

# Konstantinos Gavriilidis

Edinburgh, United Kingdom | [konsgavriil@gmail.com](mailto:konsgavriil@gmail.com) | +44 7436832708

[linkedin.com/in/konsgavriil](https://linkedin.com/in/konsgavriil) | [github.com/konsgavriil](https://github.com/konsgavriil)

## Profile

---

Artificial Intelligence Engineer with proven expertise in advanced AI research and industrial deployment, delivering impactful data-driven solutions for real-world applications. Experienced in building and deploying ML models, fine-tuning large language models (LLMs), and implementing scalable MLOps pipelines. Specialized in Natural Language Processing (NLP), AI agents, and explainable AI, with a focus on developing transparent and reliable autonomous systems.

## Core Skills

---

**Programming & Machine Learning:** Python, Java, SQL, Pandas, NumPy, PyTorch, Scikit-learn, Feature Engineering, Supervised & Unsupervised Learning, Model Evaluation

**Generative AI:** LLMs, Hugging Face, PEFT, LangChain, RAG, AI Agents, MCP, Inference Guardrails

**MLOps & Deployment:** Docker, Kubernetes, Weights & Biases, AWS, Azure, GCP, Model Deployment, CI/CD

**Explainability & Responsible AI:** SHAP, LIME, Bias & Fairness Detection, Governance, Responsible Deployment

## Experience

---

### AI Engineer – Freelance

*MedAscend, Edinburgh | Nov 2025 – Present*

- Currently working on the AI stack of MedAscend, a virtual agent platform for medical staff training. Responsible for designing evaluation tasks, fine-tuning large language models, and analysing model behaviour to improve performance and reliability.

### Research Intern – Machine Learning & Explainable AI

*SeeByte Ltd, Edinburgh | Mar–Aug 2022 & Mar–Aug 2024*

- Fine-tuned large language models to automate post-mission reporting, reducing manual workload by 70% and improving analyst efficiency.
- Designed and deployed a surrogate model framework achieving 90% accuracy to explain autonomous maritime agent decisions, enhancing operator trust and safety.
- Delivered technical workshops on LLMs and Transformer models; contributed to governance frameworks for mission-critical AI systems.

### Teaching & Research Assistant

*Heriot-Watt University, Edinburgh | Jan 2021 – Dec 2022*

- Developed a Unity-based telepresence system enabling low-latency remote control of underwater robots, improving inspection capabilities across distributed sites.
- Supported delivery of Software Engineering, Data Science, and ML courses; mentored students on applied AI and robotics projects.

### Machine Learning Engineer

*SnapDragon Monitoring Ltd, Edinburgh | Jun 2018 – Sep 2019*

- Integrated deep learning models into production pipelines, achieving 98% accuracy in product listing classification and 81% accuracy in counterfeit detection.

### Software Developer Intern

*Veltio Ltd, Thessaloniki | Nov 2015 – Apr 2016*

- Built a web-based tool for managing Docker operations, reducing deployment time and improving developer productivity.

## Relevant Projects

---

### Enabling Human-Agent Interaction with Large Language Models

- Developed and evaluated multiple styles of human-agent interaction to enhance user situation awareness.
- Fine-tuned large language models (Llama 2, Mistral, and Falcon) on separate tasks to generate three distinct types of explanations for autonomous agent decisions.
- Conducted feature engineering on robot state data, model benchmarking, and output evaluation using semantic accuracy metrics and expert-based qualitative assessments, ensuring the explanations were reliable, interpretable, and aligned with user needs.

### Surrogate Model Framework

- Designed and implemented a framework leveraging transparent surrogate models to approximate autonomous decision-making and uncover the causal factors behind an agent's actions.
- Integrated both interpretable and black-box models with SHAP-based feature attribution for robust model selection and explainability.
- Achieved a 90% fidelity in replicating the decision-making behaviour of maritime autonomous robots, enabling transparent and trustworthy model insights.

## Education

---

### PhD, Robotics and Autonomous Systems

*Heriot-Watt University, Edinburgh | 2019 – 2025*

- Thesis: Explainable Reasoning for Remote Autonomous Agents.

### MSc, Artificial Intelligence with Speech and Multimodal Interaction

*Heriot-Watt University, Edinburgh | 2017 – 2018*

- Thesis: Counterfeit Product Detection with Deep Learning.

### BSc, Informatics Engineering

*Technological Educational Institute of Central Macedonia, Greece | 2011 – 2016*

## Certifications

---

### Model Context Protocol (MCP) Course

*HuggingFace | Oct 2025*

- Gained hands-on experience in integrating large language models with external tools, data sources, and services through standardised communication protocols.

### AI Agents Course

*HuggingFace | Mar 2025*

- Acquired expertise in designing, building, and deploying intelligent AI agents.

### Machine Learning - Stanford University

*Coursera | Nov 2020*

- Developed a strong foundation in supervised and unsupervised learning algorithms, including their training, evaluation, and optimization.

## Awards & Scholarships

---

### Best Poster Award

*Scottish Research Partnership in Engineering (SRPe) Conference | Oct 2023*

- Recognised for research on explainable AI for autonomous maritime systems.

### PhD Scholarship - Edinburgh Centre for Robotics

*SeeByte Ltd & SRPe | 2019–2025*

- Fully funded doctoral scholarship in Explainable AI for autonomous decision-making.

### MSc Scholarship - Heriot-Watt University

*DataLab Innovation Centre | 2017–2018*

- Merit-based scholarship supporting postgraduate study in Artificial Intelligence.

## Publications

---

1. **Enhancing Situation Awareness Through Model-based Explanation Generation** [Paper link]  
*International Natural Language Generation Conference (INLG)*, 2024
2. **A Surrogate Model Framework for Explainable Autonomous Behaviour** [Paper link]  
*IEEE International Conference on Robotics and Automation (ICRA)*, 2023
3. **Plan Verbalisation for Robots Acting in Dynamic Environments** [Paper link]  
*International Conference on Automated Planning and Scheduling (ICAPS)*, 2021