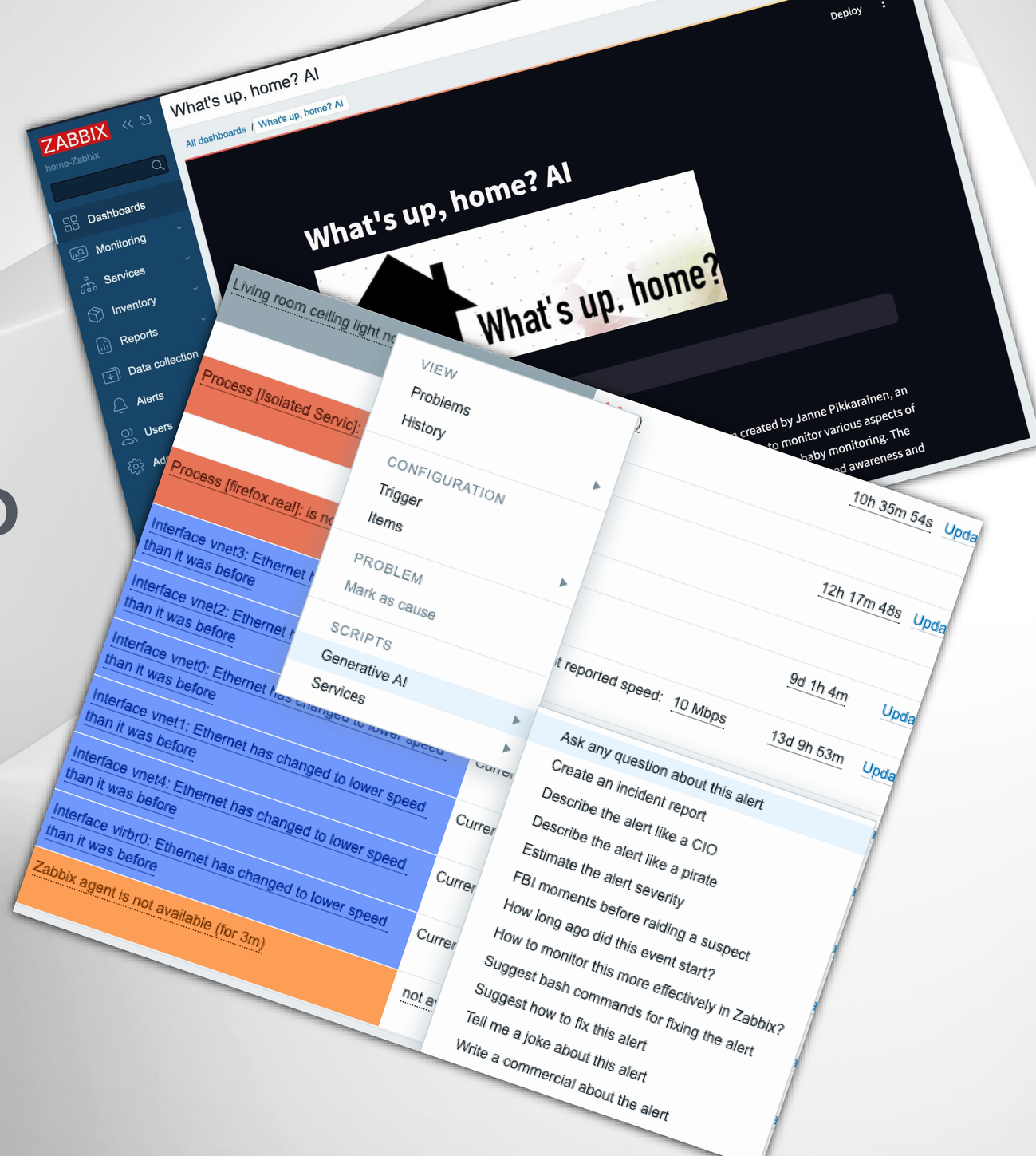
What’s up, home?

Goes Generative AI



Janne Pikkarainen <contact@whatsuphome.fi>

Who am I?

- Janne Pikkarainen from Finland FI
- Sitting in front of computers since 1986
- Doing IT for living for 25+ years now, out of that monitoring since 2001
- 10+ years at Forcepoint, currently as Lead Site Reliability Engineer
- Walking, talking Zabbix commercial

About Forcepoint

- Provides cyber security for the biggest enterprises you can imagine of, hospitals, critical infrastructure and so forth
- NGFW, SD-WAN, Web/e-mail protection, VPN, ZTNA, RBI, DLP, CASB, many other acronyms and products
- Global footprint with employees & customers spread all over the world

Quick recap of my blog

- Been doing my blog since March 2022
- I monitor my home with Zabbix
- In addition to home, I monitor and do just about anything with Zabbix
- During 2023, started to integrate Zabbix with a locally run LLM, GPT4All
- See https://whatsuphome.fi/

I monitor everything with Zabbix

More info at https://whatsuphome.fi/

AdGuard Home
Air conditioner
Air humidifier
Airport departures/arrivals
Air quality index (outdoors)
Apple Watch
Baby sleep
Baby stroller temperature
BackupPC
“Banana” color
Car location
CCTV camera
CO2 levels (indoors)
Cozify
Countdown timer
Docker
Dog in bed?
Door sensors
Elasticsearch
Electricity price
Electricity consumption
Facial cream usage
FlightGear
HAProxy
HashiCorp Vault
Headset
Home router
HP LaserJet
Jenkins
Laptop webcam
Lights
Logs
Lunch menus
Maritime traffic
Mobile data usage
Motion sensors
MySQL
Northern Lights
Philips OneBlade
Power sockets
PostgreSQL
Product prices
Raspberry Pi 4
Roomba
RSS feeds
Selenium
Smoke/fire alarm
Sonos smart speaker
Thermometers
TV
Weather
whatsuphome.fi website
whatsuphome.fi visitors
Zabbix Security Advisories

Also: Blender 3D integration, Star Wars screensaver integration, Fairy Tale media type, among other weird stuff

Today’s agenda

What is GPT4All?

Integrate GPT4All with Zabbix context menus

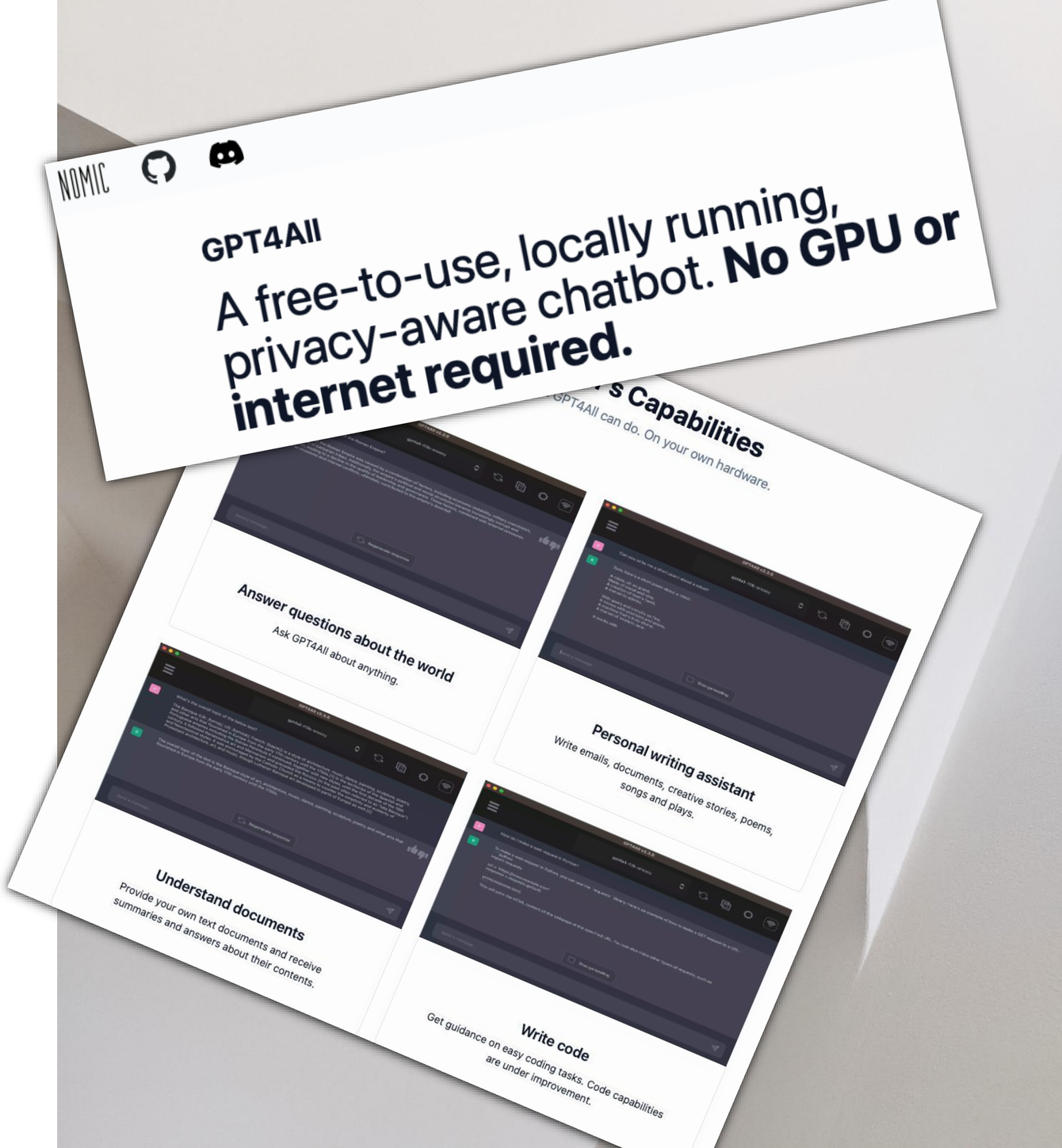
Let GPT4All write its own blog based on active Zabbix alerts

Embed your Zabbix data with GPT4All

Give Zabbix a personality through GPT4All

What is GPT4All?

- GPT4All is an open-source project to provide a Generative AI for almost any hardware
- Supports 1000+ different language models
- Locally run, no need for account or Internet, so your data is safe
- Has a GUI client, but also Python/Node/other bindings
- Through Python, very easy to integrate with Zabbix



Hello world in GPT4All

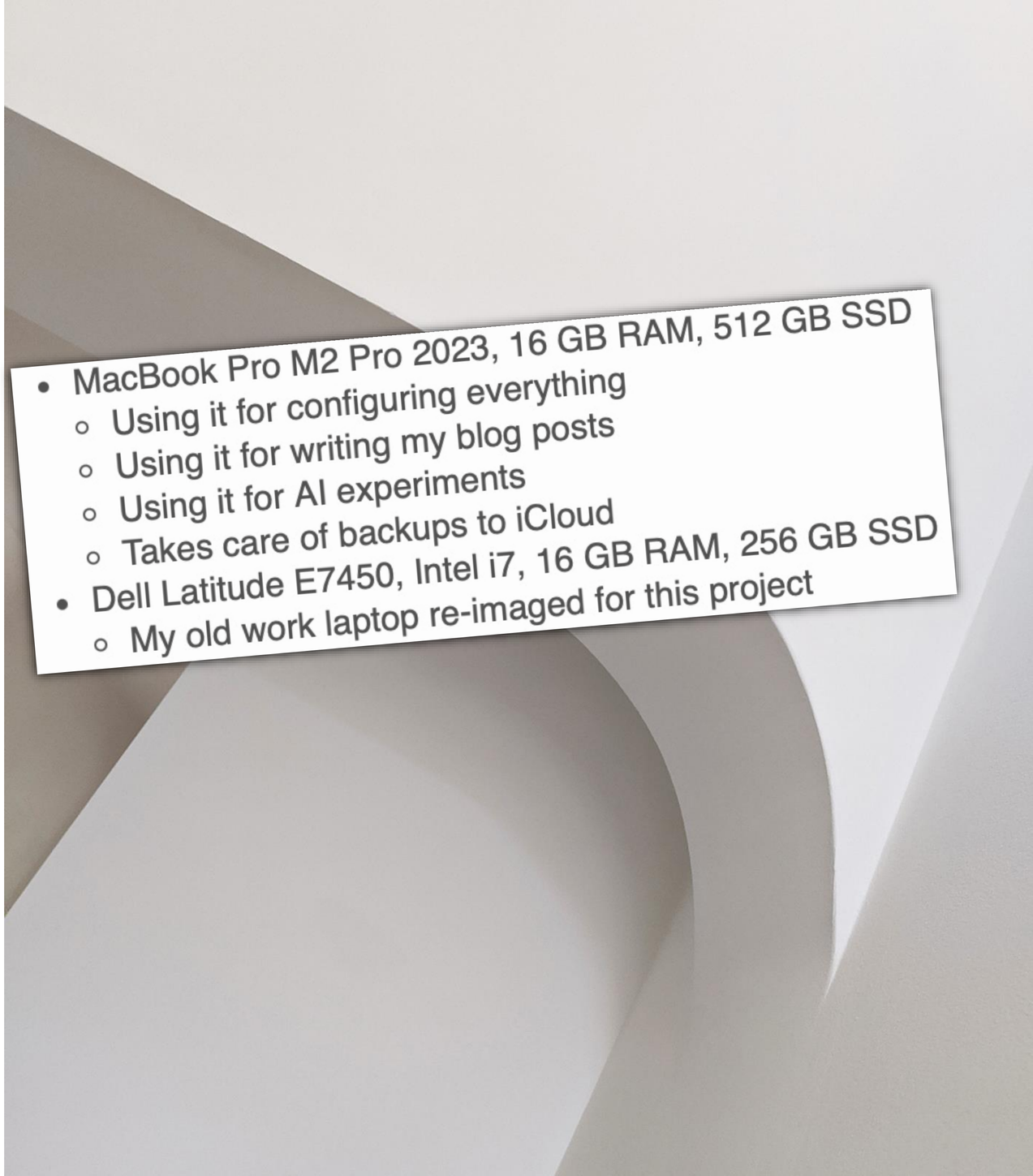
.. and Python

```
from gpt4all import GPT4All

model = GPT4All("orca-mini-3b-
 gguf2-q4_0.gguf")

output = model.generate(“Say hello
 to Zabbix Summit”, max_tokens=3)

print(output)
```



GPT4All + Zabbix context menus

Let AI to help you inside Zabbix

Enrich your alerts dashboard

How to do that?

Through Zabbix Scripts!

```
from gpt4all import GPT4All

import argparse

parser = argparse.ArgumentParser(description='Pass question to GPT4All')

parser.add_argument('-q', '--question')

args=parser.parse_args()

model = GPT4All('wizardlm-13b-v1.2.Q4_0.gguf')

system_template = 'A chat between a curious user and an artificial intelligence assistant.'

prompt_template = 'USER: {0}\nASSISTANT: '

with model.chat_session(system_template, prompt_template):

  response1 = model.generate(args.question)

  print(response1)
```

Pros and cons

- Very easy to implement and extend
- Integrates with Zabbix menus nicely
- … but the output dialog is tiny
- … no back and forth chat
- … does not support the interactive ways of Zabbix 7.0’s inter-widget communication framework

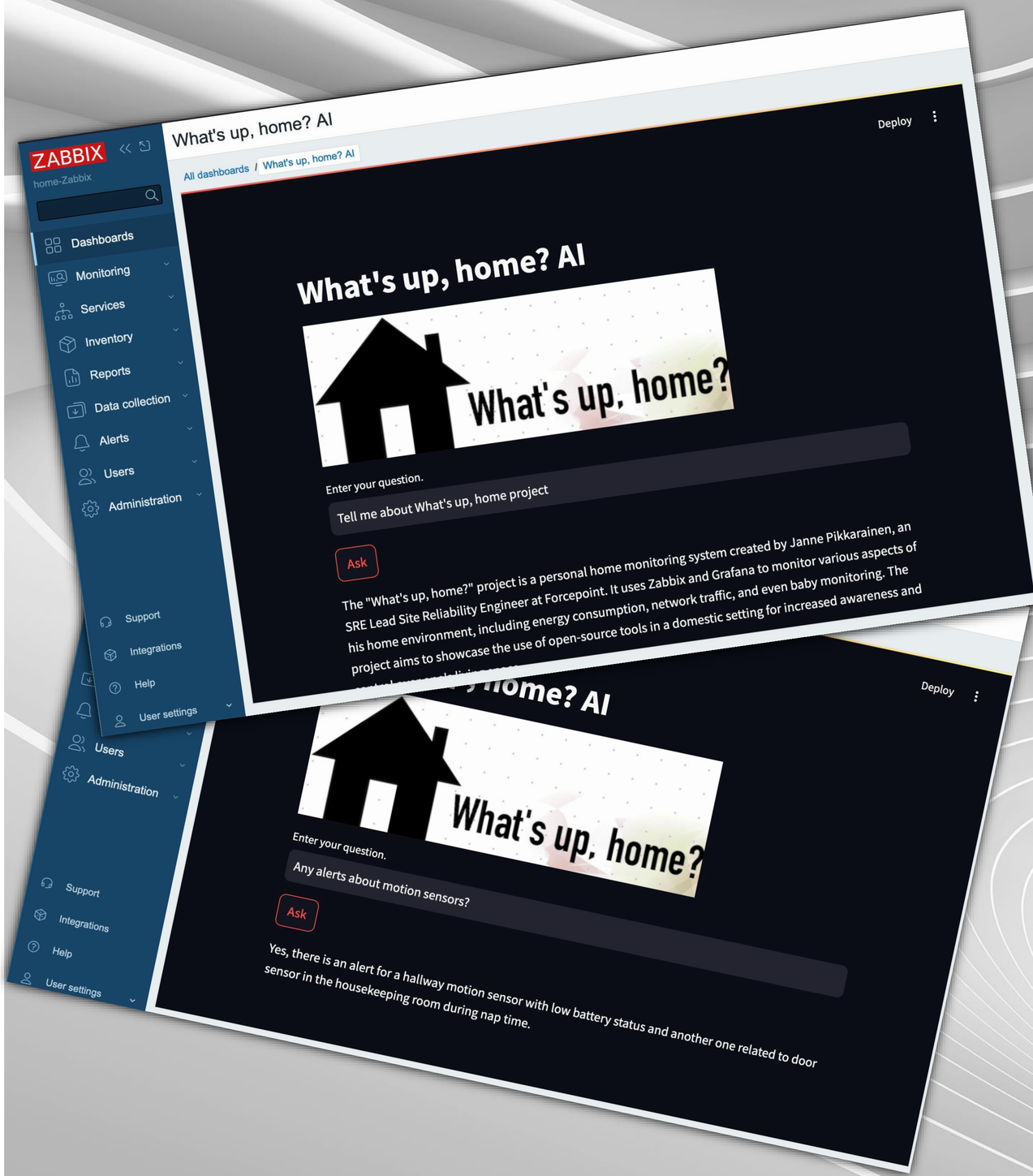
Time to blog!

Automatically, that is

My blog flow

- https://generativeai.whatsuphome.fi/ is 100% produced by GPT4All
- Python script asks Zabbix about the active alerts, and GPT4All writes a blog post about them
- Published to Drupal every morning 7am via Drupal API

Embed your Zabbix data with GPT4All



Embedding data

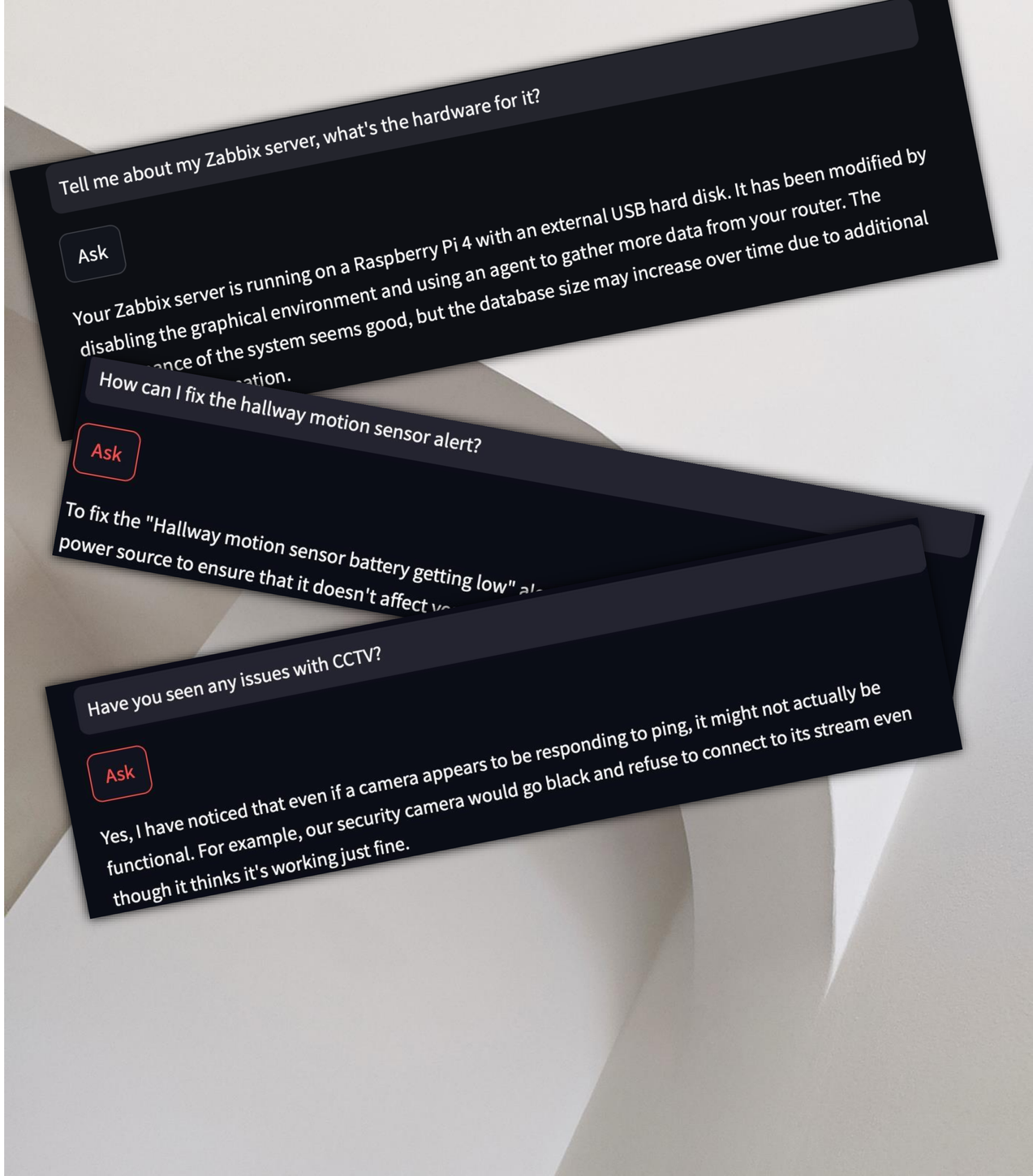
- General idea: with GPT4All, it’s easy to include custom data from local directories or remote locations
- GPT4All and Python langchain makes it easy
- Feed it Zabbix inventory, hosts, alerts, services, SLAs, playbooks, wiki…
- Create your own web UI with for example Streamlit and embed it with Zabbix URL widget or create a custom Zabbix widget

Additional data

- Zabbix user manual, other manuals you need
- In my case, my blog entries
- CVEs, Zabbix release notes
- News/change log etc about the software you care about
- This helps GPT4All to make its responses even better

Does it work?

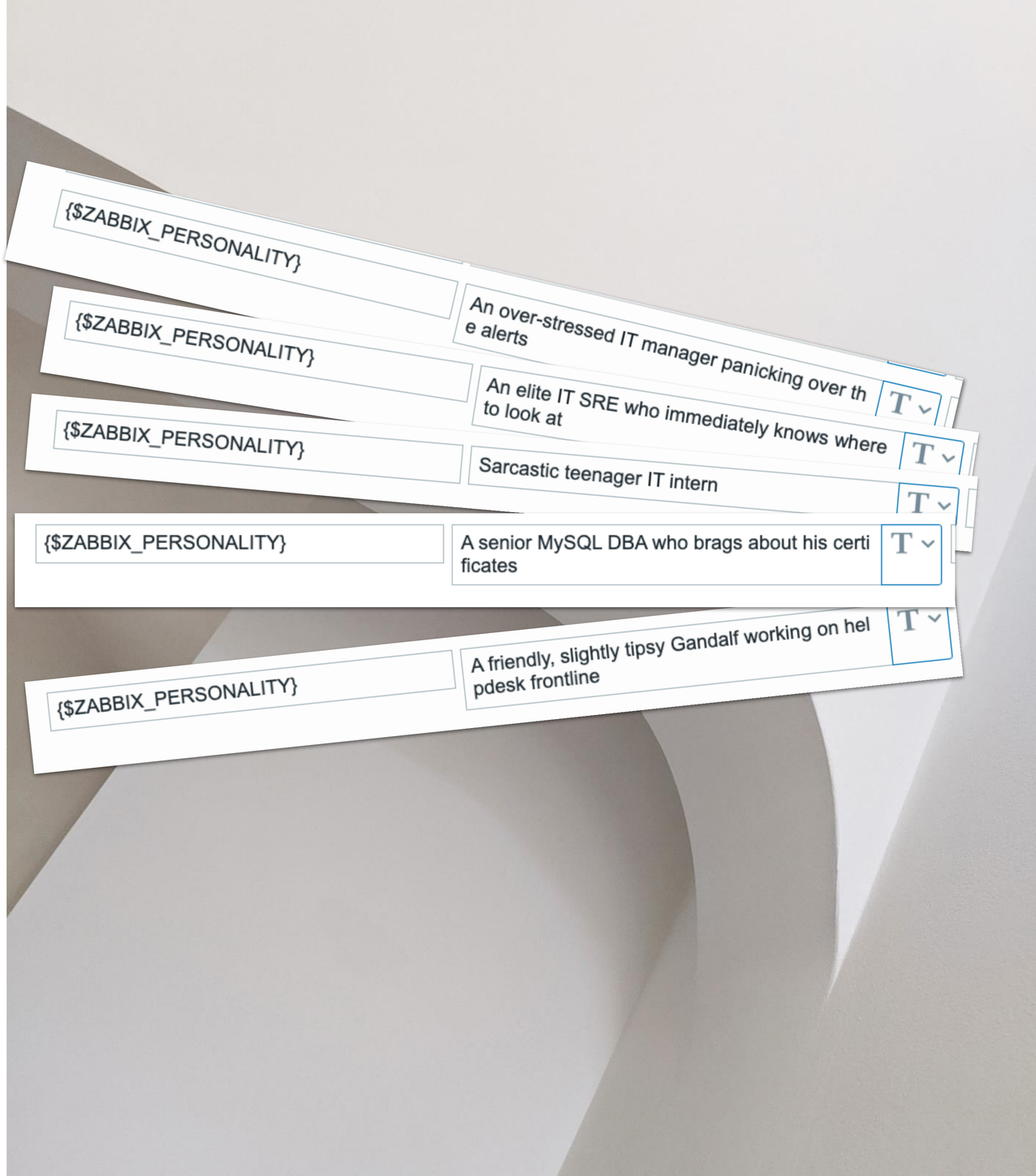
- Good for general questions about your hosts
- Good for telling if it has seen any issues with particular host or service if you feed it the past problems
- Bad with exact timelines or other time-related data
- BEWARE! As usual with GPTs, it can hallucinate



Give Zabbix a personality through GPT4All

Give Zabbix a personality

- Simple yet effective: define the personality through global Zabbix macros: {$ZABBIX_PERSONALITY}
- Use that macro as part of your alert message templates and other parts of Zabbix
- Why? Because you can. And, to make the alert contain more info about the current issue.



Use different personas

- For even deeper debugging, change the persona to some specific expert
- Add the personas to your alert context menus or just change them in the chat
- ChatGPT, GPT4All etc GUI clients provide this functionality out-of-the-box, why to use in-house Zabbix solution?
- Your own code enables Zabbix API real-time use and your own embeddings

Multiple personas

- Alternatively, give your prompt multiple personas and ask them to chat together as a virtual team
- Example on the right reads chats written to a text file by a cron job and feeds them to Zabbix item
- Result is then shown on Zabbix 7.0 item history widget
- Mostly for pure fun, but can sometimes give you new ideas why something is not working

GPT4All virtual handover call

Timestamp	GPT4All observations
2024-07-02 10:46:05 PM	**SRE:** Hey team, we've got a few Zabbix alerts to look at. Can someone take the lead on these? **DBA:** Sure thing! Let me start by looking into the server issues and database status. **Network Expert:** I'll handle the network-related alerts, such as the changing speeds of your interfaces. **DevOps Engineer:** I can take care of the Docker and Jenkins issues. **SRE:** Thanks everyone! Let's go through these alerts one by one, starting with the Zabbix server alerts: 1. Hallway motion sensor battery getting low - (Unack) 2. Outdoors lights power button off or fuse probably tripped - (Unack) 3. containerd.service: Service is not running - (Unack) **DBA:** The first two Zabbix server alerts seem to be non-critical issues related to physical infrastructure. I'll make sure they get addressed during the next maintenance window. **Network Expert:** For your interface speed changes, it could be due to network congestion or a hardware issue. Can you please check if there are any ongoing activities that might cause this?

Conclusions

Is GPT4All useful?

- A fresh approach to monitoring with completely new possibilities
- Can give you new ideas how to fix something; I say ideas, as I would consider it risky to automatically run something that's decided by AI
- With good prompting and data, can help in resolving issues
- Feels very human
- Show it to legacy monitoring software guys to impress them

Contact me

- Download the scripts from https://github.com/jannepikkarainen/whatsuphome/
- Connect with me on LinkedIn: https://linkedin.com/in/jannepik/
- Read my blog: https://whatsuphome.fi/
- Follow me on Mastodon: https://mastodon.social/@whatsuphome

THANK YOU!

Now it’s time for feedback and questions