

# STAN - Personalized HomeLLM

An agentic AI assistant with authorized access  
to your home and personal data

# AI-integration in smart devices is increasing rapidly

- Growing number of edge-devices with AI hardware
- AI is becoming more and more capable of complex tasks



# GPT-3.5 and GPT-4 performance using zero-shot and agent workflows

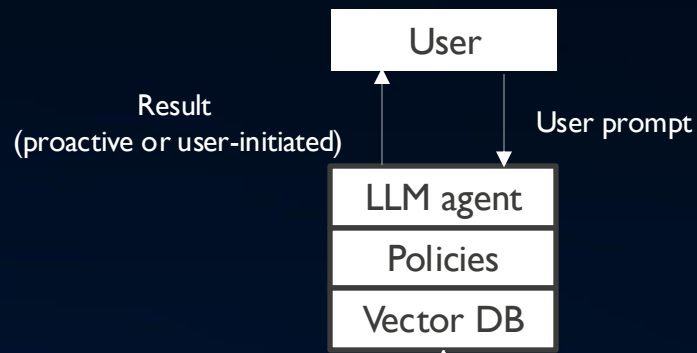






Performance of GPT-3.5 and GPT-4 (zero-shot) on HumanEval, along with algorithms that use agent workflows on top of GPT-3.5 or GPT-4. Thanks to Joaquin Dominguez and John Santerre for help with this analysis.

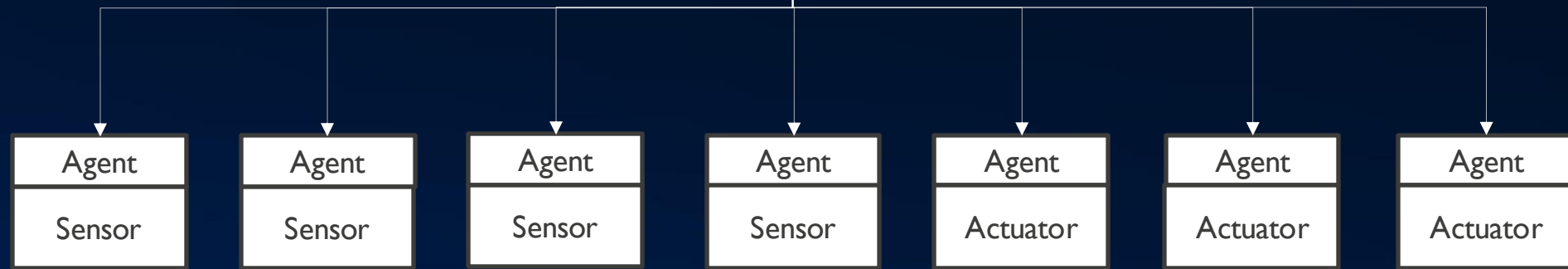
## PSX MISSION:

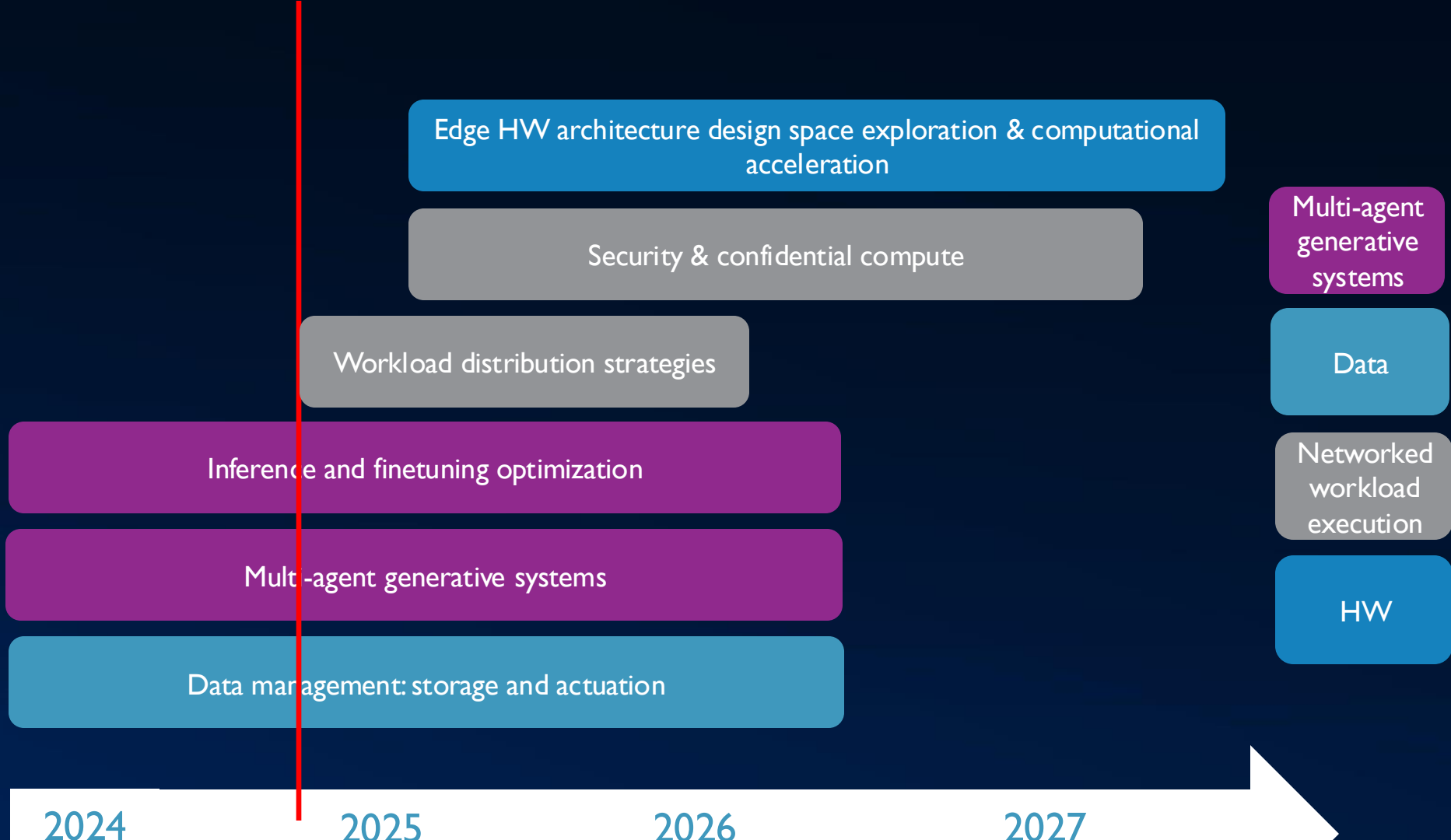
Extreme edge systems that allow demanding agentic genAI workloads to run on highly **sensitive** data while optimizing **performance**

# AGENTIC EDGE ARCHITECTURE



-  Sensor subagent
-  Actuator subagent
-  Orchestration subagent
-  Document subagent





Edge HW architecture design space exploration & computational acceleration

Security & confidential compute

Workload distribution strategies

Inference and finetuning optimization

Multi-agent generative systems

Data management: storage and actuation

Multi-agent generative systems

Data

Networked workload execution

HW

2024

2025

2026

2027

Multi-agent  
generative  
systems

Data

Networked  
workload  
execution

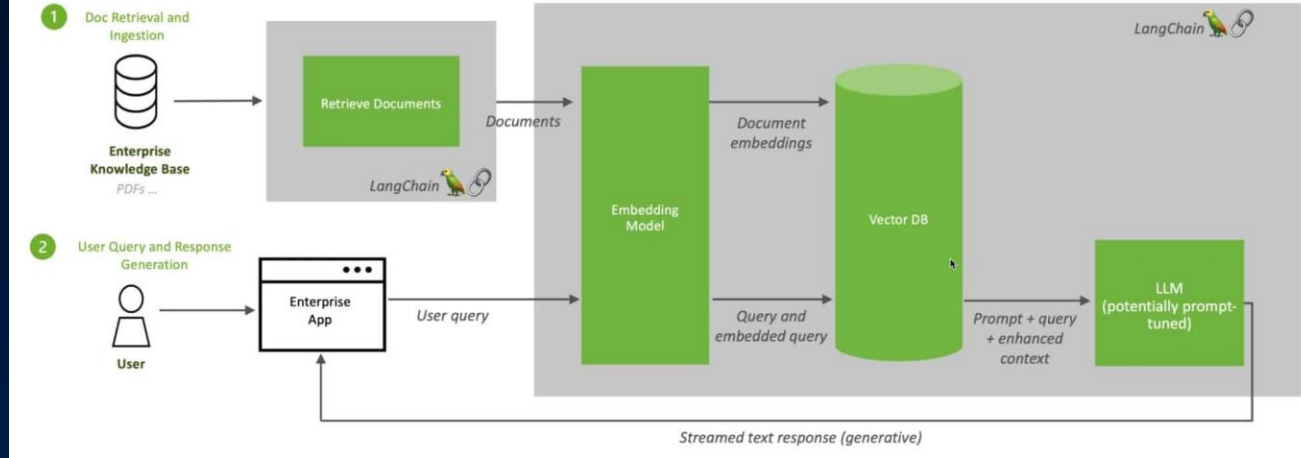
HW





Retrieval-Augmented Generation (RAG) is a neural network architecture that combines the capabilities of information retrieval and natural language generation. It is designed to enhance the generation of text by incorporating relevant information retrieved from a large corpus of data.

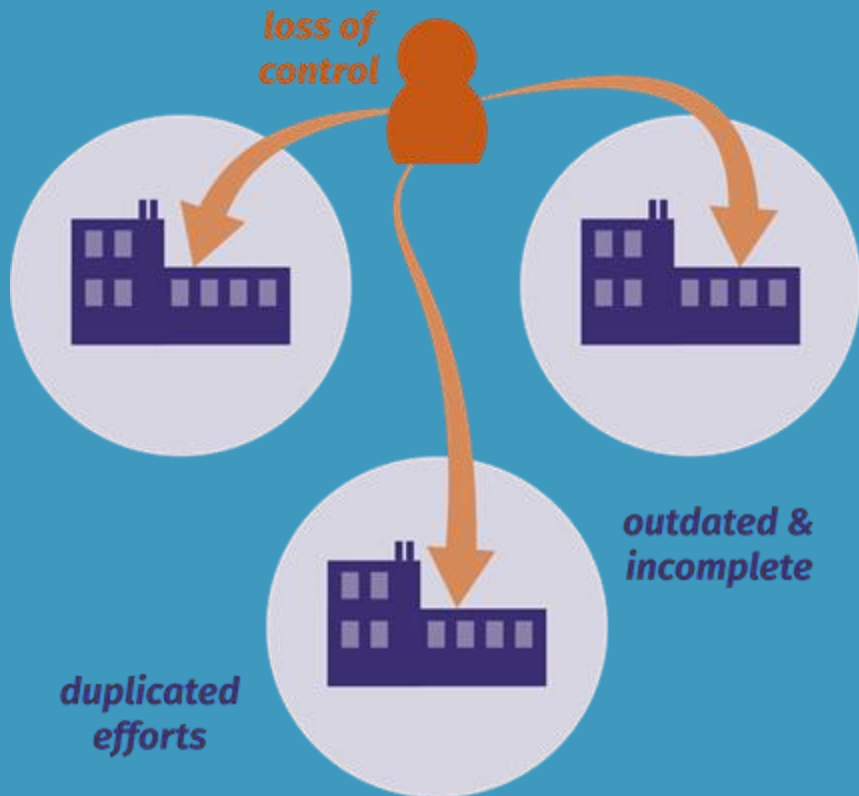
### Retrieval Augmented Generation (RAG) Sequence Diagram



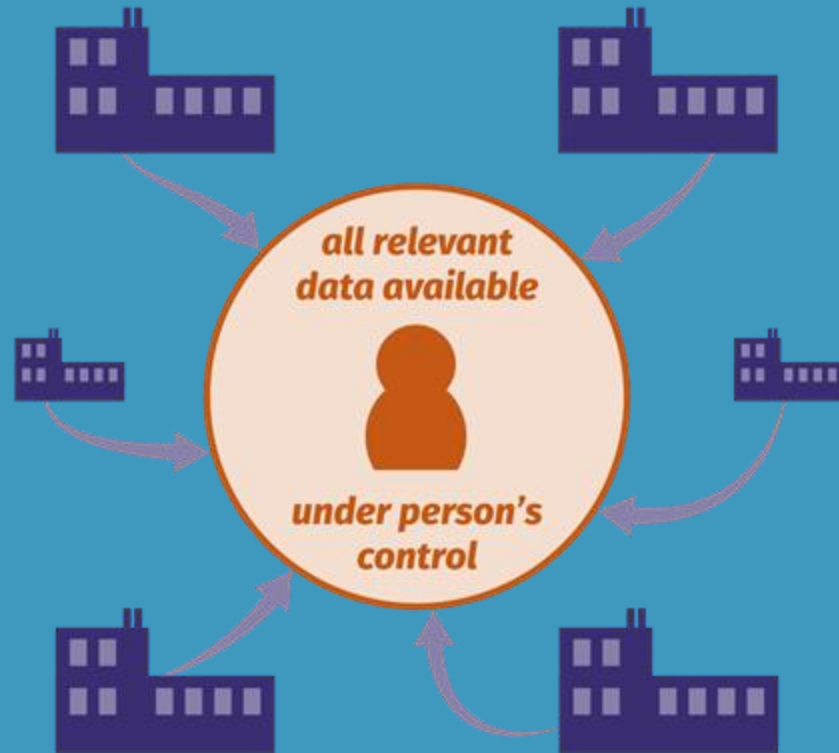
<https://blogs.nvidia.com/blog/what-is-retrieval-augmented-generation/>



## data stored by service providers



## data stored in a personal data pod



# Extreme Edge agentic networks will change the way consumers interact with their smart home ecosystem

## Usability:

Steer and manage all your devices with one Assistant using natural language/voice

## Personalization:

Unlock the personalization power of an AI assistant in any edge device, based on large data volumes

## Collective intelligence:

Enable collaboration between agentic devices to realize new and advanced applications

## Integration:

Seamlessly onboard any edge device, regardless of platform or OS



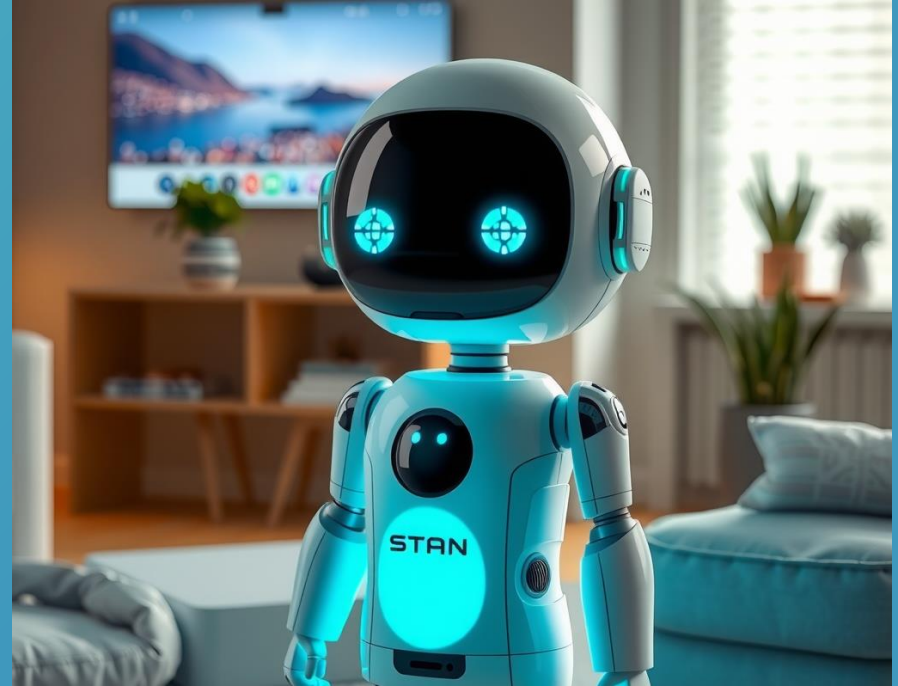
# Meet the Cohousing Trio

Sandra, Blake and Adam are three friends living together in a smart home.



# Enter STAN, the Personalized Home Assistant

STAN answers questions about the home, helps with energy management, but can also help with personal questions!



# Sharing Data with Housemates

Housemates decide what STAN can share with others.

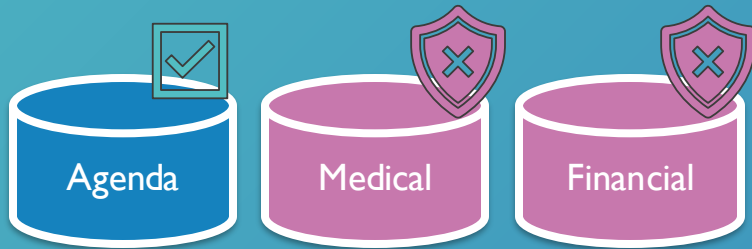


Organizing a drink? Sure, they can see my schedule.



# Privacy in a Smart Home

Sandra prefers not to share her medical info with others.



# STAN

On tour





# STAN

## Accessing my own calendar

The screenshot shows the STAN interface with a user named Sandra Johansson. Stan, the personal home assistant, has responded to the query "When's my Art history class?". The response lists four dates and times for Art History classes: October 21st, 22nd, 23rd, and 24th, all from 8:30 AM to 3:00 PM. A "PERSONAL DATA VAULT ACCESS" panel is open on the right, showing details for the user Sandra, including her calendar type, data ownership, and a granted access status.

**Stan**  
the personal home assistant

Sandra Johansson

When's my Art history class?

Based on the provided schedule, I can see that you have multiple classes scheduled for Art History. Here are the dates and times:

- October 21st: 8:30 AM - 3:00 PM
- October 22nd: 8:30 AM - 3:00 PM
- October 23rd: 8:30 AM - 3:00 PM
- October 24th: 8:30 AM - 3:00 PM

Your Art History class is scheduled for these dates and times.

**PERSONAL DATA VAULT ACCESS**

User **Sandra**  
Type **Calendar**  
Data owner  
**Sandra**  
Status **Access granted**

# STAN

## Accessing my roommate's calendar

**Stan**  
the personal home assistant

Sandra Johansson

Check Blake's calendar, when does he have free time Friday afternoon?

To answer your question, I'll need to analyze the provided data and find any events that occur on a Friday afternoon.

Here's how I can do it:

1. Filter the events by date: Since you're interested in finding free time on Friday afternoon, I'll look for events that start after 12 PM (noon) on Fridays.
2. Check if there are any overlapping events: If multiple events occur at the same time, I'll need to determine which ones Blake has committed to attending.

After analyzing the data, I found the following event:

- "Game Day with Friends" - This event starts at 17:00 (5 PM) on Friday and ends at 20:00 (8 PM).

Since this event starts after 12 PM (noon), it's a potential candidate for Blake's free time. However, since he has committed to attending the game day, his calendar is not entirely free during that time.

If you'd like me to find other events or provide more information, let me know!

PERSONAL DATA VAULT ACCESS

User Sandra	User Sandra
Type Calendar	Type Calendar
Data owner Blake	Data owner Sandra
Status Access granted	Status Access granted

Ask me about your personal data, smart home devices, or anything else

# STAN

## Accessing my roommate's financial info

Check Blake financial info, how much debt is he in?

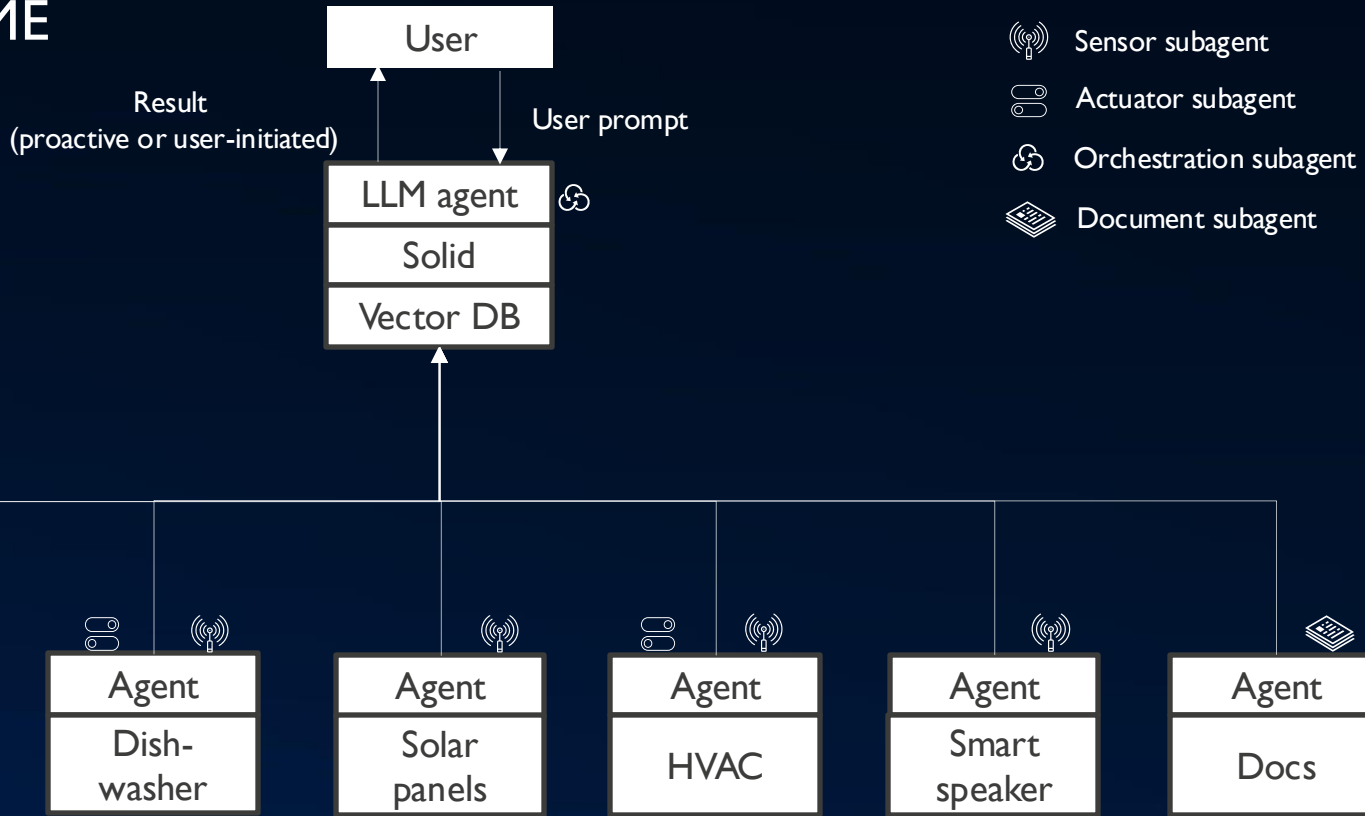
Sorry, I was unable to get the required information to answer the something related, like checking Blake's calendar or schedule if

PERSONAL DATA VAULT ACCESS

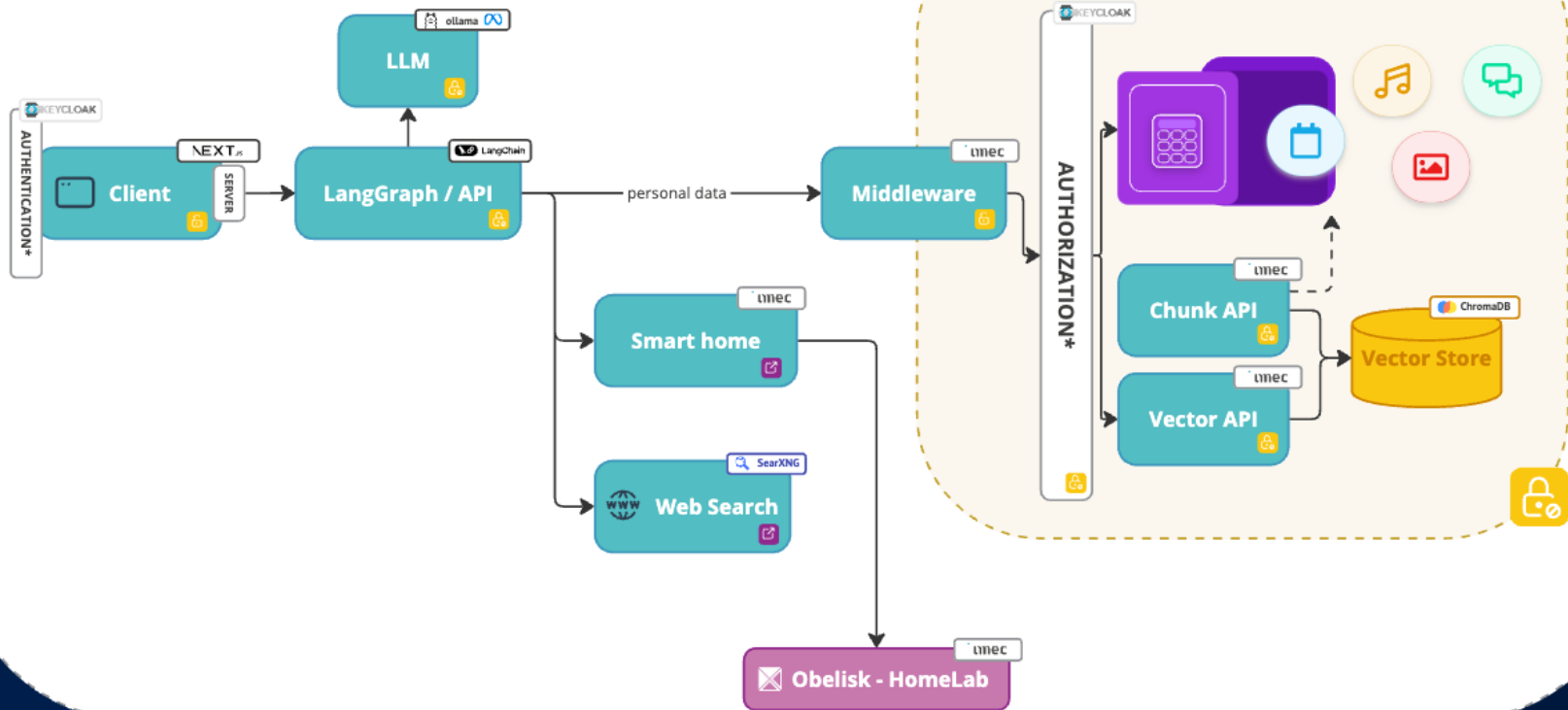
User <b>Sandra</b>	User <b>Sandra</b>
Type <b>Financial</b>	Type <b>Financial</b>
Data owner <b>Blake</b>	Data owner <b>Sandra</b>
Status <b>Access denied</b>	Status <b>Access granted</b>

Ask me about your personal data, smart home devices, or anything else

# THE AGENTIC HOME



SHORT TERM: NVIDIA JETSON  
 LONGER TERM: EXTREME EDGE NETWORK OF DEVICES



# Authentication and Authorization with Keycloak

- Log in with Keycloak to receive Access Token (JWT)
- Policies by User and Type
  - Calendar, Financial, Medical
  - Fixed Personae (Adam/Blake/Sandra)
- On request polices are validated for the user and type.  
eg. Does Blake have access on Adams calendar data? (Adam Calendar Policy)

Name	Type	Dependent permission
> Adam All Policy	User	Adam All Permission <a href="#">1 more</a>
▼ Adam Calendar Policy	User	Adam Calendar Permission
Dependent permission	Adam Calendar Permission	
> Adam Emails Policy	User	Adam Emails Permission
> Adam Financial Policy	User	Adam Financial Permission
> Adam Medical Policy	User	Adam Medical Permission
> Adam Notes Policy	User	Adam Notes Permission
> Any User Policy	User	Any User Permission <a href="#">1 more</a>
> Blake All Policy	User	Blake All Permission
> Blake Calendar Policy	User	Blake Calendar Permission

# SOLID/RAG Toolbox

- **Chunk API**
  - Splits (in chunks) different input documents to be stored as embeddings in the Vector DB
  - Different strategies can be used (eg. Paragraphs, sentences, number of characters, ..)
  - Currently only supports plain text and json
  - Type is added for the entire document (policy checking)
- **Vector API**
  - Search for document embeddings
  - Can return the entire document or only the closest found chunk
- **ChromaDB**
  - Vector DB used to store and search for embeddings

- Keycloak Middleware

- Gateway to the Vector API
  - Can be extended to include other API's/databases
- Checks policies based on request (Authorization header)
- Gets the data from the requested resource and returns it

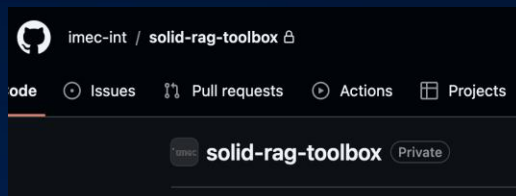
```
POST keycloak-middleware-service:8080/vector-api/invoke

Params Authorization Headers (9) Body Scripts Settings

none form-data x-www-form-urlencoded raw binary GraphQL JSON

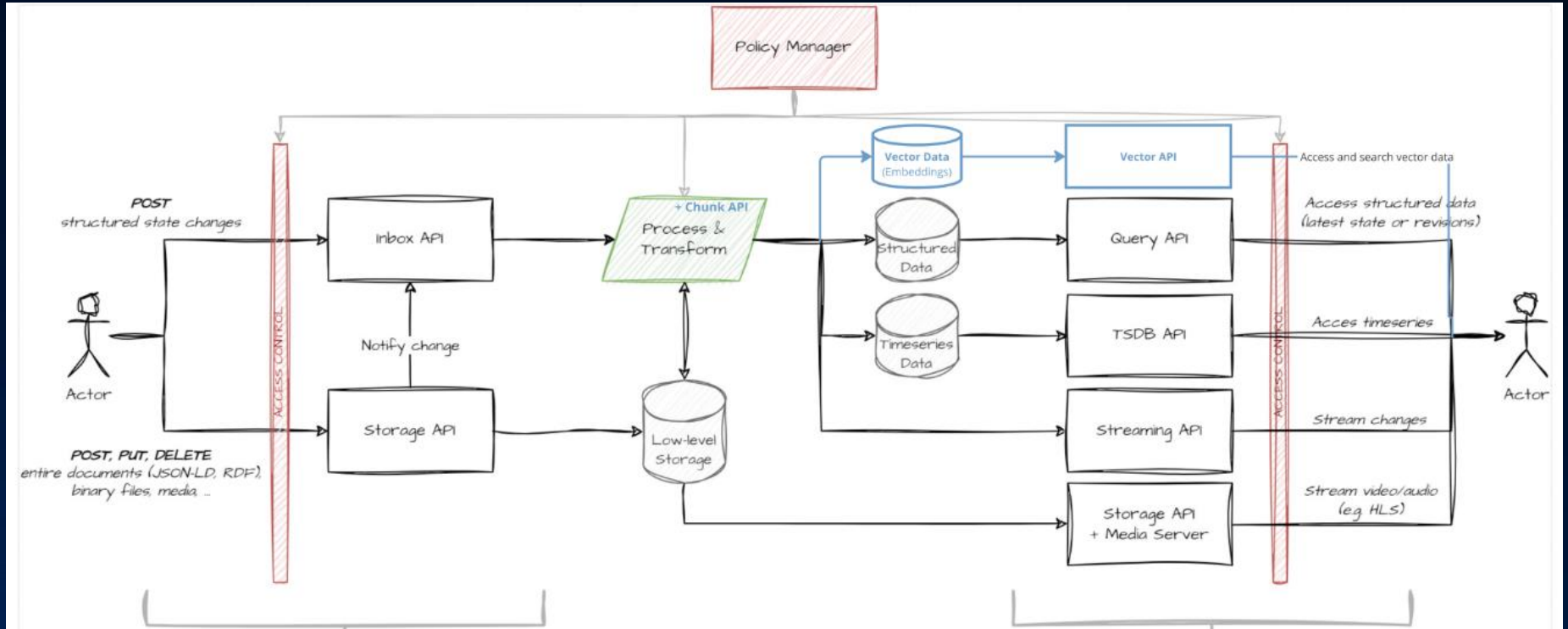
1 {
2   "question": "meetings on 15/11/2024",
3   "user_data_requested": [
4     "adam@psx.be"
5   ],
6   "type": "calendar",
7   "full_document": "True"
8 }
```

- Coming soon..





# Next steps



## For more info:

Jan Deprez

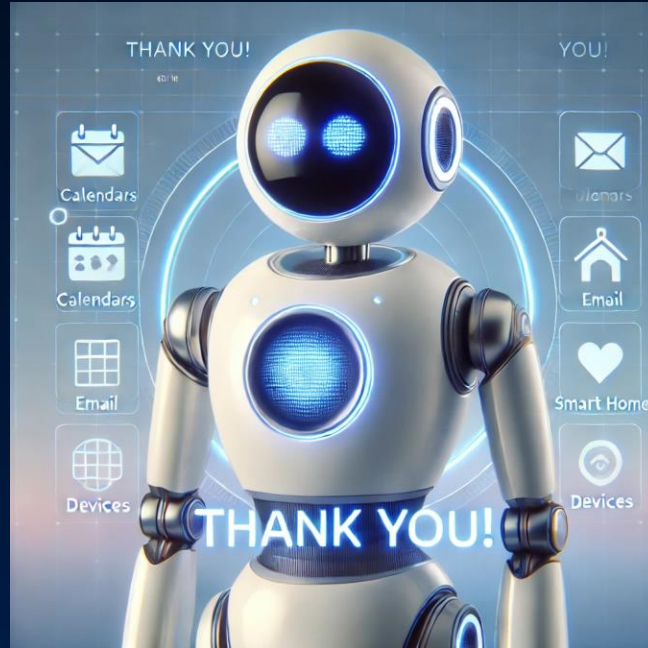
[jan.deprez@imec.be](mailto:jan.deprez@imec.be)

Maarten Taeymans

[maarten.taeymans@imec.be](mailto:maarten.taeymans@imec.be)

Tanguy Coenen

[tanguy.coenen@imec.be](mailto:tanguy.coenen@imec.be)





mtec

embracing a better life