eAssistant

Artificial Intelligence

HackYeah 2024

Team **Mniejsza Dawka Development**

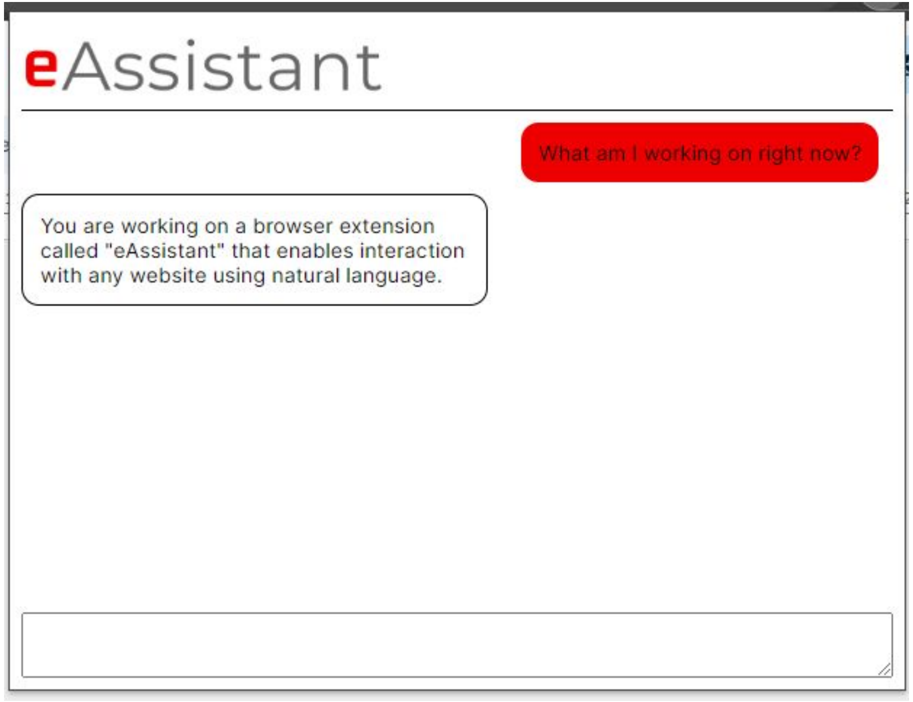
- Maciej Kaszkowiak
- Mateusz Karłowski
- Tymoteusz Jagła
- Adam Piaseczny

Background

- AI and chatbots are becoming more common*[1]* world-wide. **Despite that:**
- 1. You can’t instantly use a chatbot on any website:
  - Private websites, SPAs, are hard to integrate without APIs
  - Most of chatbots require time-consuming indexing prior to usage
- 2. You can’t tell your browser to click something for you:
  - Chatbots are typically read-only and cannot action on your behalf
- **Life would be easier if you could do both!**
  - Imagine navigating complicated gov websites with simple instructions
- We aim to tackle this problem

[1] - https://www.forbes.com/advisor/business/ai-statistics/, https://www.tidio.com/blog/chatbot-statistics/

Our solution



eAssistant

What am I working on right now?

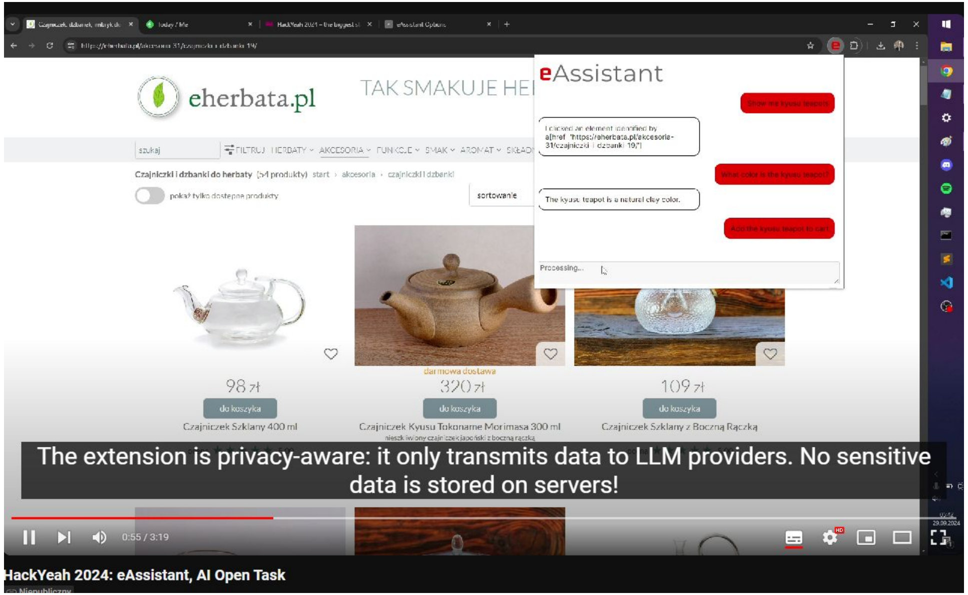
You are working on a browser extension called "eAssistant" that enables interaction with any website using natural language.

eAssistant

- A browser extension that works on any website
  - Private websites, SPAs, all of them
  - No APIs required
- You can ask questions or order actions
- Enables seamless navigation

Demo

- We have a working Chrome extension!
- You can view a short video of it in action on:
  - an online shop
  - an online RSS reader
  - the HackYeah website!
- OR you can install the Chrome extension right now!
  - see README.md for instructions


eherbata.pl
TAK SMAKUJE HE
szukaj
FILTRUJ HERBATY AKCESORIA FUNKCJE SMAK AROMAT SKŁAD
Czajniczki i dzbanki do herbaty (54 produkty) start › akcesoria › czajniczki i dzbanki
pokaż tylko dostępne produkty
sortowanie
98 zł
do koszyka
Czajniczek Szklany 400 ml
darmowa dostawa
320 zł
do koszyka
Czajniczek Kyusu Tokoname Morimasa 300 ml
109 zł
do koszyka
Czajniczek Szklany z Boczną Rączką
0:55 / 3:19

eAssistant

Show me kyusu teapots

I clicked an element identified by a[href "https://eherbata.pl/akcesoria-31/czajniczki i dzbanki 19/"]

What color is the kyusu teapot?

The kyusu teapot is a natural clay color.

Add the kyusu teapot to cart

Processing...

The extension is privacy-aware: it only transmits data to LLM providers. No sensitive data is stored on servers!

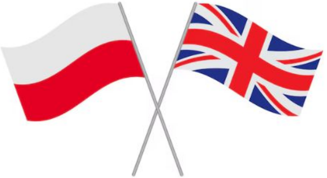
HackYeah 2024: eAssistant, AI Open Task

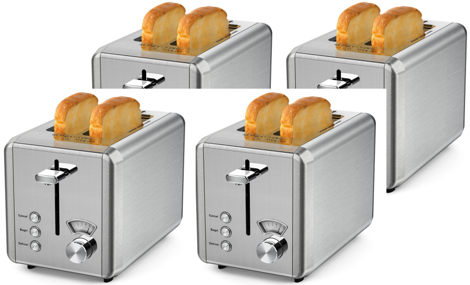
Agent behaviour - internals

- Our secret sauce is the details :)
- The agent:
  - has internal logic, which routes an user query through different prompt chains
  - can classify user’s intent into one of multiple categories
  - each category triggers different interactions between the website and the browser
  - DOM elements are identified by CSS selectors generated by LLM calls
- To improve accuracy and reliability:
  - multi-modal LLM uses screenshots and HTML to spatially identify elements in DOM
  - CoT prompting greatly improves our LLM decision making
  - JSON schemas ensure we can integrate LLM outputs with our logic

Advantages

- The extension seamlessly integrates with user's preferable browser, which doesn't disrupt their existing workflow
- **No user data collected on our servers**, as we contained logic in the client side - data is only sent to LLM
- The user could choose to manually approve sensitive actions, such as clicking, in order to prevent unwanted actions on their behalf
    - LLMs aren't perfect, and I don't want to automatically buy 10 toasters!
- Works in **Polish and English** (and many more languages)





Future improvements

We have architected a robust solution that builds upon the demo:

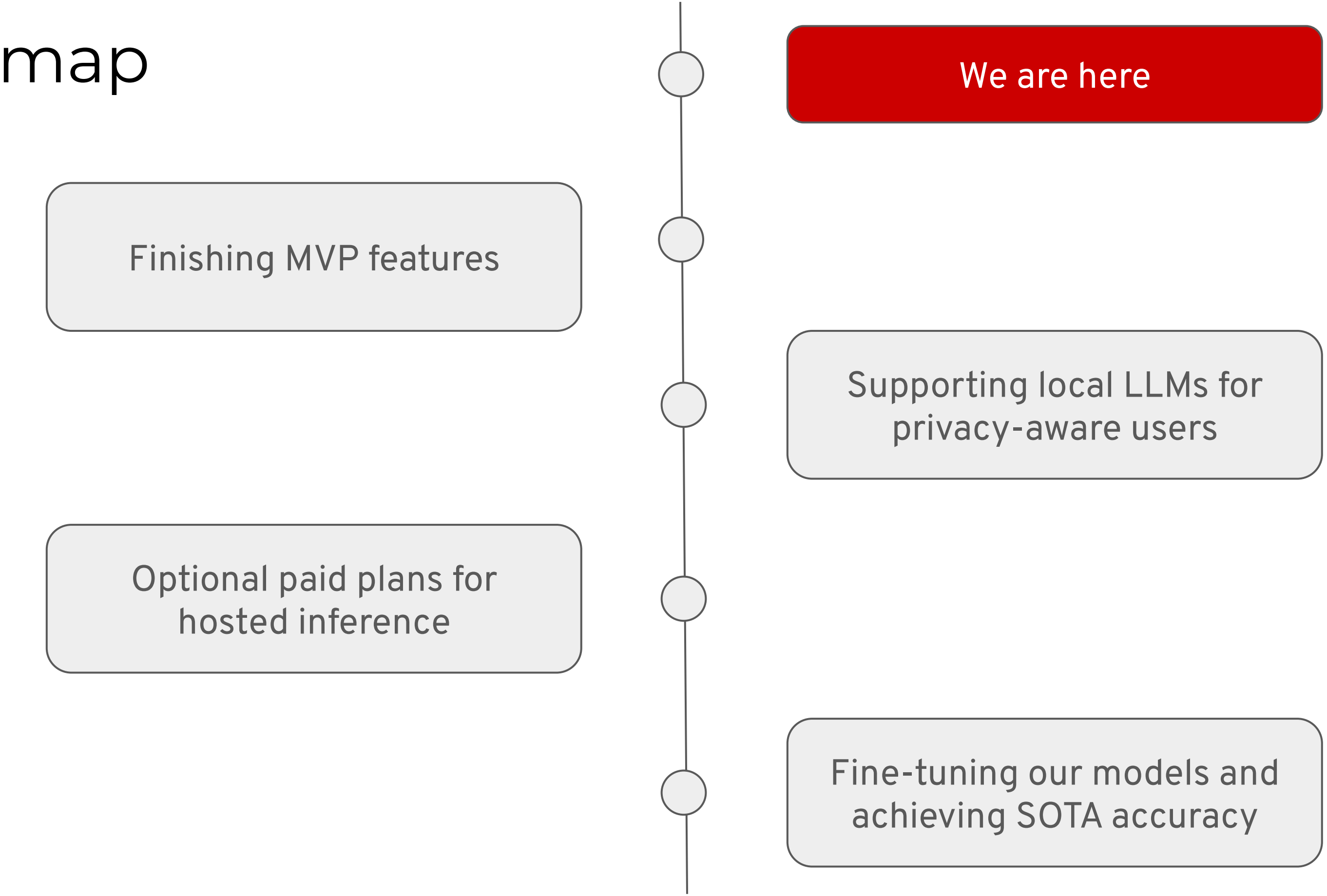
- Support for scroll and type events to allow form submission
- eAssistant can be extended for automatic navigation:
    - Allows our agent to discover complicated websites
    - Will enable actions which require **multiple actions on multiple pages**
    - On-the-fly evaluation of each path, with backtracking and pruning dead paths
- More **tight-knit integration** between Chrome and LLM
    - We could calculate elements' relative position and use it for more precise identification

Limitations

- The project utilizes LLMs, which are currently an imperfect technology
- However, their limitations become less and less of an issue with time
  - To support this statement: context length used to be a larger problem just a year ago, nowadays 128k (gpt-4o) and even 2M (Gemini 1.5) models are widely available :)
  - Our project will improve with little to no changes!
- We wish to openly acknowledge issues that can be fixed by tech progress:
  - **Costs**: repeated multimodal calls to LLM might be costly. However, inference becomes more and more cheaper, so this tool eventually become cheap for the end user;
  - **Thinking capabilities:** we might expect LLMs to become more accurate in logic-based tasks, therefore providing more actions; see recent OpenAI's o1 model [1]

[1] - https://openai.com/index/learning-to-reason-with-llms/

Roadmap



We are here

Finishing MVP features

Supporting local LLMs for privacy-aware users

Optional paid plans for hosted inference

Fine-tuning our models and achieving SOTA accuracy

Q&A

- **What LLM and why?** OpenAI's gpt-4o currently works best[1] due to multimodal capabilities, widely available API, fair pricing and Polish language support. Others can be supported, such as Llama 3.2 *(released 4 days ago!)*
- **Why offer paid plans?**
  - 1. To reach users, who can't provide an API key, or would prefer a hassle-free solution;
  - 2. To roll-out our own models, which we could later fine-tune;
  - 3. As one of potential monetization paths
- **Why allow self-hosted models?** To gain users' trust, who could be sceptical towards analysing personal data using publicly available APIs

[1] - gpt-4o also scores well on the Needle in a Needlestack benchmark, which is an improved version of the widely known NIAH benchmark: https://nian.llmonpy.ai/intro - its scores transfer to retrieving information across the whole context window, which is extremely important