

AI Research Engineer | Multimodal GenAI & MLOps

Summary

Takehiro Tsurumi

website: <https://tti440.github.io/>

Arnhem, Netherland

phone: +31 627881831 email: takehiro.tsurumi1031@gmail.com

<https://www.linkedin.com/in/takehirotsurumi/> | <https://github.com/tti440> |

<https://orcid.org/0009-0005-1405-3853>

AI Research Engineer who builds end-to-end MLOps systems, from data ingestion and ETL to model training, evaluation, and deployment across computer vision, multimodal, and generative AI. Hands-on experience with PySpark, Docker, CI/CD, Neo4j, and Kubernetes, delivering reproducible pipelines in both research and production environments. Published and award-winning research on bias and interpretability in Vision Transformers and Stable Diffusion (Best Paper AIMMES 2025, [ACM FAccT 2025](#)), with startup experience shipping ML-backed applications.

Languages: Python, SQL, C, JavaScript, C#

Tools & Frameworks: PyTorch, Hugging Face, Scikit-learn, OpenCV, Pillow, Django REST, PostgreSQL, FastAPI, Docker, Git, CI/CD, AWS S3, GraphDB, Neo4j, PySpark, Kubernetes

Soft Skills: Cross-functional collaboration, data reporting, teaching & mentoring, reproducibility & documentation, problem-solving

Vrije Universiteit Amsterdam (VU Amsterdam) - Research Assistant (Sep 2023 - Aug 2025)

- Built reproducible experimental pipelines, such as ETL pipelines, following academic MLOps best practices
- Conducted bias analysis in the interdisciplinary research from both computer vision and linguistics among commercialized models using ETLED datasets (IMDB, CelebA)
- Investigated biases in latent spaces across 10 different architectures (BEiT, Deit, MAE, SWIN, etc.) using iEAT, t-SNE, Grad-CAM, DBSCAN, and SAM2 segmentation; quantified bias at embedding and output layers. (Best Paper Award at AIMMES 2025 and published at [ACM FAccT 2025](#))
- Prepared and managed experimental environments for evaluating Stable Diffusion variants in thesis-related research.
- Integrated model interpretability techniques (Layer-wise Relevance Propagation) to analyze neural-level behavior, supporting explainable and ethical AI research.
- Prepared weekly reports for supervisors and communicated with team members from different fields, documenting progress, problems, solutions, and findings

Vrije Universiteit Amsterdam (VU Amsterdam) - Teaching Assistant (Sep 2023 - Aug 2025)

- Supported courses in Programming & Algorithms, Knowledge Graphs & Data, Agent Programming, Conversational Agents, and Modeling.
- Conversational Agents: Guided projects using Whisper, Dialogflow, and MARBEL, enabling teams to design voice-operable ML applications (ASR, NLU, NLG).
- Knowledge Graphs & Data: Assisted with graph construction, reasoning tasks, and data integration using GraphDB.
- Programming & Algorithms: maintained pytest-based testing and runtime evaluation for student assignment evaluation
- Delivered recap lectures, coding demonstrations, and assignment support, strengthening communication and collaboration.

Startup Collaboration: Magicle (Private, NDA) - AI Engineer (Oct 2024 - Mar 2025)

- Designed and implemented core components of an AI-driven outfit recommendation system, addressing cold-start scenarios through multimodal feature extraction.

- Built computer vision pipelines using SAM-2 and Swin Transformer models for apparel segmentation and visual attribute extraction.
- Developed scalable data processing workflows with Apache Spark to construct knowledge-graph entities and generate vector embeddings for semantic retrieval.
- Integrated hybrid graph and vector search using Neo4j to enable context-aware recommendation and re-ranking.
- Developed backend services and APIs within a containerized, microservices-based deployment environment with Django REST, PostgreSQL, and AWS S3.
- Implemented a frontend where users can make any collages using their own apparel items and some stamps inspired by Picsart UI using Django, Vanilla JS, and Tailwind.
- Contributed to MLOps workflows, including model versioning, automated retraining, and iterative deployment.

Waldburg-Zeil Kliniken (Germany) - Data Entry (Sep 2021 – Aug 2022)

- Communicated with dietitians about certain restrictions or conditions for patients

UNIQLO (Japan) - Retail Sales Assistant (Sep 2019 – Jan 2021)

- Developed teamwork, communication, and organizational efficiency skills in a fast-moving, high-performance environment.
- Contributed to continuous operational improvement through problem-solving and past data.

PASCO (Tokyo, Japan) - Computer Vision Intern (Jan 2018 – May 2018)

- Analyzed geospatial and satellite imagery for forest classification, coastline detection, and urban zoning.
- Built and trained CNNs from scratch (e.g., CIFAR10), gaining practical experience in model optimization, learning rates, and experimentation.
- Acquired foundational experience in applying deep learning to real-world datasets, strengthening problem-solving skills in data-driven decision-making.

Vrije Universiteit Amsterdam (VU Amsterdam)

BSc in Artificial Intelligence

Languages: English (Fluent), Japanese (Native)

Education & Languages

Projects & Awards

- Published Paper, [ACM FAccT2025: Comparative bias analysis across ViT architectures](#) (Best Paper Award, AIMMES202)
- [Stable Diffusion Bias Evaluation](#) - Designed and executed reproducible evaluation pipelines to assess fairness and bias in Stable Diffusion models across demographic attributes using controlled prompt sets and image-level analysis.
- [LLM Graph Generation](#) - Built an LLM-driven knowledge graph construction pipeline for extracting triples, refining ontologies, reasoning over graph to answer competency questions, and populating Neo4j using Ollama-based models
- [Neuron-Level Bias Analysis](#) - Developed a neuron-level interpretability pipeline by integrating Attention-LRP into Vision Transformers to analyze and quantify implicit bias at the representation level.
- [Trip Recommendation Platform \(Ongoing\)](#)
 - Developed a Dockerized AI recommendation environment integrating Spark, Neo4j, and Ollama for semantic search and itinerary generation.
 - Designed a multi-agent system using MCP (Model Context Protocol) to enable agentic, context-aware decision-making.
 - Deployed and managed distributed services using Minikube (Kubernetes) for scalable orchestration and testing.
 - Implemented vector similarity search and graph reasoning for ranking and personalized travel recommendations.