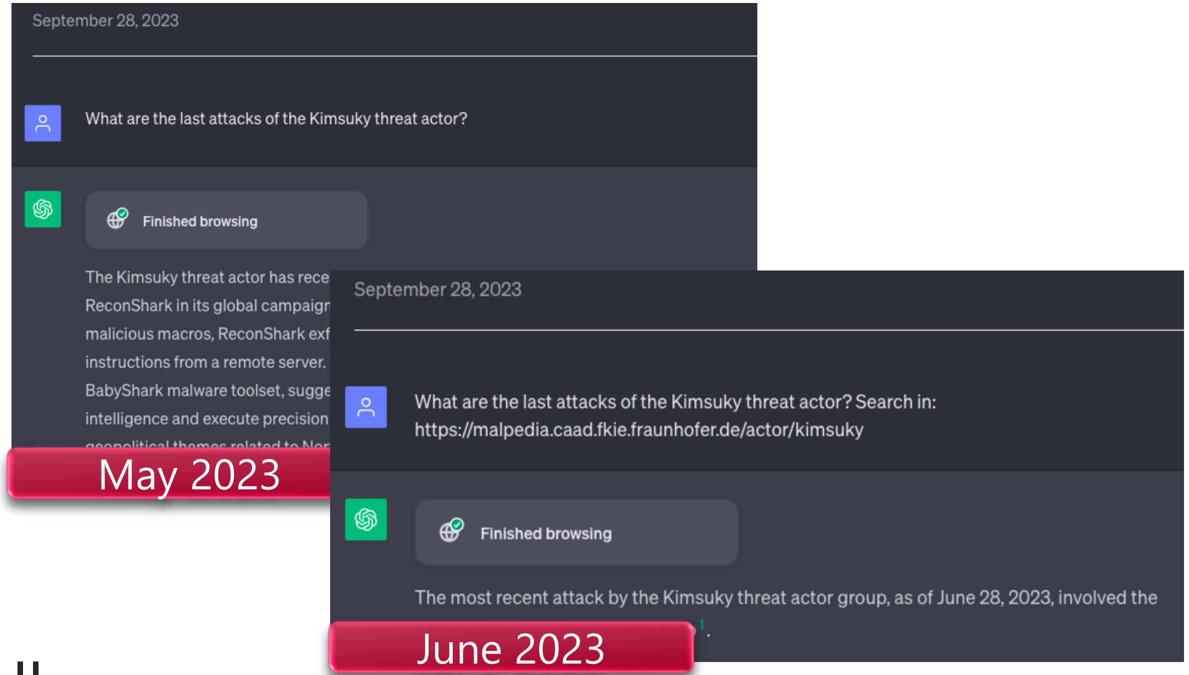
- Smaller and easier tasks -> constant behavior
- Integration with external tools and information sources
  - Access to new information and Verification

Pipeline ordering

- LLM chain contain the hardcoded instruction on ”how to do” a task

Why can’t we simply use LLMs?


DISCOVERY
BUILDING
PROMPT ENGINEERING
MODEL CAPABILITY
OVERWHELMED
Learning about LLMs that can build or do anything sparks curiosity and excitement for a new project.
Assembling the foundation of the LLM project by choosing the right tools, APIs, and training data
Struggles with coming up with the right prompts and ends up with a lot of nonsensical responses.
Can't make sense of the data and ends up making things worse.
Ready to give up and doesn't know where to turn.



September 28, 2023

What are the last attacks of the Kimsuky threat actor?

Finished browsing

The Kimsuky threat actor has rece
ReconShark in its global campaign
malicious macros, ReconShark exf
instructions from a remote server.
BabyShark malware toolset, sugge
intelligence and execute precision
geopolitical themes related to Nor

May 2023

September 28, 2023

What are the last attacks of the Kimsuky threat actor? Search in:
https://malpedia.caad.fkie.fraunhofer.de/actor/kimsuky

Finished browsing

The most recent attack by the Kimsuky threat actor group, as of June 28, 2023, involved the

June 2023

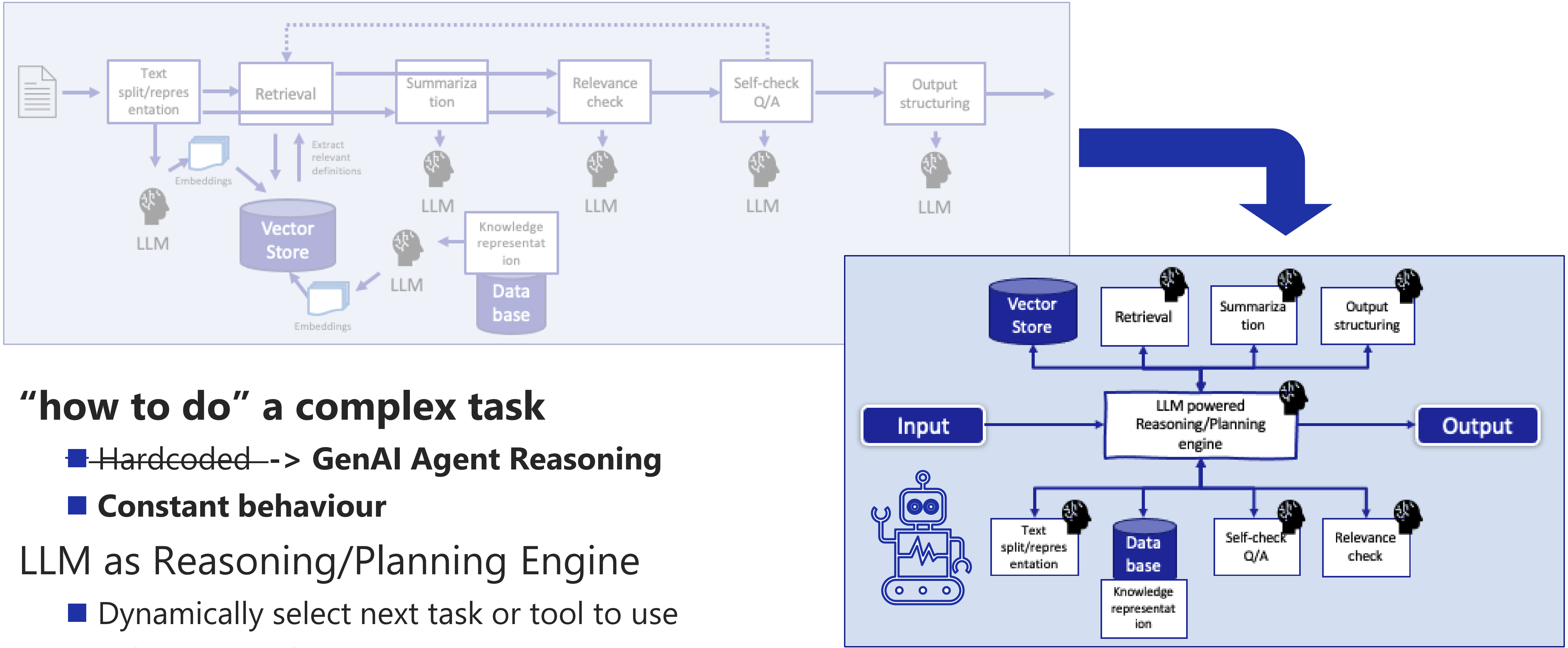
Challenge

- Easy to get a lucky prompt
  - **Extremely complex to generalize it**
- Easy to build a first prototype
  - **Extremely complex to have constant behaviour**

Challenge

- Easy to automate “What to do”?
  - E.g. Search for news related to Kimsuky
- **Difficult to automate “How to do it”**
  - E.g. Use right information sources
- **Include Human Experience into Automation**

Generative AI Agent – From Harcoded instruction to **Reasoning**



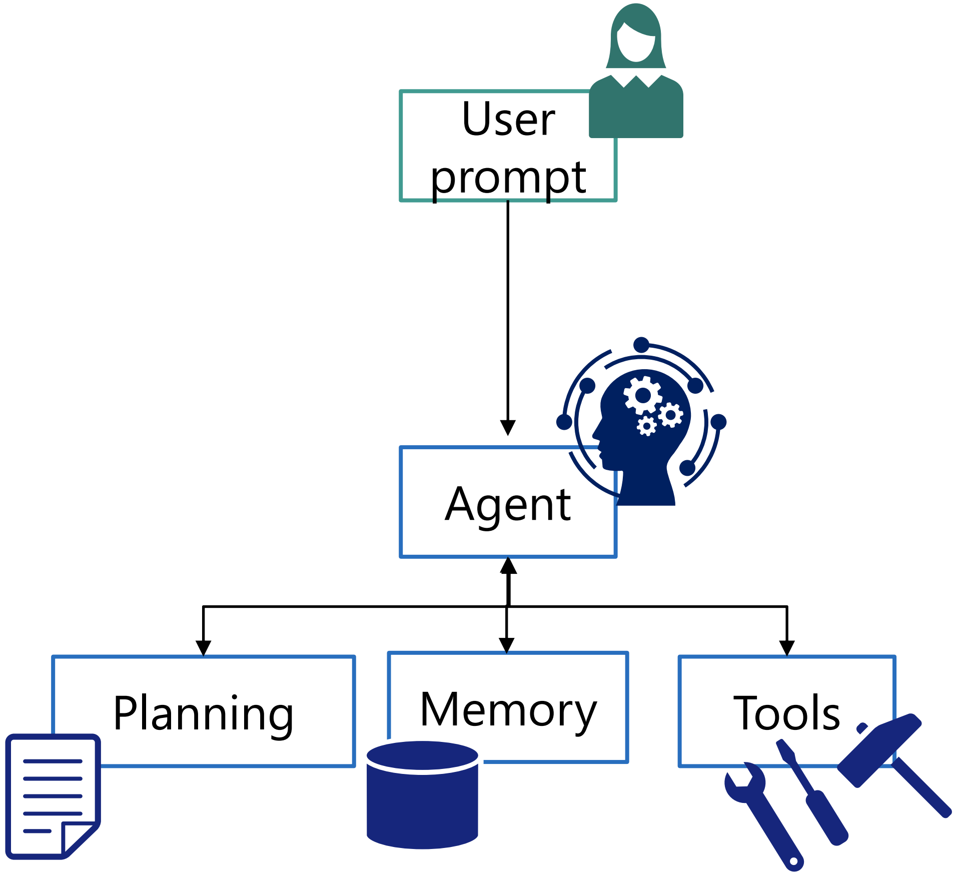
“how to do” a complex task

- ~~Hardcoded~~ -> **GenAI Agent Reasoning**
- **Constant behaviour**

LLM as Reasoning/Planning Engine

- Dynamically select next task or tool to use
- Behave according to given Best Practice

Generative AI Agents



Tools:

- Corporate Database
- Command line interpreter
- Python interpreter
- Web search
- Control a physical device
- …

- Another Agent
- …

- Access to Data Spaces

Planning

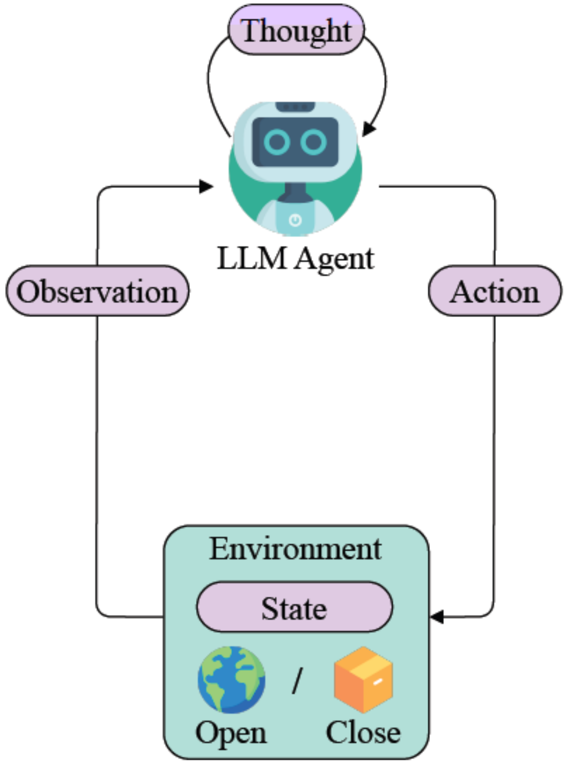
Task-decomposition

- Tree of Thoughts (Yao et al. 2023)
- Chain of thought (Wei et al. 2022)

Self-Reflection

- Reflexion (Shinn & Labash 2023)
- Chain of Hindsight (Liu et al. 2023)
- ReAct (Yao et al. 2023)

Let´s see an example of an agent doing Penetration Testing



Discover, infiltrate, and escalate privileges on the target machine.
Begin by locating the machine on the network and gaining entry as the 'student' user.
Delve into the system to elevate privileges and obtain root access.
Look within the root user's home directory for a file named 'flag'.

Different Agents

Cybersecurity

- ◆ CTI STIX extraction
- ◆ Penetration testing
- ◆ Deduplication agent
- ◆ Telegram/Darkweb crawler
- ◆ Cryptocurrency analysis agent

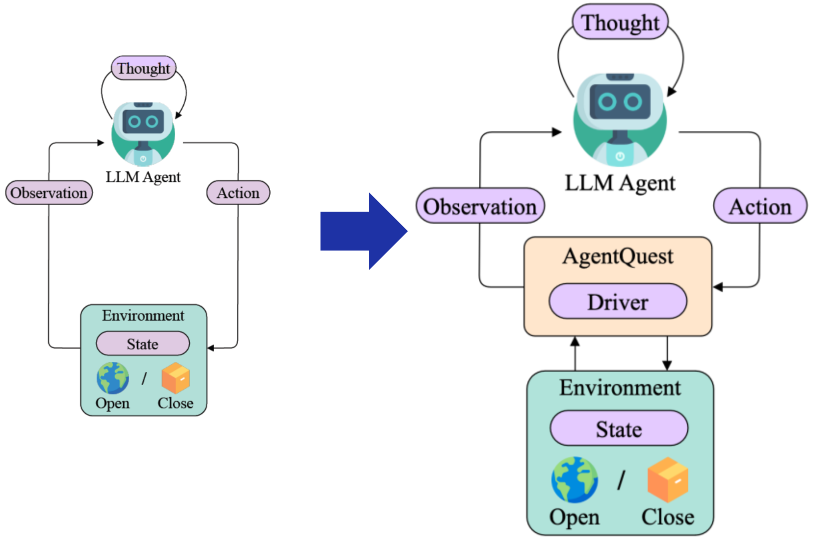
Others

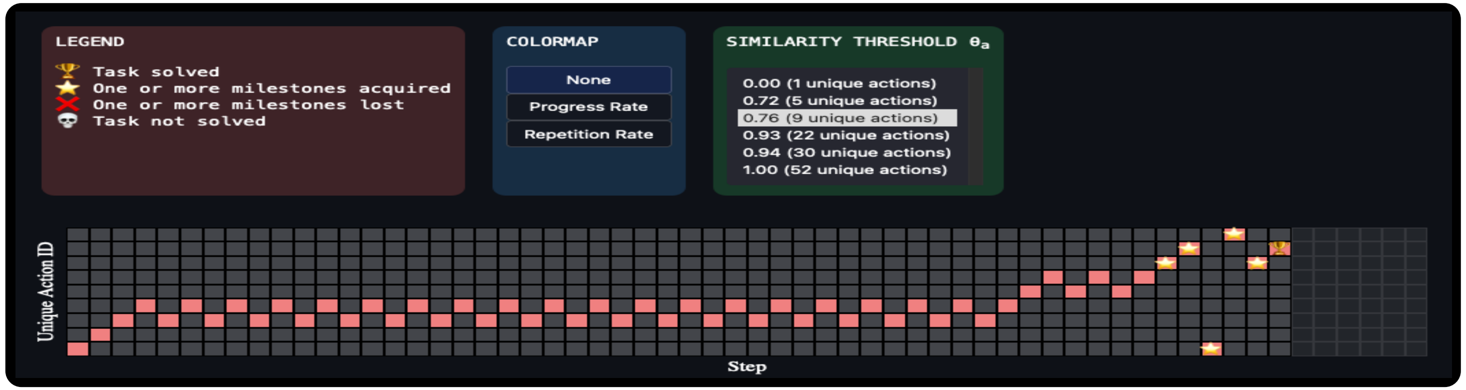
- ◆ Automated Data Analyst
- ◆ Network Configuration Agent

Research projects

Benchmarking AI agents

- Benchmarking different Agent implementations is hard
- We have developed a modular framework to support diverse benchmarks and architectures.
- Implements 2 existing metrics:
  - Success Rate (SR) and Time to Success (Steps)
- Adds two new metrics:
  - Progress Rate (PR) and Repetition Rate (RR)




LEGEND

🏆 Task solved
⭐ One or more milestones acquired
❌ One or more milestones lost
💀 Task not solved

COLORMAP

None
Progress Rate
Repetition Rate

SIMILARITY THRESHOLD θ_a

0.00 (1 unique actions)
0.72 (5 unique actions)
0.76 (9 unique actions)
0.93 (22 unique actions)
0.94 (30 unique actions)
1.00 (52 unique actions)

VIDEO

CODE

*AgentQuest: A Modular Benchmark Framework to Measure Progress and Improve LLM Agents
To appear in NAACL 2024

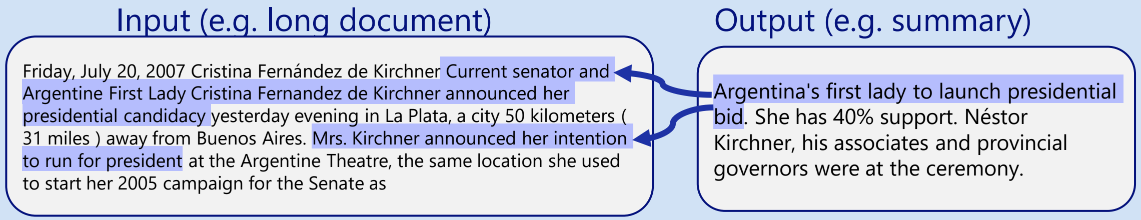
LLM Explainer

- LLM might hallucinate
- How can we make their use safer?
- Allow humans to verify if the LLM output is correct by attributing information to source / trusted references

LLM Explainer

- Run after LLM generated a text
- Link the generated text to the **input source** (e.g. the prompt, context, original document, …)

Trace back to source → verify correctness



Input (e.g. long document)

Friday, July 20, 2007 Cristina Fernández de Kirchner Current senator and Argentine First Lady Cristina Fernandez de Kirchner announced her presidential candidacy yesterday evening in La Plata, a city 50 kilometers (31 miles) away from Buenos Aires. Mrs. Kirchner announced her intention to run for president at the Argentine Theatre, the same location she used to start her 2005 campaign for the Senate as

Output (e.g. summary)

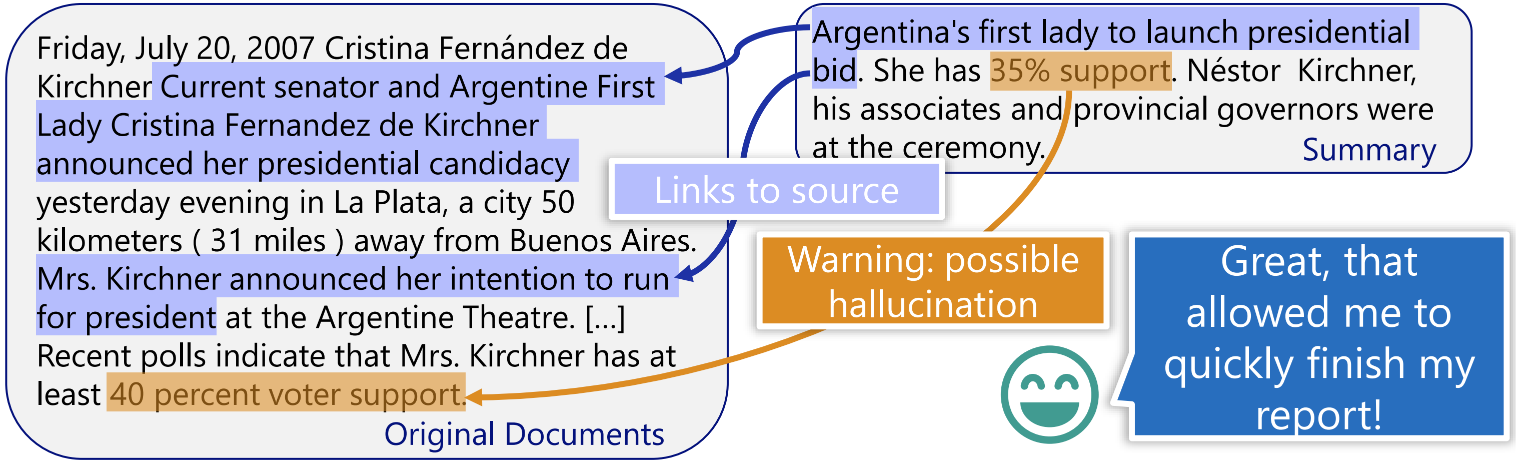
Argentina's first lady to launch presidential bid. She has 40% support. Néstor Kirchner, his associates and provincial governors were at the ceremony.

LLM Explainer

Competition

Argentina's first lady to launch presidential bid. She has 35% support. Néstor Kirchner, his associates and provincial governors were at the ceremony. Summary

This will take a long time: I have to read all these source documents to know if the summary is correct!



Friday, July 20, 2007 Cristina Fernández de Kirchner Current senator and Argentine First Lady Cristina Fernandez de Kirchner announced her presidential candidacy yesterday evening in La Plata, a city 50 kilometers (31 miles) away from Buenos Aires. Mrs. Kirchner announced her intention to run for president at the Argentine Theatre. [...] Recent polls indicate that Mrs. Kirchner has at least 40 percent voter support.

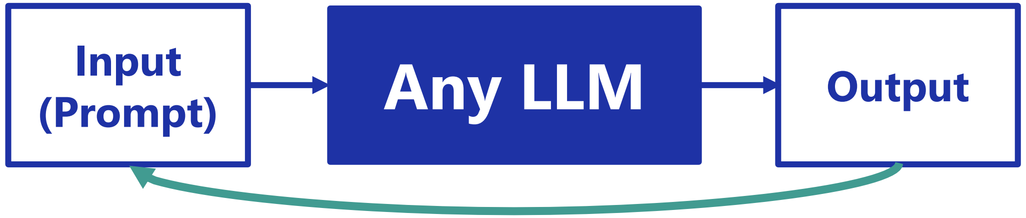
Original Documents

Argentina's first lady to launch presidential bid. She has 35% support. Néstor Kirchner, his associates and provincial governors were at the ceremony.

Summary

Great, that allowed me to quickly finish my report!

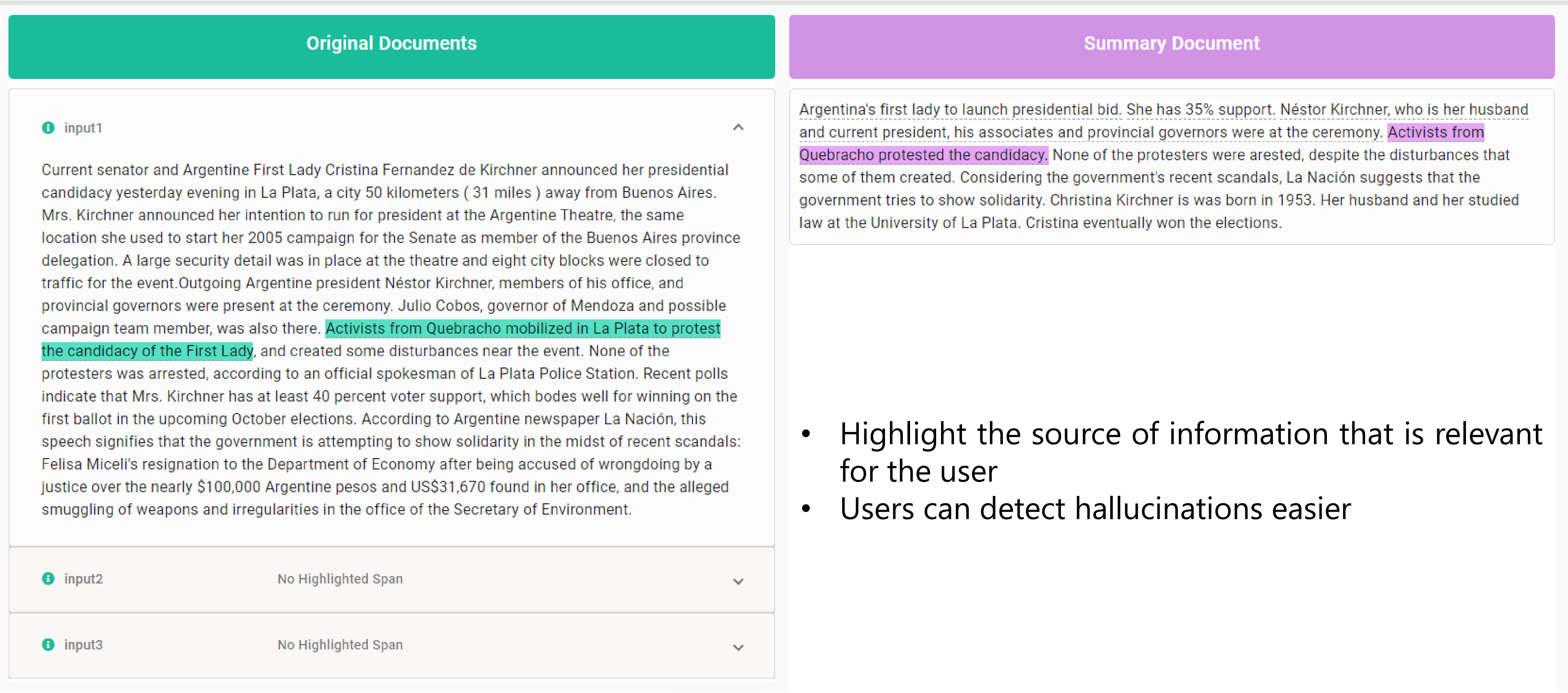
- The LLM Explainer can explain LLM summaries and answers, it does this by:
  - Creating links to the source document to allow fast verification
  - Warning of possible hallucinations for quick edits
- No known competitors for the unique feature



LLM Explainer

Explain the LLM output for quick verification

Shows where the information comes from



Original Documents

input1

Current senator and Argentine First Lady Cristina Fernandez de Kirchner announced her presidential candidacy yesterday evening in La Plata, a city 50 kilometers (31 miles) away from Buenos Aires. Mrs. Kirchner announced her intention to run for president at the Argentine Theatre, the same location she used to start her 2005 campaign for the Senate as member of the Buenos Aires province delegation. A large security detail was in place at the theatre and eight city blocks were closed to traffic for the event.Outgoing Argentine president Néstor Kirchner, members of his office, and provincial governors were present at the ceremony. Julio Cobos, governor of Mendoza and possible campaign team member, was also there. Activists from Quebracho mobilized in La Plata to protest the candidacy of the First Lady, and created some disturbances near the event. None of the protesters was arrested, according to an official spokesman of La Plata Police Station. Recent polls indicate that Mrs. Kirchner has at least 40 percent voter support, which bodes well for winning on the first ballot in the upcoming October elections. According to Argentine newspaper La Nación, this speech signifies that the government is attempting to show solidarity in the midst of recent scandals: Felisa Miceli's resignation to the Department of Economy after being accused of wrongdoing by a justice over the nearly $100,000 Argentine pesos and US$31,670 found in her office, and the alleged smuggling of weapons and irregularities in the office of the Secretary of Environment.

input2 No Highlighted Span

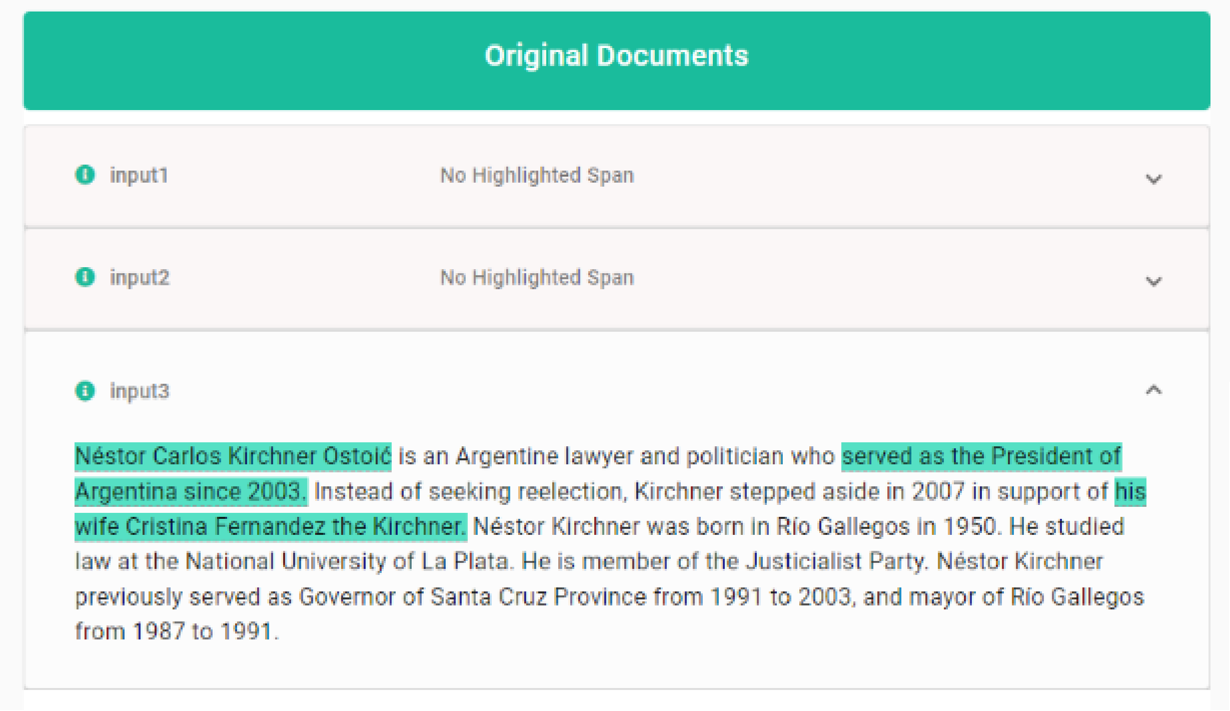
input3 No Highlighted Span

Summary Document

Argentina's first lady to launch presidential bid. She has 35% support. Néstor Kirchner, who is her husband and current president, his associates and provincial governors were at the ceremony. Activists from Quebracho protested the candidacy. None of the protesters were arested, despite the disturbances that some of them created. Considering the government's recent scandals, La Nación suggests that the government tries to show solidarity. Christina Kirchner is was born in 1953. Her husband and her studied law at the University of La Plata. Cristina eventually won the elections.

- Highlight the source of information that is relevant for the user
- Users can detect hallucinations easier

Sometimes information comes from multiple places



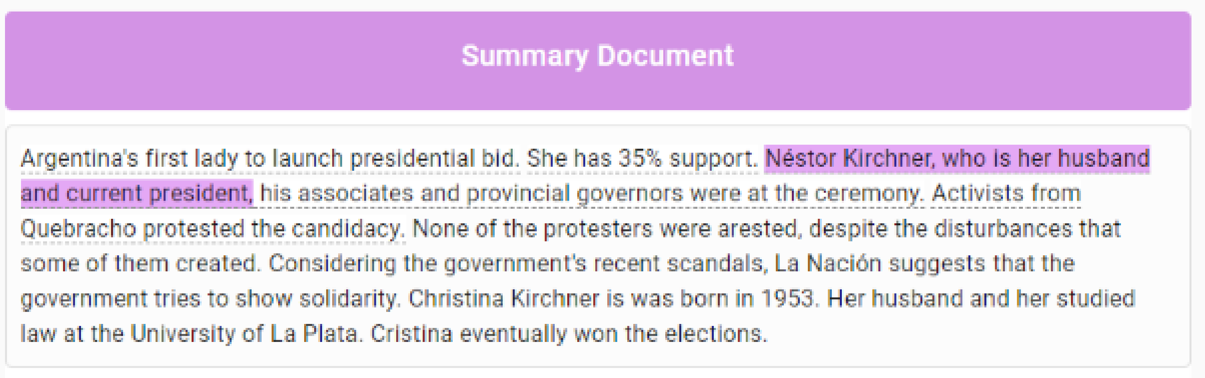
Original Documents

input1 No Highlighted Span

input2 No Highlighted Span

input3

Néstor Carlos Kirchner Ostoić is an Argentine lawyer and politician who served as the President of Argentina since 2003. Instead of seeking reelection, Kirchner stepped aside in 2007 in support of his wife Cristina Fernandez the Kirchner. Néstor Kirchner was born in Río Gallegos in 1950. He studied law at the National University of La Plata. He is member of the Justicialist Party. Néstor Kirchner previously served as Governor of Santa Cruz Province from 1991 to 2003, and mayor of Río Gallegos from 1987 to 1991.

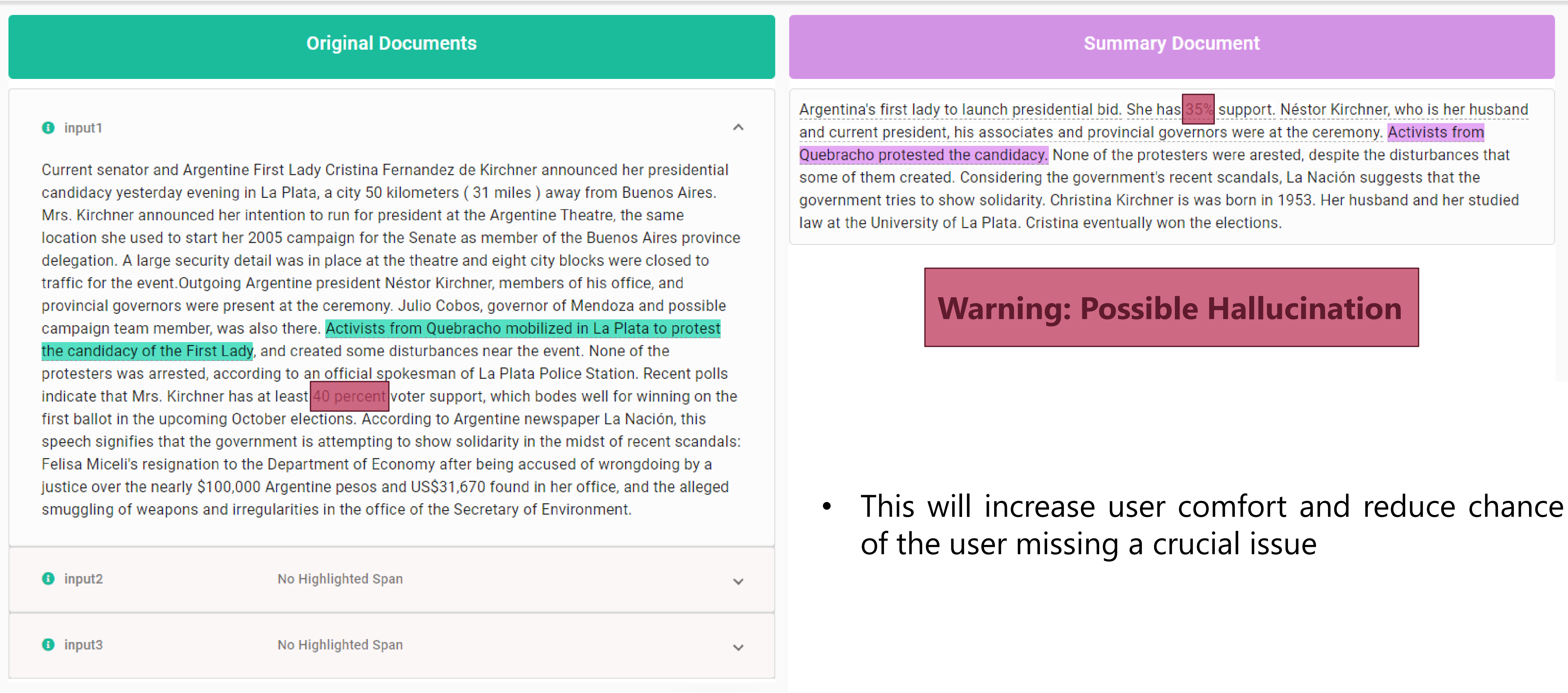


Summary Document

Argentina's first lady to launch presidential bid. She has 35% support. Néstor Kirchner, who is her husband and current president, his associates and provincial governors were at the ceremony. Activists from Quebracho protested the candidacy. None of the protesters were arested, despite the disturbances that some of them created. Considering the government's recent scandals, La Nación suggests that the government tries to show solidarity. Christina Kirchner is was born in 1953. Her husband and her studied law at the University of La Plata. Cristina eventually won the elections.

- Locate *all* information in the original documents that contributed to the content of the generated text
- This can help people to verify the information easier

Warn of hallucination



Original Documents

input1

Current senator and Argentine First Lady Cristina Fernandez de Kirchner announced her presidential candidacy yesterday evening in La Plata, a city 50 kilometers (31 miles) away from Buenos Aires. Mrs. Kirchner announced her intention to run for president at the Argentine Theatre, the same location she used to start her 2005 campaign for the Senate as member of the Buenos Aires province delegation. A large security detail was in place at the theatre and eight city blocks were closed to traffic for the event.Outgoing Argentine president Néstor Kirchner, members of his office, and provincial governors were present at the ceremony. Julio Cobos, governor of Mendoza and possible campaign team member, was also there. Activists from Quebracho mobilized in La Plata to protest the candidacy of the First Lady, and created some disturbances near the event. None of the protesters was arrested, according to an official spokesman of La Plata Police Station. Recent polls indicate that Mrs. Kirchner has at least 40 percent voter support, which bodes well for winning on the first ballot in the upcoming October elections. According to Argentine newspaper La Nación, this speech signifies that the government is attempting to show solidarity in the midst of recent scandals: Felisa Miceli's resignation to the Department of Economy after being accused of wrongdoing by a justice over the nearly $100,000 Argentine pesos and US$31,670 found in her office, and the alleged smuggling of weapons and irregularities in the office of the Secretary of Environment.

input2 No Highlighted Span

input3 No Highlighted Span

Summary Document

Argentina's first lady to launch presidential bid. She has 35% support. Néstor Kirchner, who is her husband and current president, his associates and provincial governors were at the ceremony. Activists from Quebracho protested the candidacy. None of the protesters were arested, despite the disturbances that some of them created. Considering the government's recent scandals, La Nación suggests that the government tries to show solidarity. Christina Kirchner is was born in 1953. Her husband and her studied law at the University of La Plata. Cristina eventually won the elections.

Warning: Possible Hallucination

- This will increase user comfort and reduce chance of the user missing a crucial issue

NEC brings LLMs to EU projects

DESIRE6G

Optimization of Neural Networks
Optimization of LLM training and inference

EMPYREAN

Analysis of CTI using LLMs

NATWORK

LLM based Cyber Threat Assistant

LICORICE

Privacy Preserving LLMs

Wrapping up

- We need AI to Support cybersecurity… and many other areas
- LLMs are very powerful… but it is very difficult to obtain consistent behaviour and “productize” them
  - Moreover, fighting hallucinations
- AI Agents improve the situation… but still it is difficult to benchmark them
- Agent tools need to access data!! We need:
  - Ways to check if the information is legit
  - Trusted sources -> European Data Spaces
- We are working to solve all these problems!

\Orchestrating a brighter world

NEC

Roberto.Gonzalez@neclab.eu