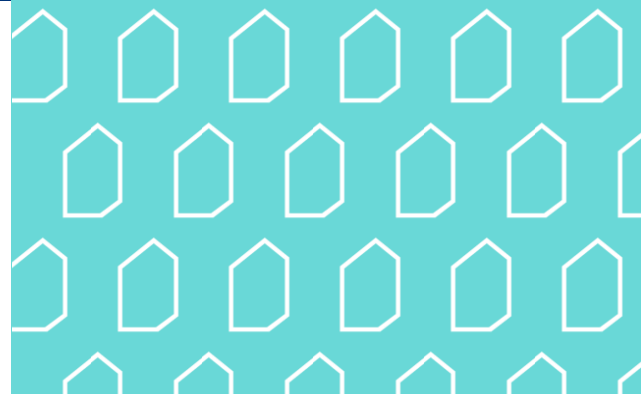


# Language Models & GEN AI In Public Administration AI – Battle!

Timo Impiö 8-2023



**DIGITAL AND  
POPULATION DATA  
SERVICES AGENCY**



# Agenda

- GPT – has it something to offer for organizations?
- AI Battle! – 3 rounds
  - Keyword extraction
  - Text summaries & styling
  - LLM integrates to your applications

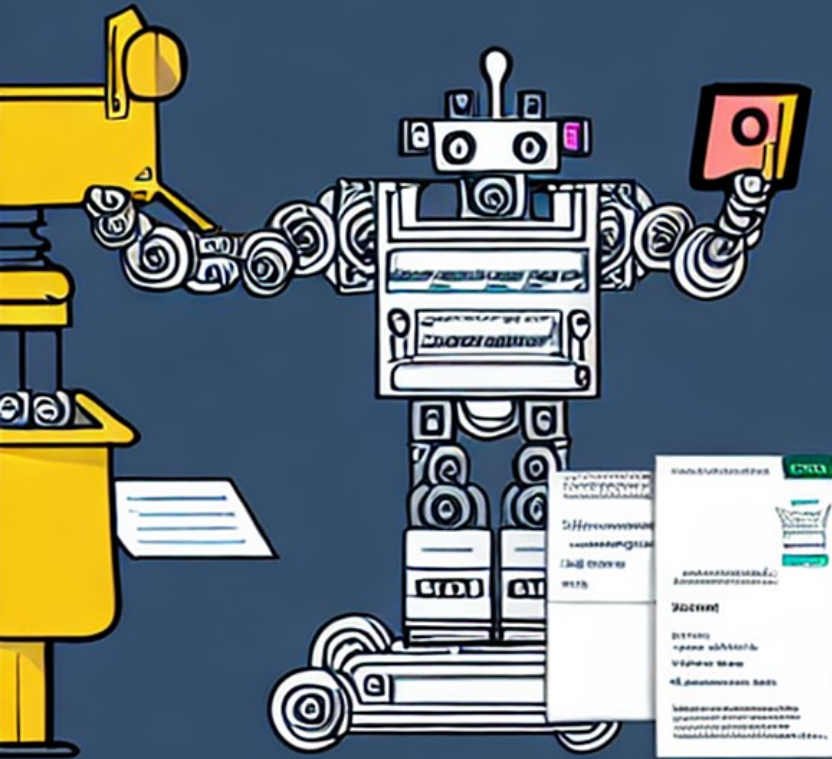
Timo Impiö

Lead Specialist – Data Analytics & Enterprise Architecture

Finland Digital and Population Data Agency



# Generative Pre-Trained Transformer (GPT)



*“design a magazine illustration of a busy GPT model as a robot that writes long text responses to user sent envelopes”*

*Image generated from text input  
(freeimagegenerator.com)*

- GPT is an advanced software product, suitable for text processing and generation
- Capable to produce new text content based on training material and "backstory"
  - Continue the story.., Make a summary of topic "ABC"
  - summaries, text classification, target styles
- Model training model requires tons of data and computing resources
  - fine-tuning models with own data is quick and cost-effective
- Value for organizational users?
  - Chatbots and Customer service assistant can be trained on customer content
  - Fine-tuning with your own data is quick and cheap (model pricing varies)
  - Fine-tuned model can be integrated into application like any other external API (new measurable value to existing apps)
  - More relevant document search (text structure and relevance)

ai\_battle

11.05.2023

**TURKU**  
**(FIN)**

**Turku University**  
**TurkuNLP: "GPT3"**  
**trained with**  
**finnish content**  
**(Markus H.)**

**VS**

**The Rest**  
**of the World**

**OpenAI: "GPT 3.5 Turbo"**  
**trained with openly**  
**accessible information**  
**(Teemu O.)**

# Round 1

## Extract keywords from text

A distinctive feature of GPT models is to produce new text, which is based on the given input.

Let's test the language model's ability to identify keywords from a single public service description.

data source: Finnish Service Catalogue for public services (FSC)



- **TurkuNLP Open GPT3:**  
Model is fine-tuned with service descriptions & keyword examples – by Markus Holopainen
- **OpenAI Ada**  
Model is fine-tuned with service description & keyword examples – by Teemu Okkonen

## Keywords, the hidden gold inside text

Service 'Playgrounds' in Finnish Service Catalogue has description:

“The purpose of the city’s playgrounds is to provide safe, diverse and stimulating play environments. The safety of the playgrounds is monitored regularly according to the playground safety standard EN 11767. This includes an annual inspection, a functional inspection and routine inspections. Regular maintenance activities include maintaining cleanliness, changing sandbox sand, adding and fluffing safety sand, and removing swing seats for the winter. The playground inspectors and maintainers have completed the playground inspector qualification. The playgrounds are intended for children and their guardians, but the visitors cannot be limited. There are 49 playgrounds in Hyvinkää. The distance from the front door to the residential area playground should be about 500 m for children and about 200 m to the block playground. The playgrounds and location map can be found on the Hyvinkää city website.”

- **TurkuNLP (GPT3)** generated keywords:

play area, playground, sports facility, play area, park, playground, outdoor park, green area

- **OpenAI ADA** generated keywords

play activity, playground activity

- **Actual manually given keywords (test set for reference):**

play activities, playground activities, playgrounds, exercise and sports





## Round 2

Summaries and  
style transformations

GPT models can summarize and  
simplify long texts.

Models can be trained with examples to  
transform text into the desired style.

Our instructions for the language model:

*“Simplify the service description so that even a  
primary school student can understand it”*



- **TurkuNLP GPT3:**  
Fine tuned with descriptions and example summaries - Markus Holopainen
- **OpenAI text-davince-003:**  
Fine tuned with descriptions and example summaries - Teemu Okkonen

**Service:** “Building Permission”, **description:** “The building supervisory authority usually requires marking the location and height of a building in connection with new construction. The need for the required measurement is indicated in the documents of the granted building permit. The exact location and height of the building are marked on the terrain before construction begins. The customer orders the marking work from the city’s technical department after the building permit application has been approved. Before the precise marking of the building site required by the authorities and after the height level has been brought to the plot, the builder can make an approximate measurement for digging the foundation and constructing so-called wire bucks. If necessary, the Technical Department of the City also performs an approximate measurement for which a separate fee based on working hours is charged. After the foundation or base of the building has been completed, the customer orders a location inspection from the Technical Department of the City if required by the building permit. The location inspection ensures that the location and height level of the building are in accordance with approved drawings. A certificate is written for the inspection and a fee according to tariff is charged.” (197 words)

## Simplified summary for primary school students:

### TurkuNLP “GPT3”:

“A building permit is required for new construction, expansion of a building or an increase in floor area counted in it as well as repair and alteration work if it may have an impact on safety or health conditions of building users. A building permit is not required for repairing a building or minor expansion of a residential building.” (59 words, relevance:OK, quality:OK, savings: -71%)

### OPENAI text-davince:

“If you want to build a new house, you must first notify the city. The city may request a measurement to verify that the house has been built correctly. If you need help, you can buy a measurement service from the city. When the house is ready, you must notify the city and get confirmation that the house has been built correctly” (62 words, clarity:great, relevance:good, savings: -69%)



# Round 3

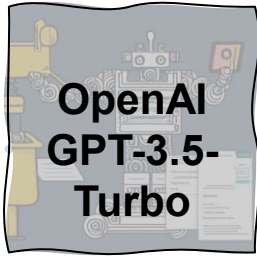
Assistant:

Search relevant local public services from external API

"Be polite and speak in finnish"

Generate a summary and show links to relevant services

Initiate a chat session :  
"Discuss politely in finnish language"



User:  
"I am looking for an apartment in Helsinki (FIN)"

recognize location

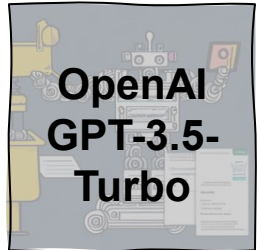


Freetext search with regional filter "Helsinki"



List of regional public services

Use LLM to review and rank services by relevance



Summary of relevant, regional, services based on user request



Taustatarina: "Keskustele ystävällisesti suomen kielellä, odota ihmisen lisävalintoja"

OpenAI  
GPT-3.5-  
Turbo

Hei! Olen tekoäly, eli AI. Tehtäväni on auttaa sinua löytämään tietoa palveluista Suomessa. Käytän apuna OpenAI-nimisen yrityksen palveluita. Muistathan, ettet kerro minulle mitään henkilökohtaista!

Hei kuinka voit?

Käyttäjä:  
"Onkohan Sotkamossa uurnalehtoa"

Paikkakunnan  
tunnistus  
syötteestä

OpenAI  
ADA

Kiitos kysymästä! Minähän olen tekoäly, joten minun toimintani ei varsinaisesti riipu omasta hyvästä voinnistani. Mutta voin varmasti auttaa sinua parhaani mukaan - mistä voisoin olla avuksi tänään?

Onkohan Sotkamossa uurnalehtoa?

AuroraAI-tekstihaku  
tunnistetulla aluerajauksella  
"Sotkamo"

AuroraAI  
fasttext-  
kielimalli

Voin etsiä tiedon, jos Sotkamosta löytyy uurnalehtoa. Mutta ensin minun täytyy tarkentaa hieman: tarkoittko uurnalehdolla hautausmaata, joka on tarkoitettu vain uurnien hautaamiseen, vai uurnalehtoa, joka on urneille tarkoitettu lokerikko hautausmaalla?

Etsi Sotkamon uurnalehto.

Palvelujen lista

Valitse käyttäjälle sopivimmat  
palvelut

OpenAI  
GPT-3.5-  
Turbo

Kaikki kuvaukset eivät vastaa kysymykseen uurnalehdosta Sotkamossa, mutta yksi vastaus on näistä mahdollinen: Huovishuone esittelee kirjailija Veikko Huovisen työ- ja elinympäristöä, tuotantoa ja niiden taustoja. Sijainti: sotkamo Huovishuone on kuitenkin museo, eikä uurnalehto. Käyttämäni palvelutietovaranto oli digi- ja väestötietoviraston (DVV) palvelutietovaranto, josta ei valitettavasti löytynyt tietoa Sotkamon uurnalehdosta. Suosittelen tarkistamaan paikallisten hautaus toimistojen ja Sotkamon kunnan hautaus toimiston sivuston lisätietojen saamiseksi.

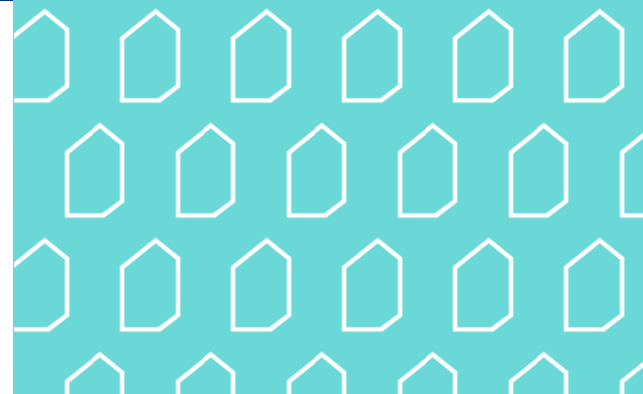
Kirjoita tähän

Tiivistelmä alueen  
sopivimmista palveluista

# Assistants & chatbots with organization defined content



**DIGITAL AND  
POPULATION DATA  
SERVICES AGENCY**



# Finland Digital and Population Data Agency: Text analytics and Chatbots

**3X10D Service recommendation API**

STAT, Cos-similarity

**Freetext search for public service information**

LM, FASTTEXT

**Re-ranking search results by earlier recommendations**

ML, XGBOOST

**Digiturva Chat**

Langchain + LLM

**www.DVV.fi content assistant (CHATBOT)**

Langchain + LLM

**www.suomi.fi Content assistant (CHATBOT)**

Langchain + LLM

**DVV Process Guide assistant (Search by NLP)**

LLM

**TurkuNLP Sentence Bert vs. Facebook Fasttext POC for text search**

Open Source vs. Commercial LLM

**AI-Battle Commercial cloud-LLM vs in-house-driven Open Source LLM**

Open Source vs. Commercial LLM

**National Cyber Crime Center Topic modelling on websites by content**

Python NLP, clustering, visual



## Digiturva-chat

Kysy digiturvasta

Vastauksia perustuen 80 julkiseen asiakirjaan

**Sinä:** Mikä on digiturva?

**Avustaja:** Digiturva tarkoittaa digitaalisen toimintaympäristön turvallisuutta, jota usein kutsutaan myös kyberturvallisuudeksi. Se on jo nyt erottamaton osa turvallisuutta ja tulee olemaan yhä keskeisemmin tulevaisuudessa. Digiturvallisuus sisältää viisi eri osa- aluetta, joita ovat tiedon suojaaminen, tietoturvallisuus, tietosuoja, tietoturvallisuus ja tietoturvallisuus.

**Sinä:** Mikä on organisaation johdon vastuu digiturvan osalta?

**Avustaja:** Organisaation johdon vastuu digiturvasta on huolehtia riittävästä henkilöstön tietoturvasosaamisesta ja - tietoisuudesta, toimia esimerkkinä mahdollistaen turvallisuuteen liittyvän asenteen ja kulttuurin syntymisen ja kehittymisen, sekä olla vastuutettu ja nimetty keskeiset digiturvallisuuden roolit, esimerkiksi tietoturvasosaamisen ja tietosuojan osalta.

Kirjoita kysymys tähän

Lähetä

## Digiturva Chat

Tiedonhallintalautakunnan kokoama digiturvallisuuden dokumenttikokoelma chat-palveluna.

Rakennettiin avustaja jolta voi kysyä tietoturvasta –vastaukset syntyvät julkisten digiturvaohjeiden pohjalta.

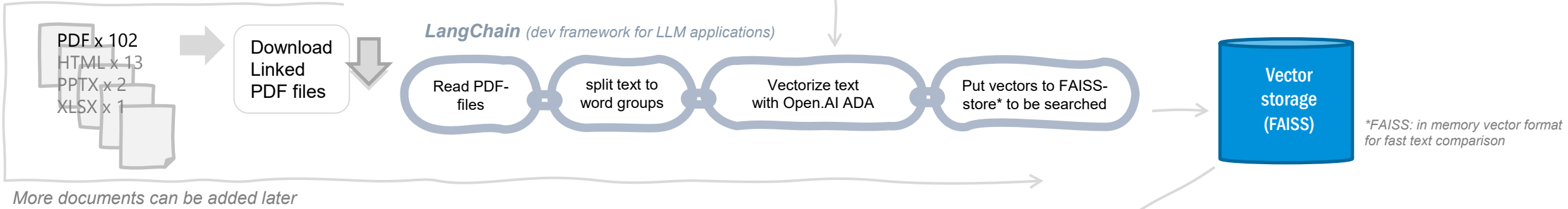
Käyttäjärühmänä suomenkieliset kansalaiset

**Sinä:** Mitkä ovat viisi tärkeintä asiaa henkilöstön huolehtia tietoturvan osalta?

**Avustaja:** Henkilöstön tietoturvasosaamisen viisi tärkeintä osa- aluetta ovat kulunvalvonta, tilojen lukitukset, kassakaapit, tietoaineiston käyttäjän tietoturvatietoisuuden ja osaamisen kehittäminen sekä käyttöoikeuksien hallintaprosessi.



Source material: A collection of digital security documents compiled by the Finland Information Management Board

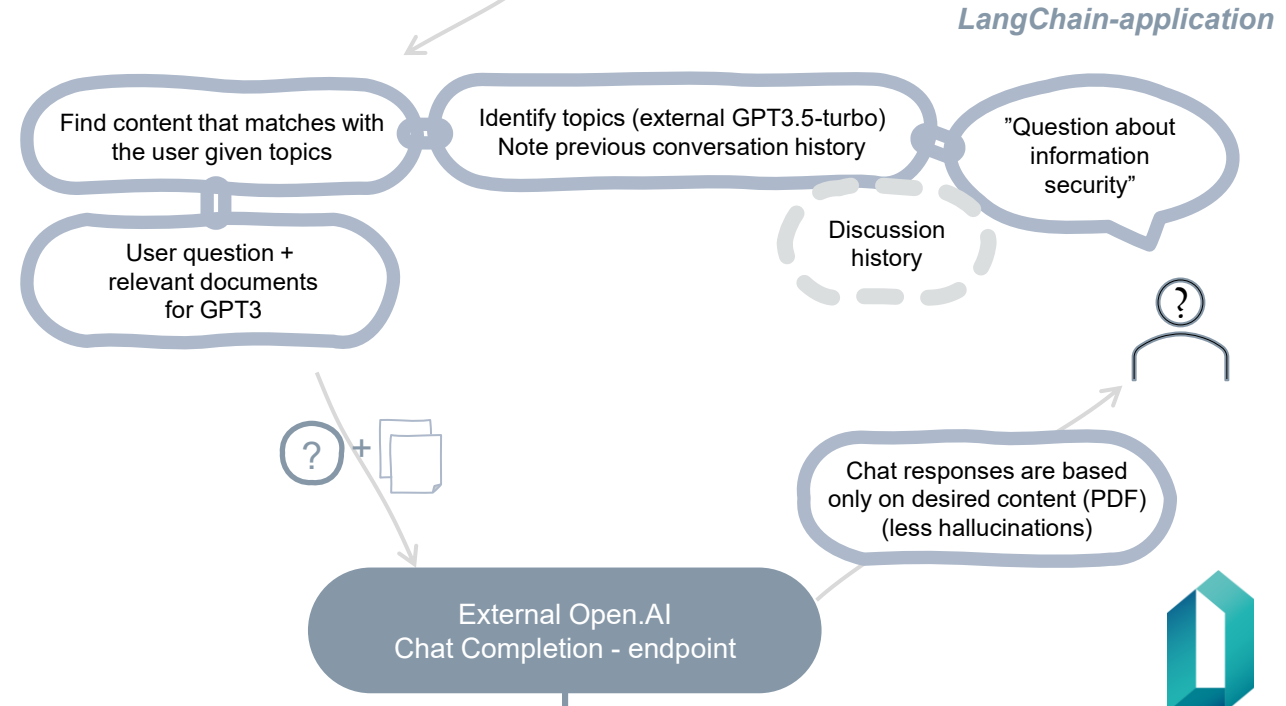


## Example question & response: "chat with documents"

**Sinä:** Mikä on organisaation johdon vastuu digiturvan osalta?

**Avustaja:** Organisaation johdon vastuu digiturvasta on huolehtia riittävästä henkilöstön tietoturvallisuusosaamisesta ja -tietoisuudesta, toimia esimerkkinä mahdollistaen turvallisuuteen liittyvän asenteen ja kulttuurin syntymisen ja kehittymisen, sekä olla vastuutettu ja nimetty keskeiset digiturvallisuuden roolit, esimerkiksi tietoturvallisuuden ja tietosuojan osalta.

Kirjoita kysymys tähän Lähetä





**Sinä:** Mikä on organisaation johdon vastuu digiturvan osalta?

**Avustaja:** Organisaation johdon vastuu digiturvasta on huolehtia riittävästä henkilöstön tietoturvasosaamisesta ja -tietoisuudesta, toimia esimerkkinä mahdollistaen turvallisuuteen liittyvän asenteen ja kulttuurin syntymisen ja kehittymisen, sekä olla vastuutettu ja nimetty keskeiset digiturvallisuuden roolit, esimerkiksi tietoturvasuojan ja tietosuojan osalta.

Kirjoita kysymys tähän

Lähetä



Videokaappaus [digiturvachatin](#) koekäyttöliittymästä

[Digiturva Chat \(dvv.fi\)](#)



# Suomi.fi website assistant

## Assistant application and user group:

An assistant application where you can ask what the suomi.fi website knows about specific question,  
The user group of the trial is DVV's customer service experts

## Material:

Content of the suomi.fi website (all services, instructions and guides in text format), (Languages: FI EN, SE)

## Data collection: Machine reading with a "crawler robot".

Content can be renewed at desired intervals to ensure content is up-to-date

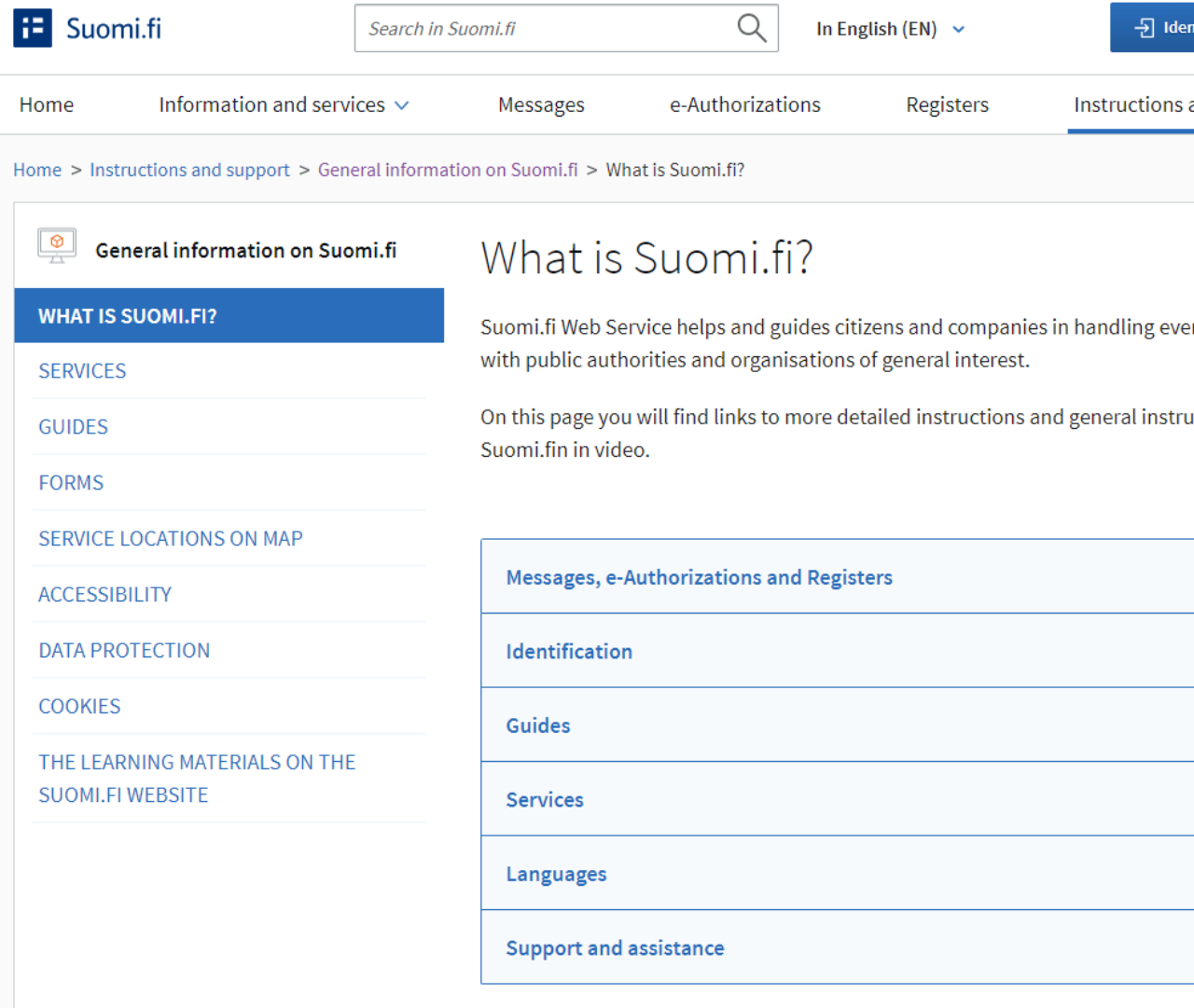
## Data processing:

Html pages are automatically collected from the site using the BeautifulSoup and Selenium libraries. The collected text documents have been converted into raw text documents with NLP libraries and vectorized into numerical format using Huggingface's open model. These have been exported to the language model using langchain's FAISS indexing, based on which the most suitable pages for the user's search query can be found using similarity comparison.

## To use the application:

Content search works as a semantic search, i.e. the entered question is treated as a complete sentence. The sentence is matched with the content of the suomi.fi website. So the search doesn't work like a single word or \*wildcard search.

The application can be used, for example, in a customer service situation or in any application where information about the content of websites is desired.

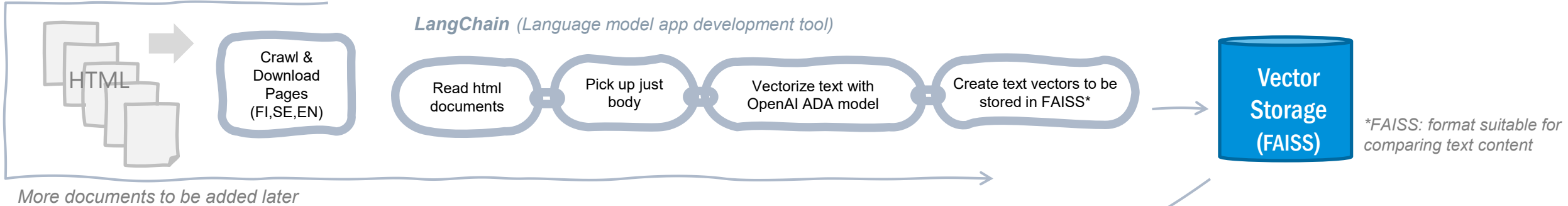


The screenshot shows the Suomi.fi website interface. At the top, there is a search bar with the text "Search in Suomi.fi" and a magnifying glass icon. To the right of the search bar, it says "In English (EN)" with a dropdown arrow. Below the search bar, there is a navigation menu with links for "Home", "Information and services", "Messages", "e-Authorizations", "Registers", and "Instructions and support". The "Instructions and support" link is highlighted. Below the navigation menu, there is a breadcrumb trail: "Home > Instructions and support > General information on Suomi.fi > What is Suomi.fi?". The main content area is titled "General information on Suomi.fi" and features a blue header "WHAT IS SUOMI.FI?". Below this header, there is a list of links: "SERVICES", "GUIDES", "FORMS", "SERVICE LOCATIONS ON MAP", "ACCESSIBILITY", "DATA PROTECTION", "COOKIES", and "THE LEARNING MATERIALS ON THE SUOMI.FI WEBSITE". To the right of this list, there is a section titled "What is Suomi.fi?" with the text: "Suomi.fi Web Service helps and guides citizens and companies in handling even... with public authorities and organisations of general interest." Below this text, there is a paragraph: "On this page you will find links to more detailed instructions and general instru... Suomi.fi in video." At the bottom of the page, there is a vertical list of links: "Messages, e-Authorizations and Registers", "Identification", "Guides", "Services", "Languages", and "Support and assistance".



# Suomi.fi website content assistant

Reference: [www.suomi.fi](http://www.suomi.fi) web pages (~200 000)



## Example question and answer:

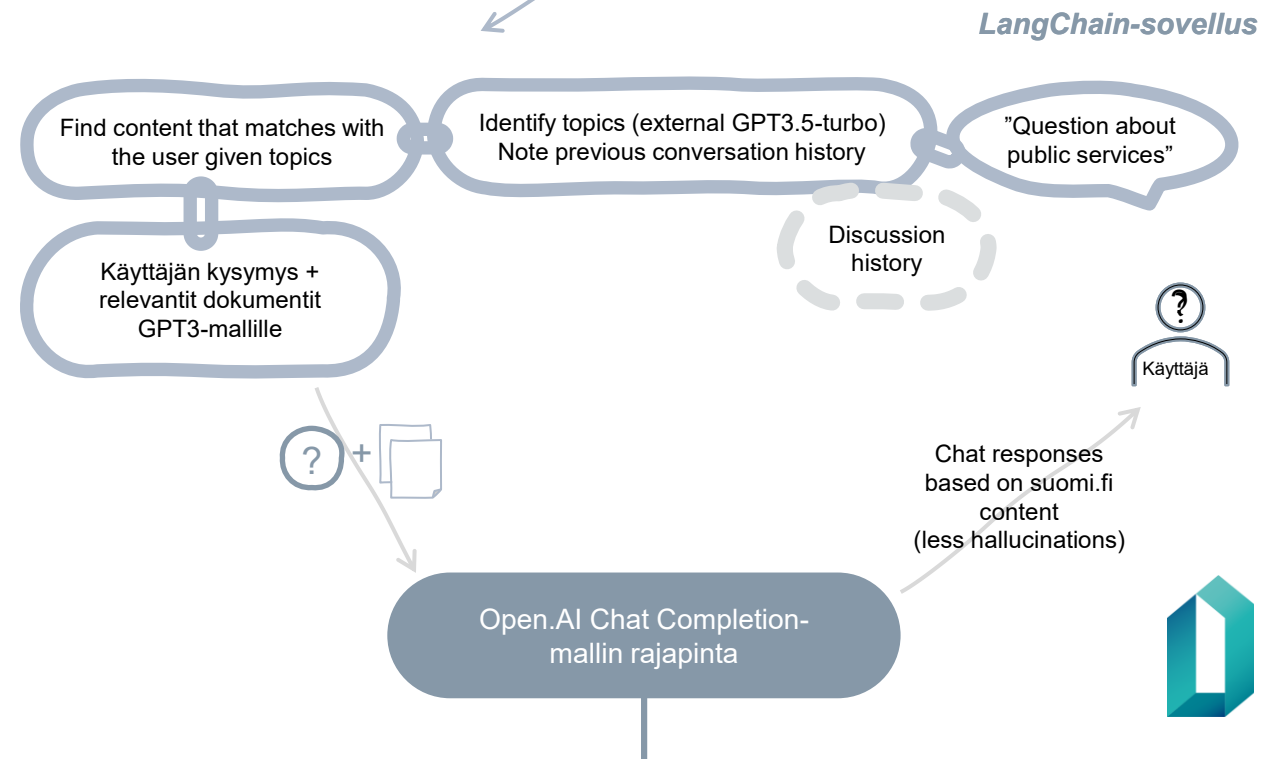
**Chat**  
Tietoa suomi.fi:n palveluista suomeksi, ruotsiksi ja englanniksi

**Sinä:** How to find a library in helsinki

**Avustaja:** To find a library in Helsinki, you can visit the following public libraries: Arabianranta Library, Jätkäsaari Library, Herttoniemi Library, Helsinki Central Library Oodi, Kallio Library, Rikhardinkatu Library, Puistola Library, Pitäjänmäki Library, Kannelmäki Library, Paloheinä Library, Laajasalo Library, Vallila Library, Tapulikaupunki Library, Oulunkylä Library, and Kontula Library.

**Lähteet:**  
[Library Services - Suomi.fi](#)  
[Search - Suomi.fi](#)  
[Search - Suomi.fi](#)  
[Search - Suomi.fi](#)

Kirjoita kysymys tähän



## Chat

Tietoa suomi.fi:n palveluista suomeksi, ruotsiksi ja englanniksi

**Sinä:** How to find a library in helsinki

**Avustaja:** To find a library in Helsinki, you can visit the following public libraries: Arabianranta Library, Jätkäsaari Library, Herttoniemi Library, Helsinki Central Library Oodi, Kallio Library, Rikhardinkatu Library, Puistola Library, Pitäjänmäki Library, Kannelmäki Library, Paloheinä Library, Laajasalo Library, Vallila Library, Tapulikaupunki Library, Oulunkylä Library, and Kontula Library.

Lähteet:

[Library Services - Suomi.fi](#)

[Search - Suomi.fi](#)

[Search - Suomi.fi](#)

[Search - Suomi.fi](#)

Kirjoita kysymys tähän

Lähetä

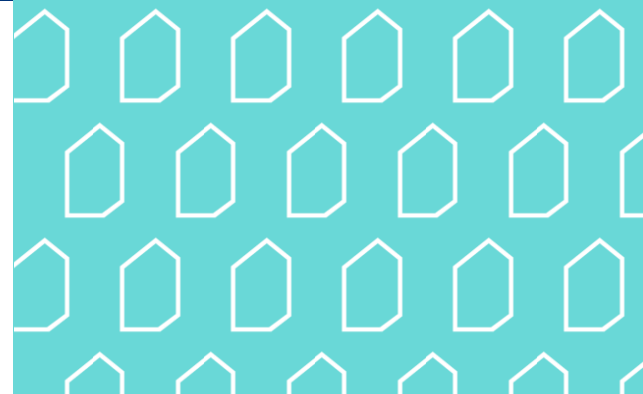
[Suomi.fi-chat](#)



# Bots and assistants



**DIGITAL AND  
POPULATION DATA  
SERVICES AGENCY**



# Botit ja avustajakokeilut

Kansalais-  
neuvonnan  
chatbot (2019)

CHATBOT

VTJ-otteet ja  
sukuselvitykset  
Boost.AI

CHATBOT

HOLLE Laskujen  
kirjausrobotti

RPA

HOLLE  
Asiakasbotti?

Ohjepankki QA-  
avustaja

LLM

DVV.fi  
hakuavustaja

LLM

Suomi.fi  
hakuavustaja

LLM

 Digiturva Chat

Langchain + LLM

Vaativuusmäärittely

Mitkä palvelut?  
Mitkä sisällöt?  
Mitkä kielet?

ASPA-avustaja

2020

2021

2022

2023

2024

