AI Agent in the Workflow Orchestrator

Peter Boers – peter.boers@surf.nl

10 April, 2026

SURF

> whoami

- Peter Boers
- 37 years old
- 10th year at SURF
- Background in Software and Network (SNE @ UvA)
- Started as a Trainee…
- …. moved on as Technical Product Manager A&O
- Tech lead of the Workflow Orchestrator Programme

> cat ~/.zsh_history

- Programming languages: Python, React, Angular
- Databases: Postgres, Mongo, MySQL, Redis
- Streaming: Kafka, Risingwave
- Platform: Kubernetes, ArgoCD, GitlabCI, Github
- Opensource: Workflow Orchestrator Programme
- **Current focus:** Influx, PyTorch, Snellius + ML, **AI Agents**

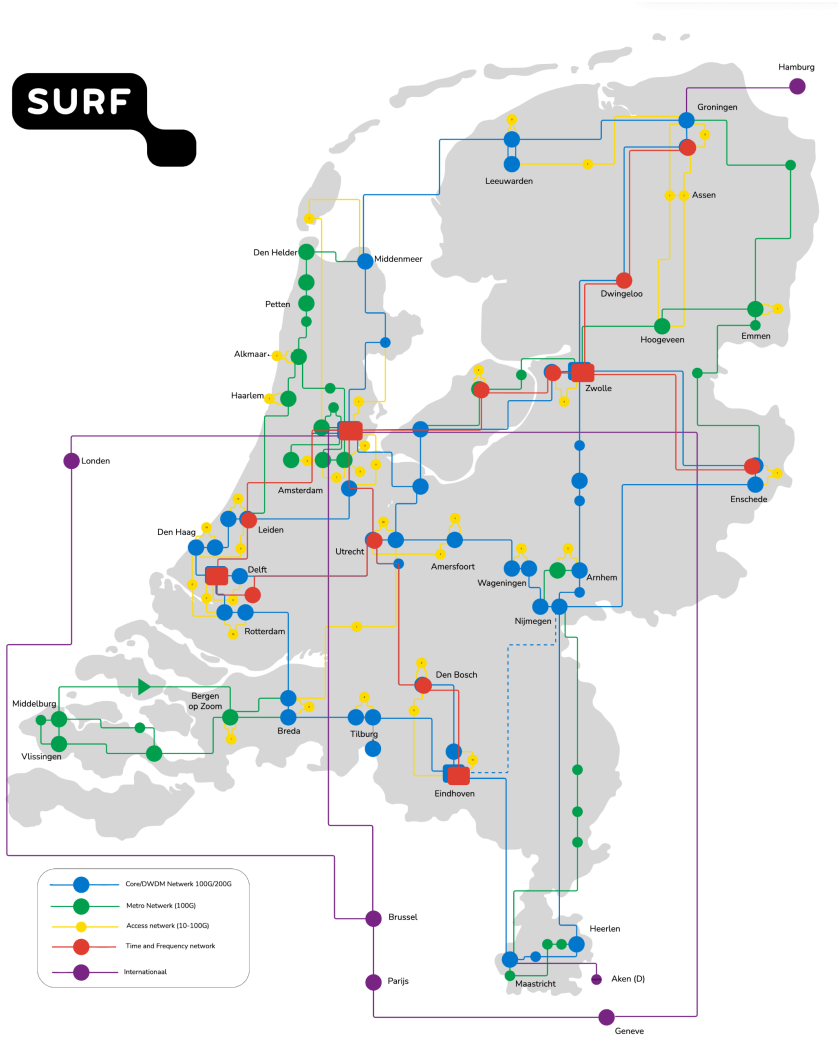
Automation -> Orchestration -> Intelligent Networks

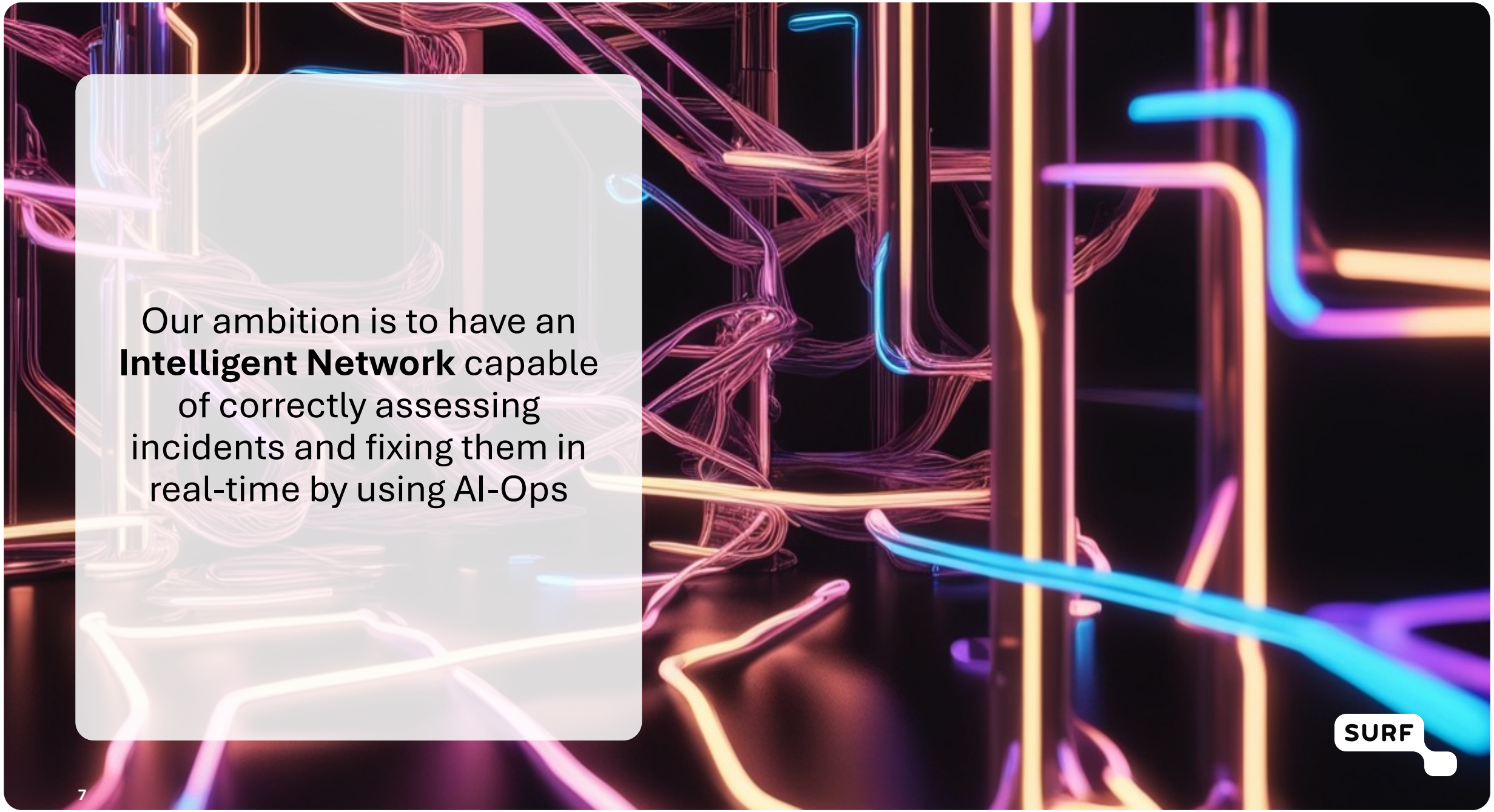
- Since 2018, SURF has been automating the network.
- Fully automated delivery of services with the arrival of SURFnet8
- Results:
- +/- **30 thousand automated** changes
- > **14 million automated tasks**
- **5554** active services (from BGP to Interfaces)
- **15k/month** downloads on PyPi for orchestrator-core
- **Self-Service** in the Network Dashboard
- Open source Software developed by SURF in use by **12 to 13** other coporations

SURFnet8 -> SURFnet∞

SURFnet8 -> SURFnet∞

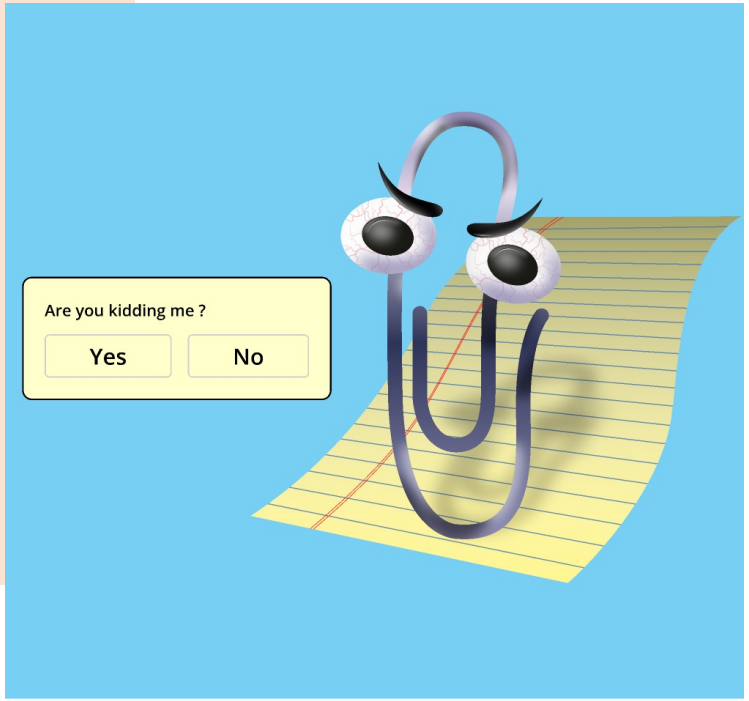
- Opportunity to revitalize the network and the software that controls it
- From the tender(s) it is possible to get a Multi Vendor/Multi-OS architecture
- Diversity and complexity are increasing
  - Campus network offering
  - TFT
- Higher level of integration is required.
- Developments follow in rapid succession and the tools we use must grow with them

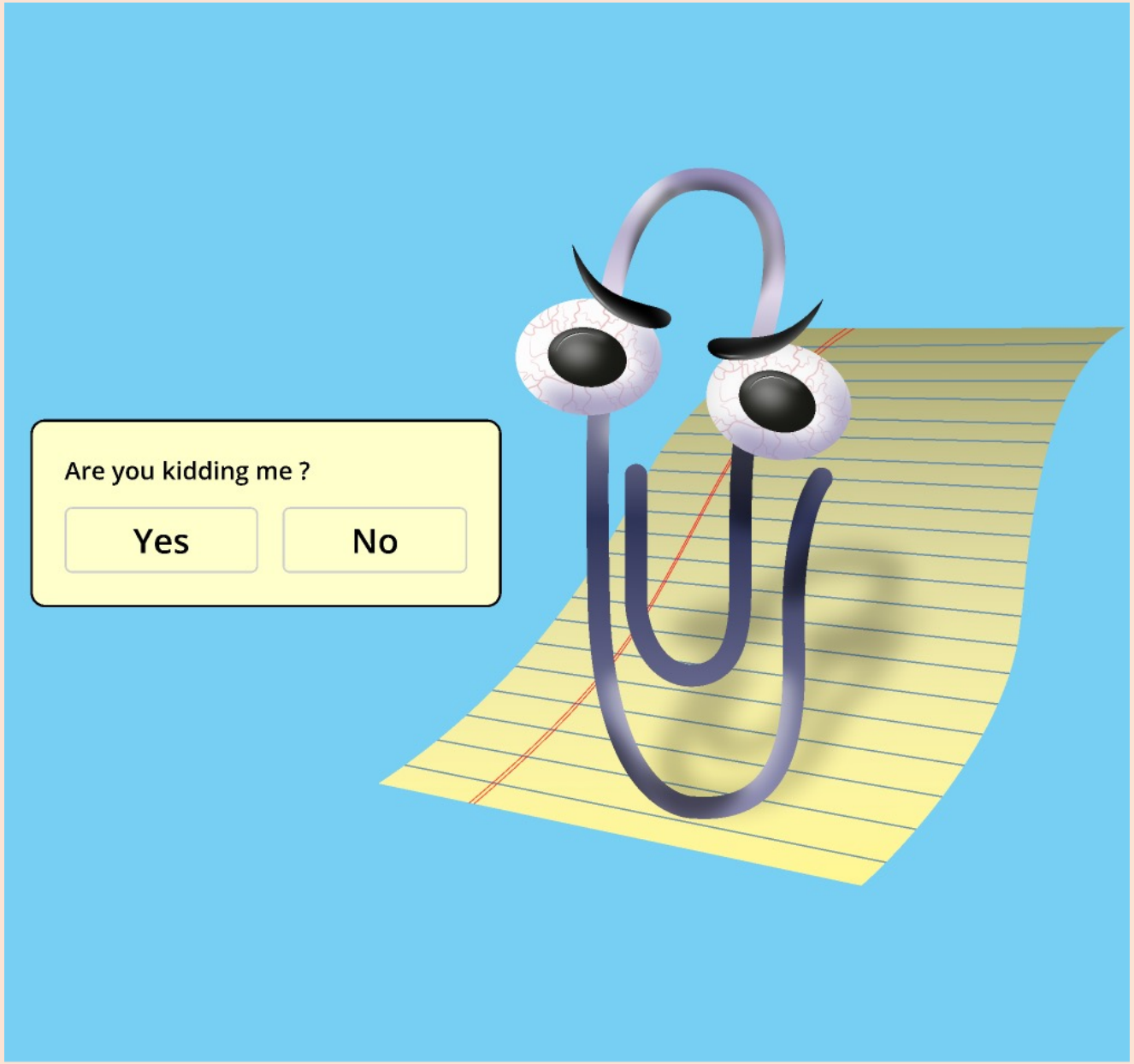



Our ambition is to have an
Intelligent Network capable
of correctly assessing
incidents and fixing them in
real-time by using AI-Ops

Network evolution, not revolution

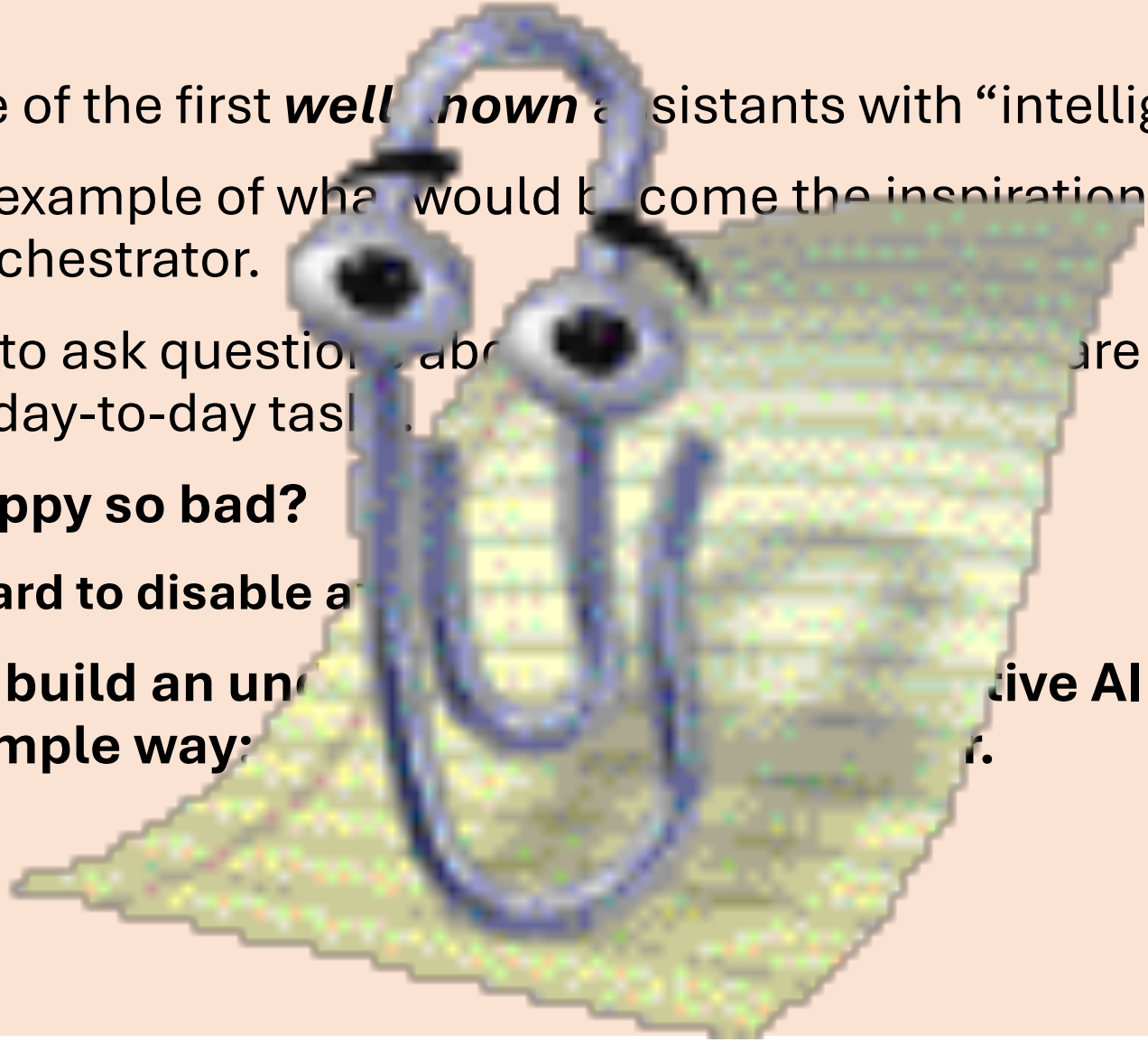
- With the continued development of SURFnet8, there is an opportunity to work with new(er) technologies
- Two strategies are emerging for AIOps
  - Machine Learning
  - LLM integration




Are you kidding me ?
Yes
No

What does clippy have to do with AIOp's and WFO?

- Arguably one of the first ***well known*** assistants with "intelligence"
- A very naïve example of what would become the inspiration for an AI Agent in the Workflow Orchestrator.
- An interface to ask questions about … are working in designed to help with day-to-day tasks.
- **Why was clippy so bad?**
  - **Intrusive, hard to disable a…**
- **Goal was to build an un… tive AI can be used in a relatively simple way: …r.**



SURF

Sidenote: Who remembers November 30th 2022?

- Perhaps just as significa
  - **Release of the first Oper**
- Since 2022 the technolo has fundamentally changed.
- All about **Retrieval Auge**
- **RAG** stands for the proce s the data from external/internal sources and then genera
- Key part of the **Retrieval** documents.

- The RAG architect nt in WFO

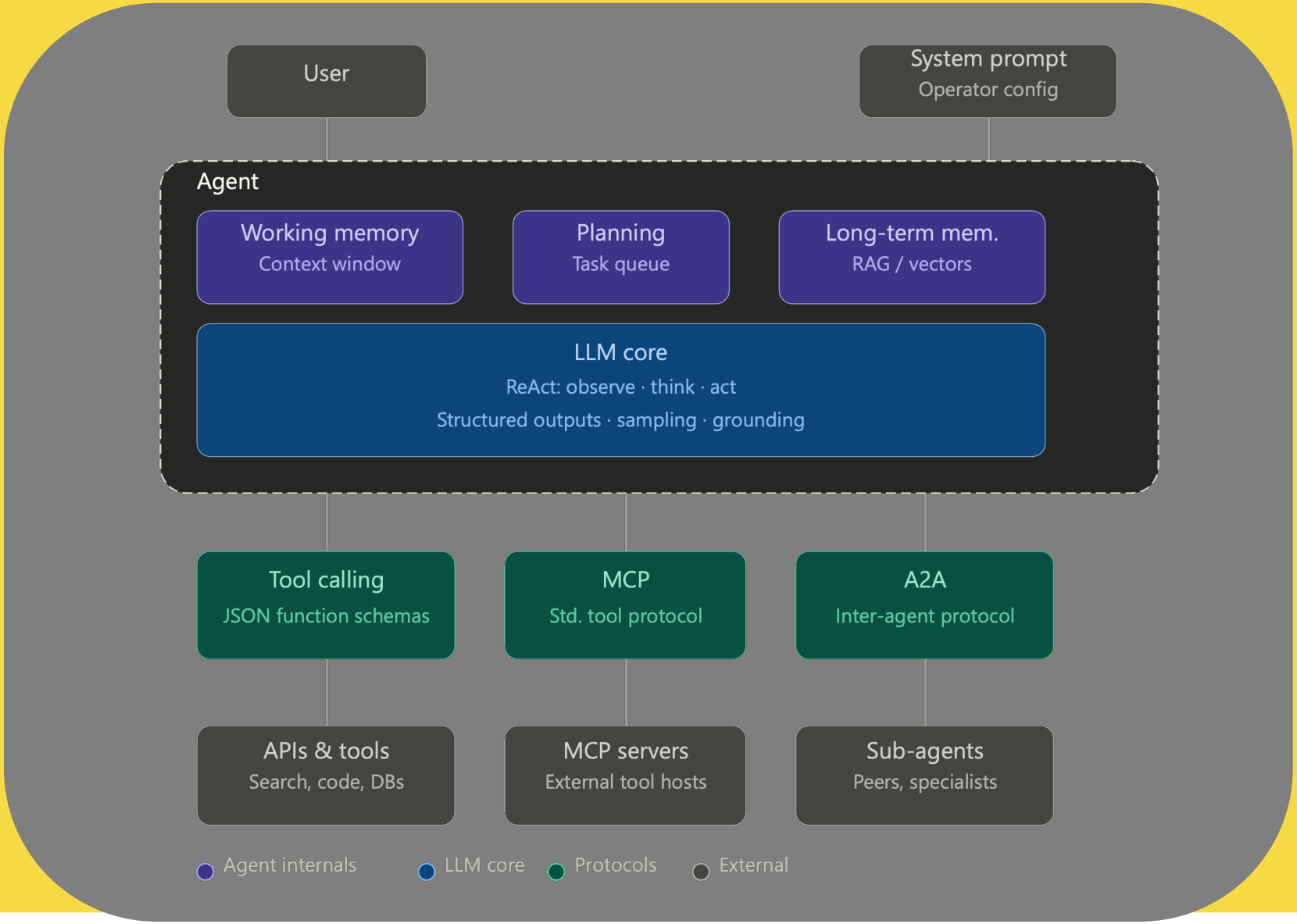
SURF

What is an AI Agent?

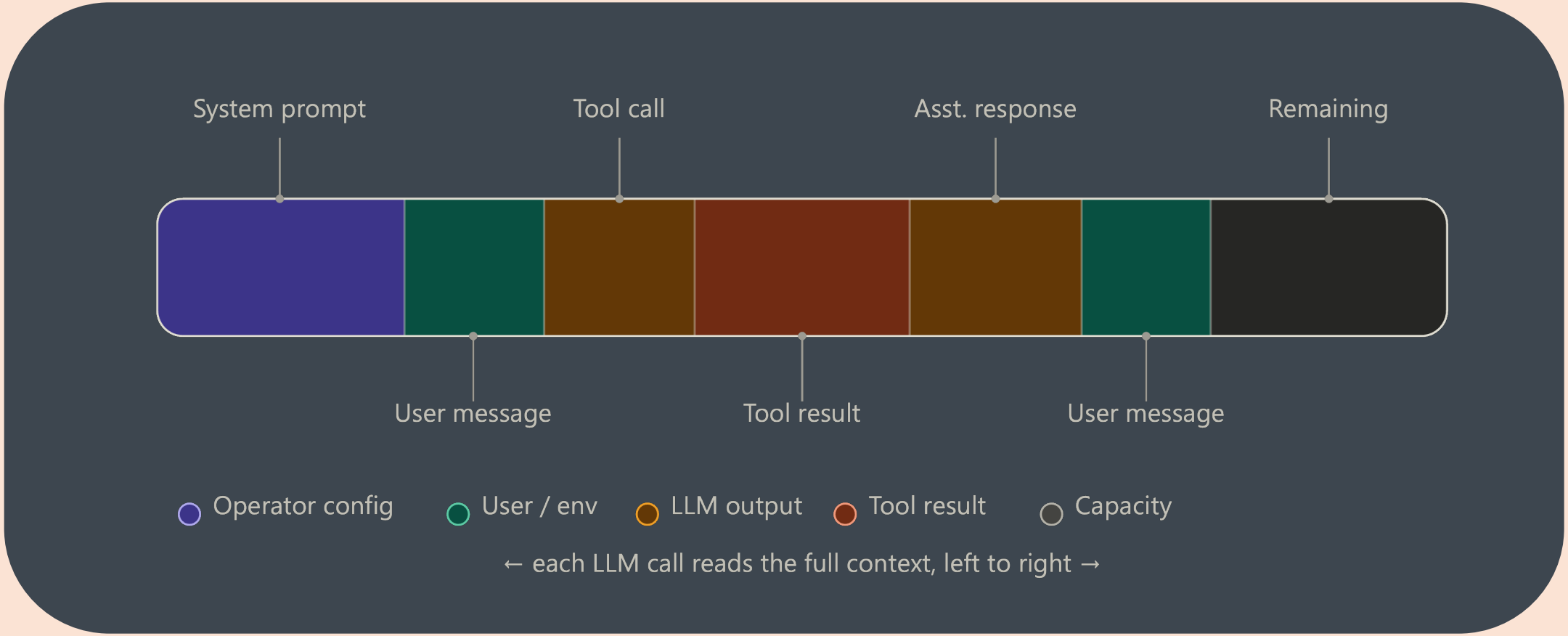
- “An AI agent is a system where an LLM drives a loop — perceiving inputs, deciding on actions, calling tools, and observing results — until a goal is reached.” **– Claude Code (2026)**
- The **loop** mentioned above is usually the **ReAct** (Reasoning + Acting) loop.
  - The LLM first reasons about the prompt
  - Then Acts upon it by calling tools and/or generating a result.
- The **goal** the agent tries to achieve is the **System Prompt** combined with the **User Prompt**
  - **System Prompt:** You are a useful AI agent and an expert in search the orchestrator, these are your tools….
  - **User Prompt:** Look up subscription xxx
- LLM’s are stateless/have no database, they have a bounded memory they can consume during each call/session, **their context window**

SURF

What is an AI agent?



Context window

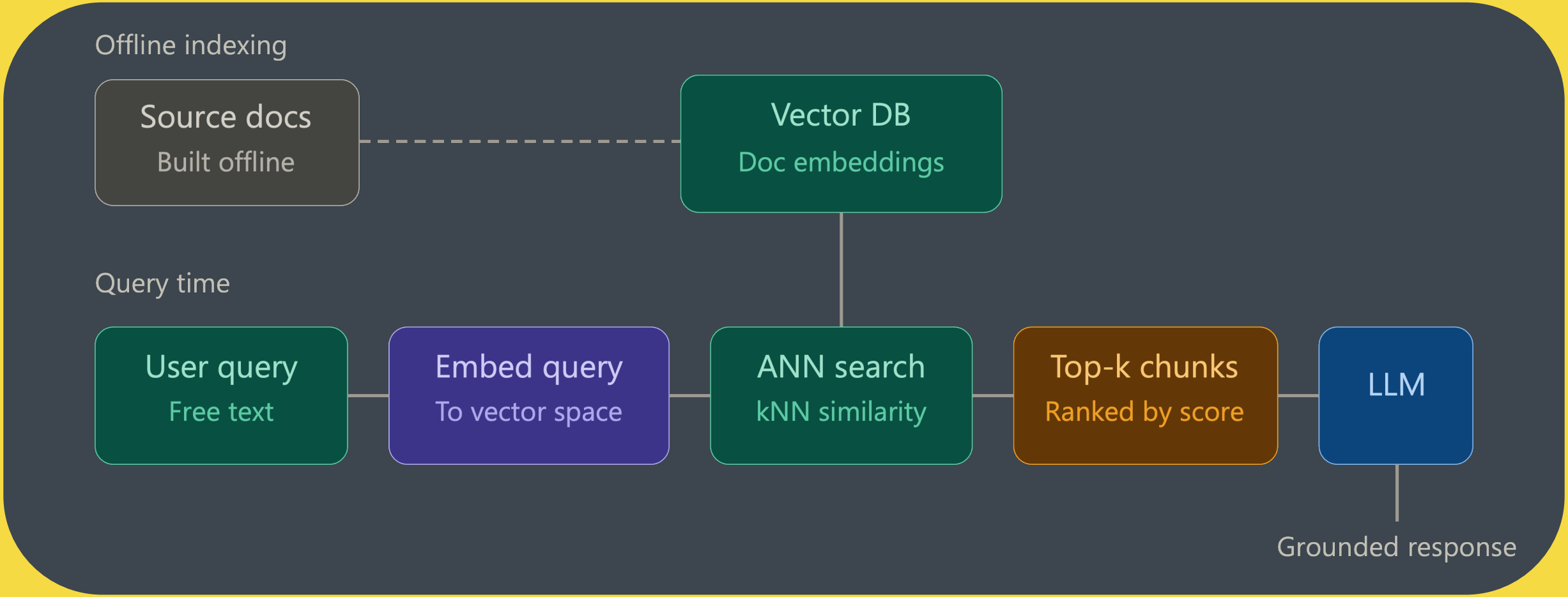


SURF

How does a RAG agent work?

- LLM based retrieval needs a **vector database** to enable **semantic retrieval**
- A **vector** is an object that has both a **magnitude** and **direction**, in contrast to a **scalar** which only contains a **magnitude**.
  - v = [1223444, -122344222, 53259, -12234875, …]
- A **vector database** maps v to a database entity or document
- Why is this useful; consider the following scenario
  - An image of a **horse** generates: v = [1,2,3,4,5,6]
  - An image of a **donkey** generates: v = [1,2,3,7,9]
- When querying the database for a ”**four legged animal**”
  - Search **v = [2,3,4,5,6,7]**
- Both objects will return with a **metric** describe **how close** they match to the document determining the **rank**

RAG pipeline

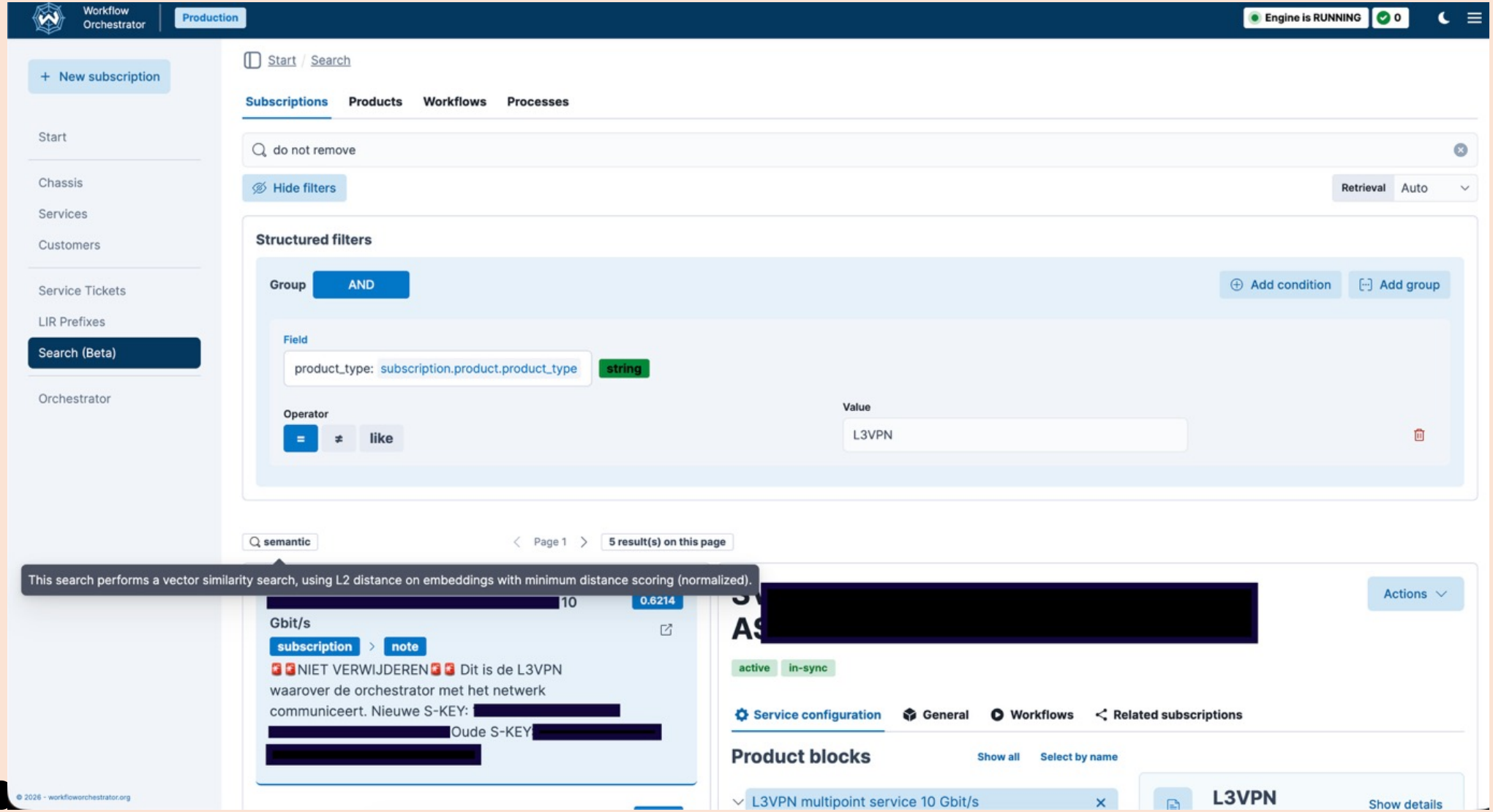


SURF

RAG use cases

- **Primary usecase:** Retrieval of documents
- **Secondary usecases implemented through tool calling:**
  - Aggregation
  - Counting
  - Exporting
  - ...
- By exposing tools to the AI Agent you can make it more **intelligent** and **reliable** in the way it returns results.

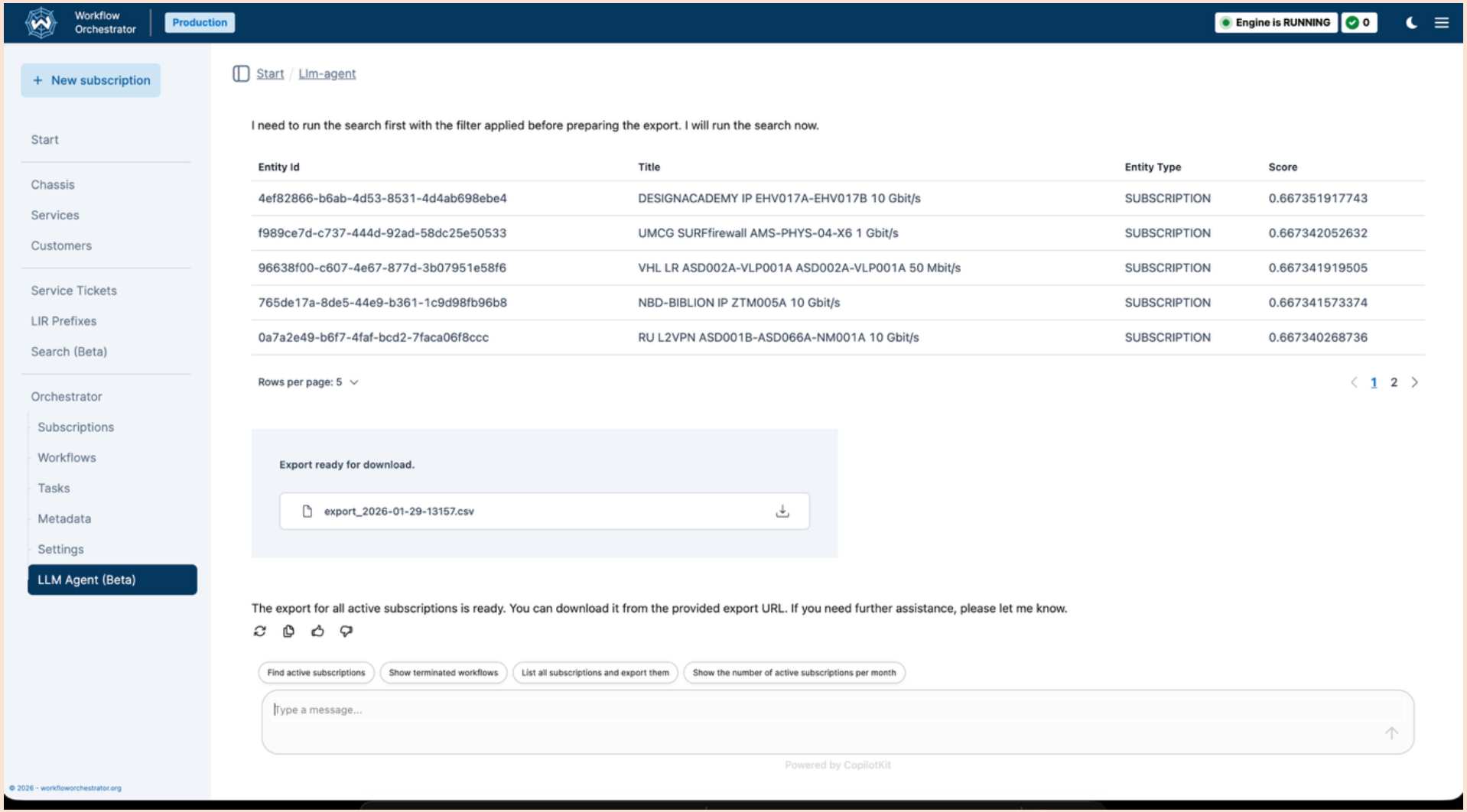
Results:


Workflow Orchestrator
Production
Engine is RUNNING
0
+ New subscription
Start
Chassis
Services
Customers
Service Tickets
LIR Prefixes
Search (Beta)
Orchestrator
Start / Search
Subscriptions
Products
Workflows
Processes
do not remove
Hide filters
Retrieval
Auto
Structured filters
Group
AND
Add condition
Add group
Field
product_type: subscription.product.product_type
string
Operator
=
≠
like
Value
L3VPN
semantic
Page 1
5 result(s) on this page
This search performs a vector similarity search, using L2 distance on embeddings with minimum distance scoring (normalized).
[redacted]10
0.6214
Gbit/s
subscription
note
NIET VERWIJDEREN Dit is de L3VPN waarover de orchestrator met het netwerk communiceert. Nieuwe S-KEY: [redacted] Oude S-KEY [redacted]
Actions
[redacted]
active
in-sync
Service configuration
General
Workflows
Related subscriptions
Product blocks
Show all
Select by name
L3VPN multipoint service 10 Gbit/s
L3VPN
Show details
© 2026 - workfloworchestrator.org

Results:


Workflow Orchestrator
Production
Engine is RUNNING
0
+ New subscription
Start
Chassis
Services
Customers
Service Tickets
LIR Prefixes
Search (Beta)
Orchestrator
Subscriptions
Workflows
Tasks
Metadata
Settings
LLM Agent (Beta)
Start / Llm-agent
The active subscriptions per month are counted and available, with the data spanning multiple months and years. If you need a specific format or further details, please let me know.
can you make it cumulative?
I'm setting temporal grouping
Count
Count Cumulative
5000
4000
3000
2000
1000
0
2018-10-01
2019-04-01
2019-09-01
2020-02-01
2020-07-01
2020-12-01
2021-05-01
2021-10-01
2022-03-01
2022-08-01
2023-01-01
2023-06-01
2023-11-01
2024-04-01
2024-09-01
2025-02-01
2025-07-01
2025-12-01
subscription start date month
The active subscriptions per month are now shown as a cumulative count. If you need anything else, please let me know.
Type a message...
Powered by CopilotKit
© 2026 - workfloworchestrator.org

Results:


Workflow Orchestrator
Production
Engine is RUNNING
0
+ New subscription
Start
Chassis
Services
Customers
Service Tickets
LIR Prefixes
Search (Beta)
Orchestrator
Subscriptions
Workflows
Tasks
Metadata
Settings
LLM Agent (Beta)
Start / Llm-agent
I need to run the search first with the filter applied before preparing the export. I will run the search now.
Entity Id
Title
Entity Type
Score
4ef82866-b6ab-4d53-8531-4d4ab698ebe4
DESIGNACADEMY IP EHV017A-EHV017B 10 Gbit/s
SUBSCRIPTION
0.667351917743
f989ce7d-c737-444d-92ad-58dc25e50533
UMCG SURFfirewall AMS-PHYS-04-X6 1 Gbit/s
SUBSCRIPTION
0.667342052632
96638f00-c607-4e67-877d-3b07951e58f6
VHL LR ASD002A-VLP001A ASD002A-VLP001A 50 Mbit/s
SUBSCRIPTION
0.667341919505
765de17a-8de5-44e9-b361-1c9d98fb96b8
NBD-BIBLION IP ZTM005A 10 Gbit/s
SUBSCRIPTION
0.667341573374
0a7a2e49-b6f7-4faf-bcd2-7faca06f8ccc
RU L2VPN ASD001B-ASD066A-NM001A 10 Gbit/s
SUBSCRIPTION
0.667340268736
Rows per page: 5
1 2
Export ready for download.
export_2026-01-29-13157.csv
The export for all active subscriptions is ready. You can download it from the provided export URL. If you need further assistance, please let me know.
Find active subscriptions
Show terminated workflows
List all subscriptions and export them
Show the number of active subscriptions per month
Type a message...
Powered by CopilotKit
© 2026 - workfloworchestrator.org

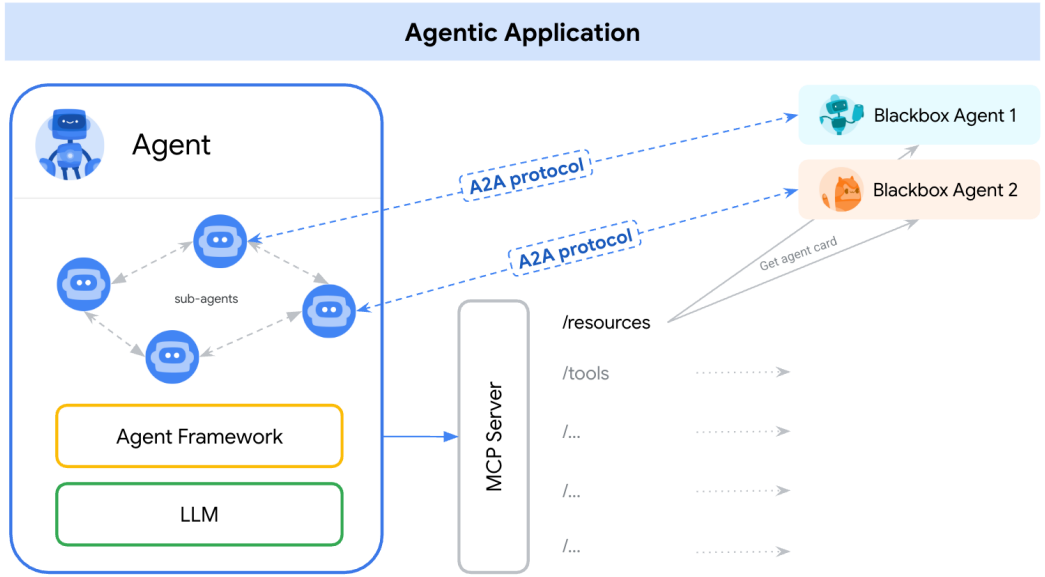
What did we learn?

- When creating an AI agent, managing the context of your conversation is key.
  - First iterations had poor performance as the workflow described in the System prompt was too complex
- Not all LLM's are suitable for **Agentic** architectures
  - Older opensource models are not trained to drive the **ReAct** loop.
- Delegate deterministic tasks to stateless tool calling as much as possible
- **Libraries and standards are moving targets**
  - **Difficult to judge whether you have made the "right" choice**
- ***We want more!***

SURF

A solution to the context problem? A2A

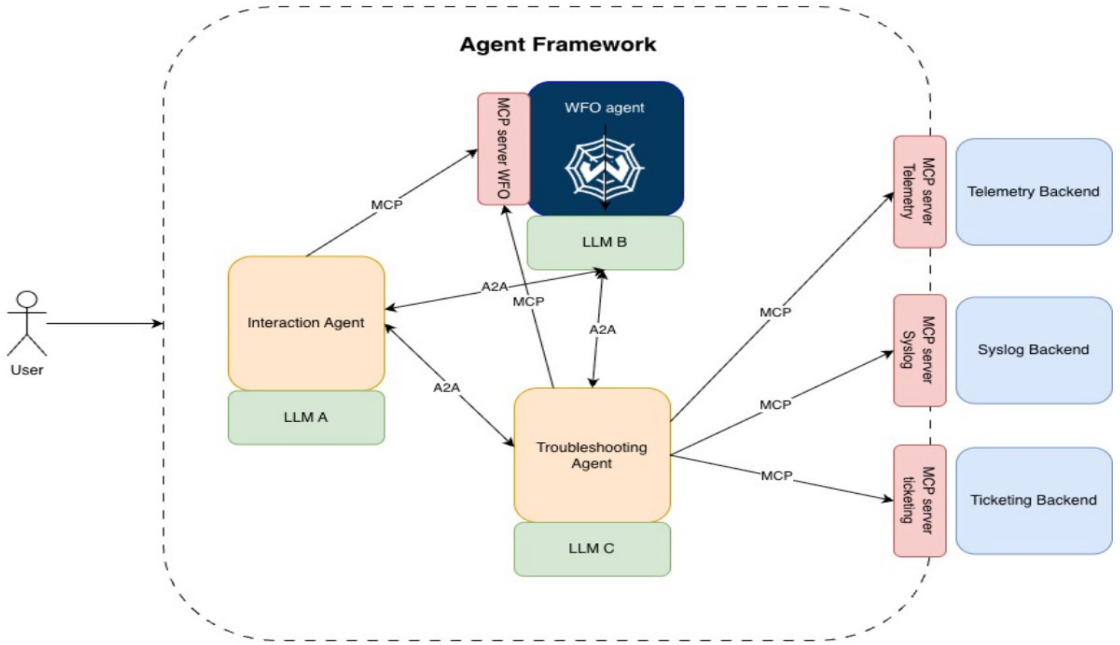
- Agent2Agent (A2A) protocol is designed by Google to build and design an Agentic Application
- Each Agent within the constellation of Agents can communicate its capabilities in a standardised way through a service discovery mechanism
- Sessions, artefacts, memory and context are formalised constructs that may be shared between agents.
- Together the Agents work in a large **ReAct** loop to solve a certain task. However the context that each agents uses remains small
- **Microservice architecture for agents**



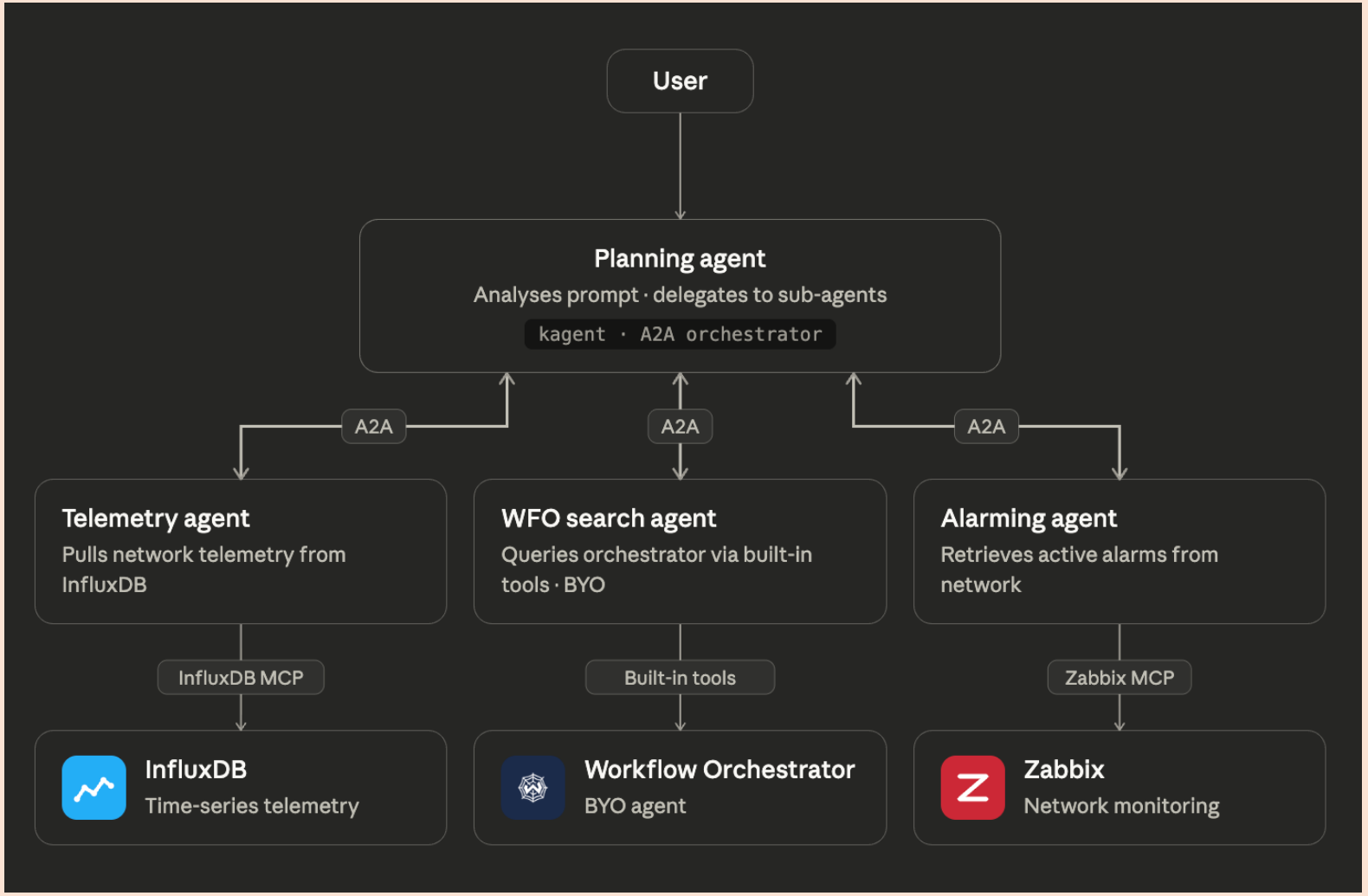
a2aprotocol.ai

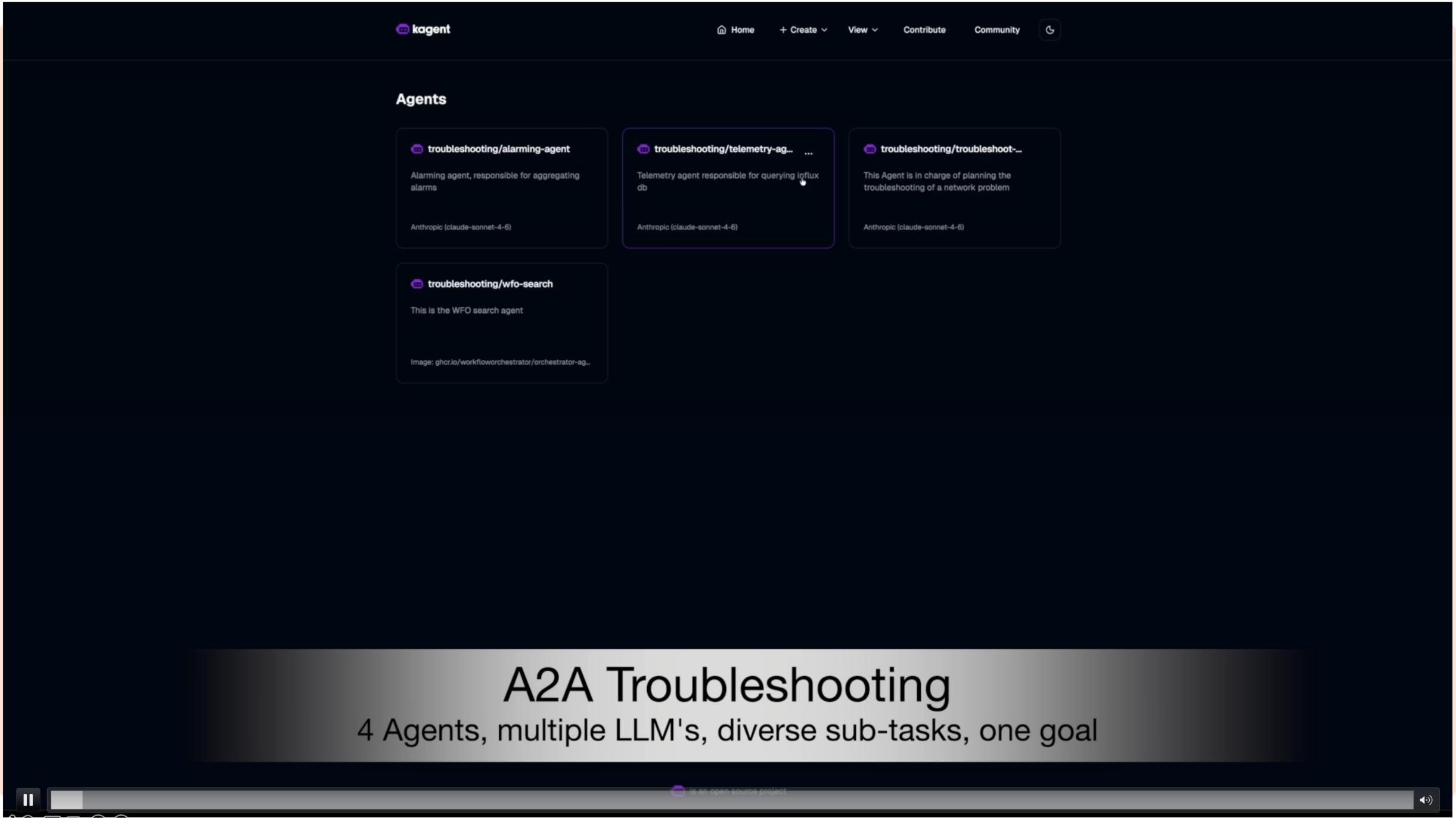
GN-52 is sponsoring SURF to build an A2A extention to the Workflow Orchestrator

- **Goal: Enable high quality consistent reasoning, across multiple agents by leveraging, Agent delegation, tool use and MCP servers to troubleshoot network incidents.**
- Interactions with the Agent architecture should be:
  - Observable
  - Reliable
  - Deterministic
  - Consistent
  - User Friendly
  - **Time Saver**

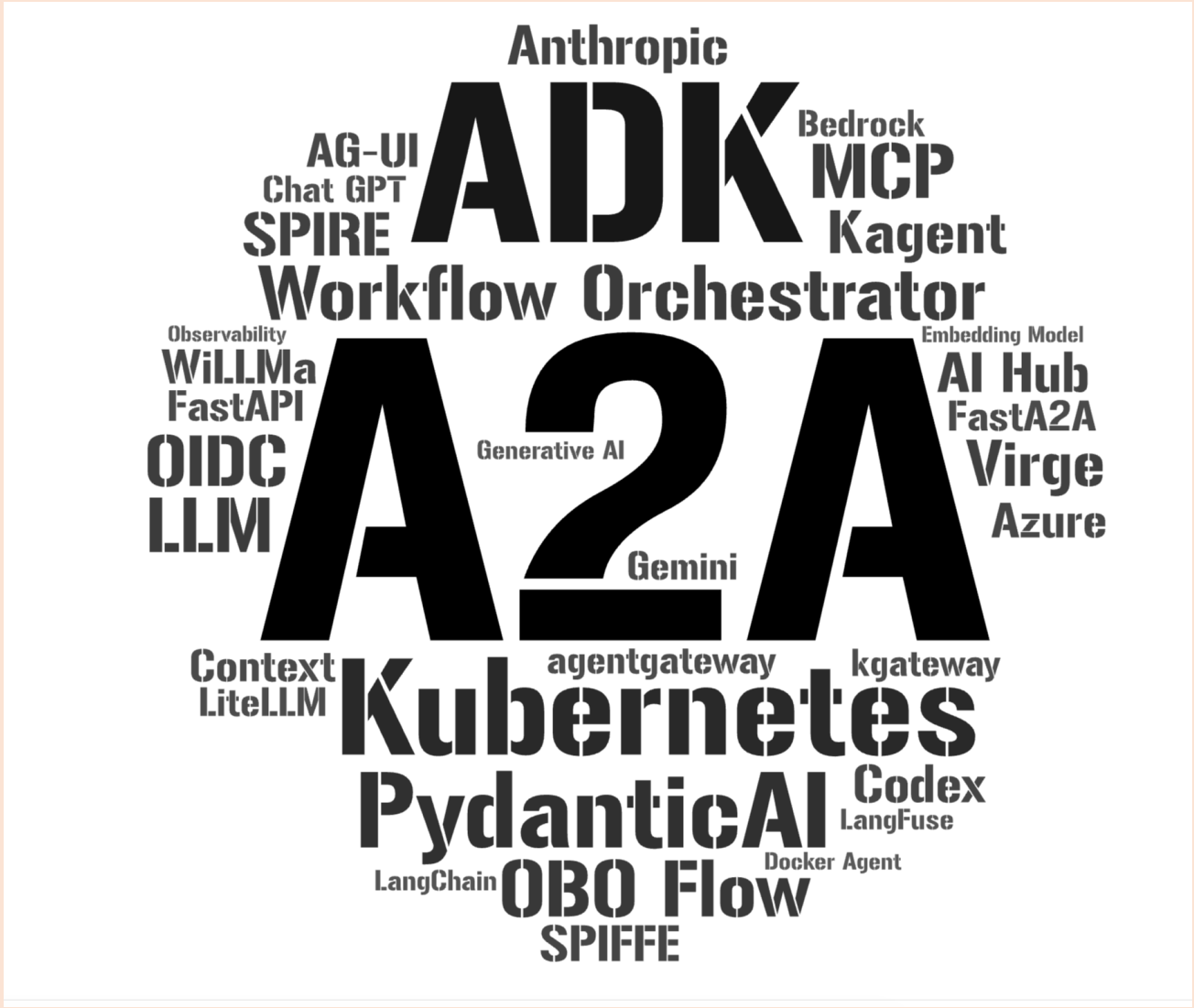


Demo: Current state (1)

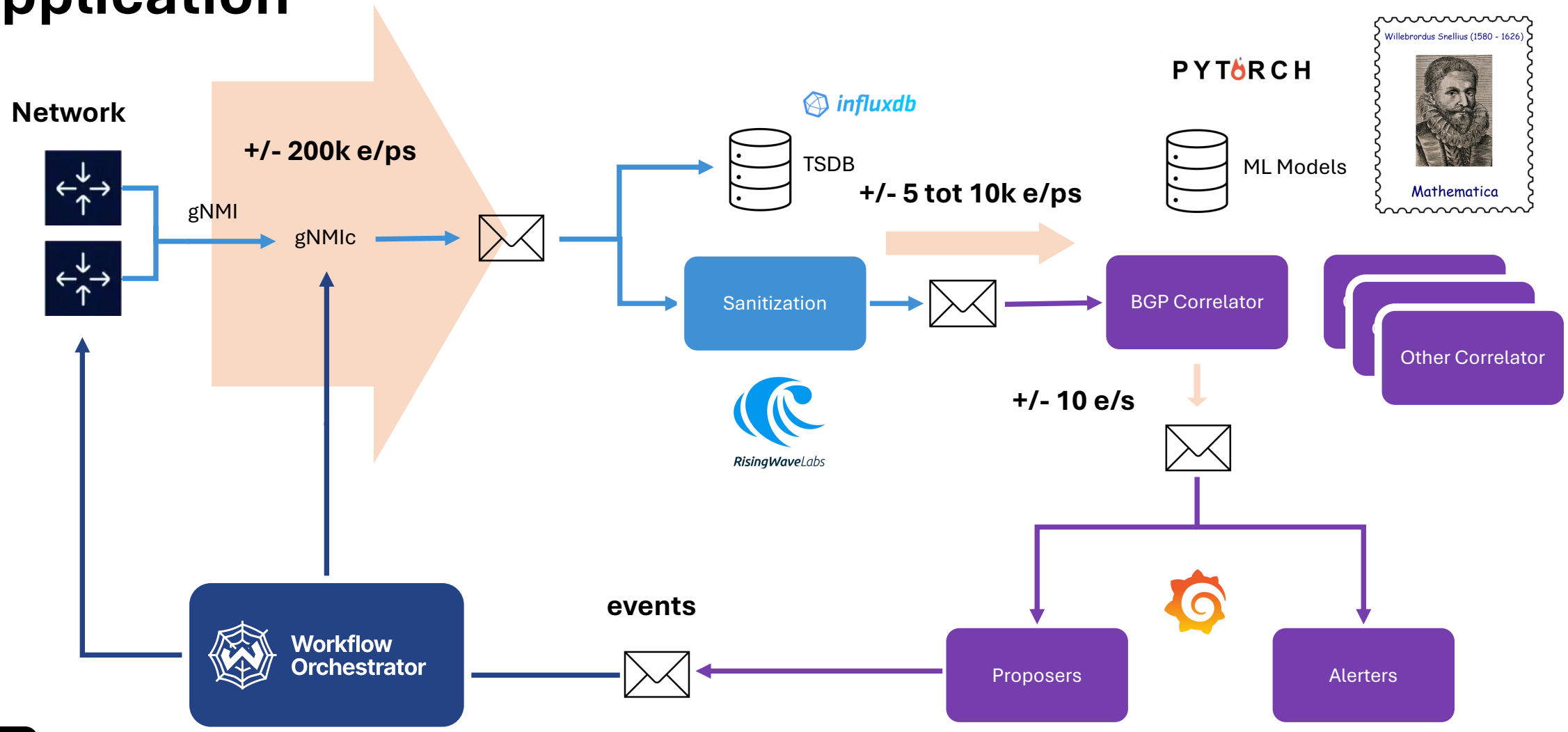



kagent
Home
+ Create
View
Contribute
Community
Agents
troubleshooting/alarming-agent
Alarming agent, responsible for aggregating alarms
Anthropic (claude-sonnet-4-6)
troubleshooting/telemetry-ag...
Telemetry agent responsible for querying influx db
Anthropic (claude-sonnet-4-6)
troubleshooting/troubleshoot-...
This Agent is in charge of planning the troubleshooting of a network problem
Anthropic (claude-sonnet-4-6)
troubleshooting/wfo-search
This is the WFO search agent
Image: ghcr.io/workfloworchestrator/orchestrator-ag...
A2A Troubleshooting
4 Agents, multiple LLM's, diverse sub-tasks, one goal
is an open source project

Current state (3)

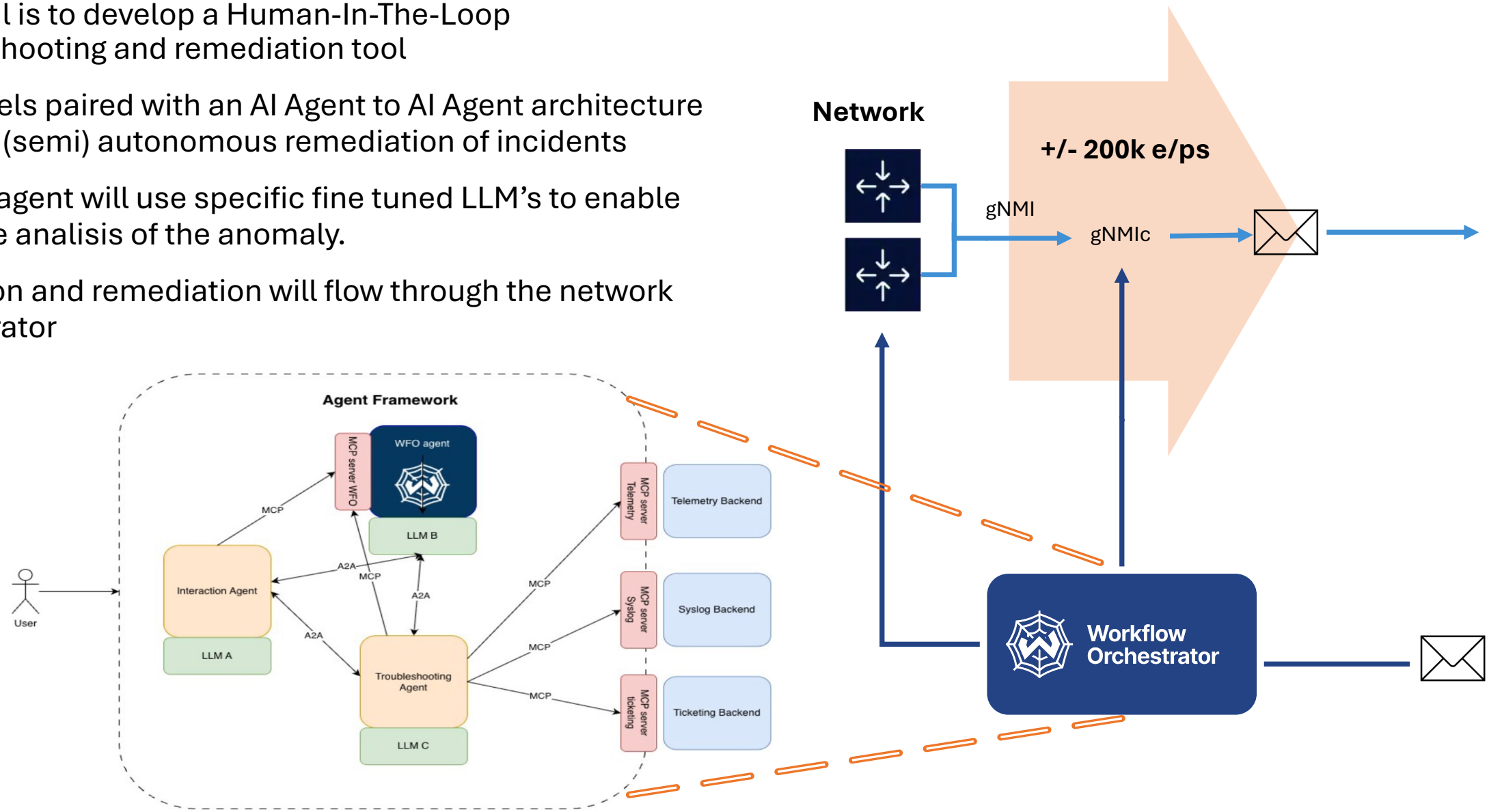


Tying the ML pipelines together with the Agentic Application



2026: 1st iteration of an Intelligent network

- First goal is to develop a Human-In-The-Loop troubleshooting and remediation tool
- ML models paired with an AI Agent to AI Agent architecture enables (semi) autonomous remediation of incidents
- Each AI agent will use specific fine tuned LLM's to enable accurate analisis of the anomaly.
- Validation and remediation will flow through the network orchestrator



Conclusion (so far)

- Fast moving landscape
- Some issues are still dificult to solve, Auth(n|z)
- More experimentations to be done:
  - **TNC** GNA-G automation WG. A2A troubleshooting between NREN's???
- **We still need to be able to pivot to different frameworks and technology fast. AI helps with that.**

Are there any questions?

- Scan QR code for Workflow Orchestrator resources

Source Code

edu.nl/eqgwt

Discord

edu.nl/nbnj6