

# Susumu Okazawa

📍 Stockholm, Sweden    ✉ susumu.okazawa@leastaction-ab.com    🔗 susumu-okazawa-b954421b5

## Experience

<b>Least Action AB</b> , CEO	Sweden
<ul style="list-style-type: none"> <li>Founded a startup to develop LLM powered legal assistant to support patent attorneys</li> <li>Open to working with clients from any industry who want to develop AI applications</li> </ul>	Oct 2025 – present
<b>Savantic AB</b> , Senior Data Scientist	Sweden
<ul style="list-style-type: none"> <li>Recommended as the most efficient developer that can deliver valuable machine learning solutions despite the short project duration</li> <li>Experienced with projects of all sizes and all ML phases</li> <li>Highly regarded not only for the coding skills, but also for the ability to prioritize tasks</li> </ul>	Aug 2020 – present
<b>TMI Associates</b> , Patent Attorney	Japan
<ul style="list-style-type: none"> <li>Expanded business of patent prosecution by 200% in the field of machine learning and quantum computer</li> <li>Presented patent strategy in the field of machine learning to pharmaceutical company and IT startups</li> <li>Supervised 3 junior attorneys</li> </ul>	Oct 2016 – present
<b>HARUKA Patent &amp; Trademark Attorneys</b> , Patent Engineer	Japan
<ul style="list-style-type: none"> <li>Documented more than 100 patent specifications of software, semiconductor device, etc.</li> <li>Assisted clients with obtaining more than 100 patents in Japan, the US, Europe, China, and Korea</li> </ul>	Apr 2013 – Oct 2016
<b>Japan Society for the Promotion of Science</b> , Researcher	Japan
	Apr 2012 – Mar 2013

## Education

<b>PhD</b> <b>The Graduate University for Advanced Studies</b> , Physics	Japan
<ul style="list-style-type: none"> <li>Thesis: Non-equilibrium aspects of the black hole thermodynamics</li> <li>Scholarships: Research Fellowship for Young Scientists</li> </ul>	Apr 2008 – Mar 2013
<b>BS</b> <b>Tokyo Institute of Technology</b> , Physics	Japan
	Apr 2004 – Mar 2008

## Skills

**Programming:** Python, OpenAI, Transformers, PyTorch, FastAPI, Qdrant, Streamlit, Gradio, Apache Spark, Databricks, SQL, BigQuery, Kafka, Elasticsearch, Google Analytics, Power BI, Docker, Git, VS Code, Cursor, CI/CD, Agile, Scrum

**Cloud Platforms:** Microsoft Azure, Google Cloud, Amazon Web Services

**Languages:** English (fluent), Japanese (native)

## Certifications

<b>Certified ScrumMaster (CSM)</b>	June 2024
<b>Databricks - Data Science Professional</b>	June 2021
<b>Databricks - Delta Lake Essentials</b>	June 2021
<b>Japanese Patent Attorney</b>	Apr 2016

## Publications

---

### **Workload assessment: Time to emanate from accurate conclusions instead of preconceived notions**

2023

Dan Hasson, Susumu Okazawa, Karin Villaume

[10.1111/joop.12436](https://doi.org/10.1111/joop.12436) [↗](#)

### **Natural language processing as work support in project tendering**

2022

Linda Cusumano, Rasmus Rempling, Robert Jockwer, Ricardo Alencar Saraiva, Mats Granath, Nilla Olsson, Susumu Okazawa

[10.1201/9781003348443-258](https://doi.org/10.1201/9781003348443-258) [↗](#)

## Extracurricular Activities

---

- 2024-05 - present: A board member of Japanese School Association in Stockholm

## Projects

---

### **LLM Agents for Geodata**

Oct 2025 – present

Gave advice on developing LLM agents to interact with geospatial data through chat interface.

- Shared best practices of developing LLM agents
- Helped to design the architecture of the agentic system interacting with PostGIS database
- **Azure OpenAI, SQL, Model Context Protocol (MCP), Docker**
- *LLM agents, Agile*

### **LLM Agents Development**

May 2025 – present

Led the team to develop a backend for an agentic system integrated in the existing client's application.

- Delivered strong results in 2 days hackathon as a starting point
- Designed the architecture of the agentic system using MCP server
- **Azure OpenAI, FastAPI, OpenAI Agents SDK, MCP, Docker**
- *LLM agents, Agile, Hackathon*

### **Agentic Application Development**

Jan 2025 – Aug 2025

Developed an agentic application to interact with private data through chat.

- Developed agents system on top of an open source chat interface
- Led the project to deliver a progress in fast iterations
- **Azure OpenAI, Open Web UI, FastAPI, PydanticAI, BigQuery, Docker**
- *LLM agents, Agile*

### **LLM-based Document Generator**

Jan 2025 – Mar 2025

Mentored a colleague to build a proof of concept application using ChatGPT for document generation.

- Built a RAG system to generate a document based on the past documents
- Led the project to deliver the proof of concept application in 2 months
- **Azure OpenAI, Gradio, Chroma, Docker, Microsoft Azure**
- *Retrieval Augmented Generation (RAG), LLM agents, Agile*

### **Water Leakage Prediction**

Oct 2024 – May 2025

Mentored a developer to get started with machine learning to predict water leakages in the water distribution network.

- Mentored a developer to create a machine learning model to predict water leakages in the water distribution network
- Predicted leak time and location based on the given network topology and the pressure sensor data
- **scikit-learn, SciPy, NumPy, plotly**
- *Exploratory data analysis, Data visualization*

### **Fruit Quality Assessment Device**

Sept 2024 – Feb 2025

Led a team to develop hardware, GUI, and ML models for non-destructive fruit quality assessment.

- Facilitated communication to achieve the fastest and the best quality of delivery
- **Raspberry Pi, Sensors, 3D printing, Python**
- *IoT, Infrared spectroscopy, Agile, Scrum*

### **RAG-based Chatbot**

Sept 2024 – Nov 2024

Led a team to develop a chatbot prototype that generates answers based on internal documents.

- Mentoring new hires to get up and running as quickly as possible
- Created a CI/CD pipeline to automate container deployment to Azure
- **Azure OpenAI, Azure AI Search, Azure Document Intelligence, Azure App Service, Gradio, GitLab CI/CD, Docker**
- *Retrieval Augmented Generation (RAG), Agile, Scrum, Code review*

### **LLM-based Consultant Profile Generation**

Feb 2024 – Apr 2024

Supervised a master's thesis on generating consultant profiles using LLMs such as Mistral.

- Gave advice on open source vector databases and LLMs
- Supported setting up infrastructures such as Azure OpenAI and RunPod (serverless GPU)
- **Azure OpenAI, Qdrant, Mistral, Serverless GPU**
- *Retrieval Augmented Generation (RAG), Prompt engineering*

### **Data Pipeline Modernization**

May 2023 – June 2024

Developed and modernized data pipelines using Azure Databricks and CI/CD tools.

- Introduced several modern features of Databricks such as Unity Catalog, Delta Live Tables and Databricks Asset Bundles
- Developed DAX code for Power BI to process the data for dashboarding
- **Apache Spark, PySpark, SparkSQL, Databricks, Unity Catalog, Delta Live Tables, Databricks Asset Bundles, Github Actions, Power BI, DAX**
- *Data modeling, CI/CD, Agile, Scrum*

### **ChatGPT Document Generator**

Apr 2023 – May 2023

Built a proof of concept application using ChatGPT for document generation and vector database for PDF embeddings.

- Built a vector database to store text embedding vectors extracted from PDF documents
- Developed the prompt for ChatGPT to include the relevant past document in the context
- **ChatGPT, Qdrant, Streamlit, Docker, Microsoft Azure**
- *Prompt engineering, Agile*

### **ChatGPT Workshop**

May 2023 – May 2023

Delivered a one-day workshop on using ChatGPT with vector search, including Python and F# code samples.

- Prepared sample code in both Python and F# because the client mainly works on .NET
- **ChatGPT, .NET, F#**
- *Workshop*

### **Streaming Data Processing Research**

Dec 2022 – Dec 2022

Explored streaming data processing using Apache Kafka for an internal hackathon.

- Prepared a stack of Confluent Kafka on an internal server for an internal hackathon
- Prepared an IoT device as a data source as well as a sample MQTT consumer
- **Apache Kafka, MQTT, Docker**
- *IoT, Hackathon*

### **LIA Internship Supervision**

Sept 2022 – Nov 2022

Supervised a LIA (Lärande i Arbete) internship focused on practical machine learning and data analysis.

- Guided the interns on how to perform exploratory data analysis on the real data
- Provided advice on how to present findings to a broader audience
- **pandas, scikit-learn, plotly**
- *Exploratory data analysis, Code review*

### **Machine Translation API and ML Pipelines**

Sept 2022 – Apr 2023

Developed a machine translation API, ML pipelines, and a dashboard for ChatGPT testing.

- Built the machine learning pipeline using Amazon SageMaker
- Built the translation API using FastAPI and Amazon Lambda
- **FastAPI, AWS SageMaker, AWS Lambda, Amazon ECR, Docker**
- *Machine translation, MLOps, Agile*

### Mobile Device Location Estimation

Nov 2021 – June 2022

Developed ML models and algorithms for mobile device location estimation and statistics aggregation.

- Developed new machine learning models to estimate the location of mobile device based only on the ping time series
- Developed new gridding algorithms which are used in the aggregation of statistics
- Applied geospatial functions in BigQuery, taking care of the amount of data, which is a few hundred GB per day per country
- **BigQuery, OpenStreetMap, GeoPandas, NumPy, Shapely, NumPyro, PyStan, Google Cloud, Docker**
- *Exploratory data analysis, Geospatial analysis, Bayesian hierarchical model, Agile*

### Physics Parameter Optimization

Feb 2022 – Apr 2022

Supported strategy and used Optuna for optimal physics parameter search in fluid property models.

- Suggested using Optuna to find the optimal physics parameters in the model which estimate the fluid properties flowing through pipe
- **Optuna, scikit-learn, SciPy, NumPy, pandas**
- *Exploratory data analysis, Hyper parameter tuning*

### Travel Mode Detection Model

Sept 2021 – May 2022

Developed an unsupervised model to estimate public transport share and travel time ratios, with dashboard presentation.

- Presented the results in a dashboard developed by Plotly Dash
- Developed an unsupervised model to estimate public transport share and travel time ratios
- **scikit-learn, SciPy, NumPy, pandas, Plotly, Dash, Docker**
- *Exploratory data analysis, Data visualization*

### Spare Parts Return Analysis

Nov 2021 – Dec 2021

Analyzed returns of spare parts and created SQL queries for data aggregation and visualization.

- Created SQL queries which aggregate data from a bunch of tables
- Visualized the results to grasp an overview of the returns
- **Microsoft SQL Server, pandas, Plotly**
- *Exploratory data analysis, Agile*

### GPT-2 Chatbot for Domestic Abuse Support

Sept 2021 – Dec 2021

Designed and trained a GPT-2 chatbot for domestic abuse support, deployed as a web API.

- Designed the training data for a GPT-2 based chatbot which consults people suffering in domestic abuse
- Trained the GPT-2 based chatbot on a GPU in Azure
- Deployed the model in Azure Functions as a web API, which can be invoked from the mobile app
- **pandas, GPT-2, Plotly, Streamlit, Docker, Microsoft Azure**
- *NLP, Training on GPU, Agile*

### Procurement Document Analysis Tools

May 2021 – June 2021

Developed text analysis and visualization tools for procurement documents, published a paper with the client.

- Developed text analysis and visualization tools using Elasticsearch, Kibana, and Streamlit for procurement documents in PDF
- Deployed the applications as containerized web applications on Azure
- Published [a paper in proceeding](#) with the client
- **Elasticsearch, Kibana, Streamlit, Docker, Word2Vec, Plotly, Microsoft Azure**
- *NLP, Agile*

### What-if Tool for Public Transport

Mar 2021 – Apr 2021

Built a prototype what-if tool for analyzing driver and target values, with Bayesian modeling.

- Built a prototype of a what-if tool which analyzes relationship between various driver values, such as punctuality, and target values, such as customer satisfaction index
- Developed the dashboard in one week, including the requested feature and two novel features
- Developed Bayesian hierarchical models to estimate relationship between the drivers and the targets for each user group
- **NumPy, pandas, scikit-learn, NumPyro, GPy, Plotly, Dash, Voilà, Heroku**
- *Exploratory data analysis, Data visualization, Bayesian hierarchical model*

### **Stress Management Data Analysis**

Jan 2021 – Feb 2021

Analyzed stress management app data, identified personality clusters, and published a peer-reviewed paper.

- Identified clusters of personality traits that have strong confounding relationships with stress related variables
- Published [a peer reviewed paper](#) with the client
- **NumPy, pandas, scikit-learn, tslearn, GPy, Matplotlib, Plotly**
- *Statistical test, Time series analysis*

### **Customer Journey Analysis**

Nov 2020 – Dec 2020

Analyzed customer journey factors using NLP and Bayesian models.

- Used Natural Language Processing (NLP), which involves text classification, topic modelling, and feature extraction
- Gained insights from time series analysis using Bayesian probabilistic models
- **Gensim, PyTorch, NumPyro, Matplotlib**
- *NLP, Exploratory data analysis, Bayesian probabilistic model*

### **Swedish Question Answering System**

Oct 2020 – Nov 2020

Developed a closed domain QA system based on BERT for English and Swedish, published dataset and model.

- Published [the Swedish SQuAD 2.0](#) dataset and [the fine-tuned model](#)
- Established better scores on question answering task compared to the publicly available model
- **Transformers, Google Cloud Translation, Tensorflow, PyTorch**
- *Training on GPU*