Jared Joselowitz

49 Otter Close, London, E152PZ, UK
☎ +44 744 0408647 ✉ jaredjoss123@gmail.com linkedin.com/jaredjoselowitz github.com/jaredjoss

AI Research Engineer with an MSc in Applied Machine Learning from Imperial College London, experienced in developing and deploying ML models, collaborating with cross-functional teams, and solving real-world challenges using cutting-edge AI techniques. Currently working with LLMs in the clinical domain, focusing on building reliable, interpretable, and steerable AI systems that prioritize safety and deliver societal benefits.

EDUCATION

Imperial College London 2023-2024
MSc in Applied Machine Learning *London, United Kingdom*

- **Grade**: Distinction.
- **Relevant Coursework:** Machine Learning (linear regression, logistic regression, SVMs, Knn), Deep Learning (CNN, RNN, RL, Autoencoders, VAE, GAN), Advanced Deep Learning Systems, Topics in Large Dimensional Data Processing, Computer Vision and Pattern Recognition (PCA, LDA, decision trees), Optimization.
- **Dissertation (Advisor: Dr. Sonali Parbhoo):** Inverse Reinforcement Learning For Large Language Model Interpretability. Fine-tune open source LLMs to reduce toxicity using RLHF and extract the reward model using IRL.

University of the Witwatersrand 2018-2021
BSc in Electrical (Information) Engineering *Johannesburg, South Africa*

- **Grade**: Distinction. 22 distinctions overall.
- **Relevant Coursework:** Software Development I/II/III, Data Structures and Algorithms.
- 1st place in 4th year and Top 5 in 3rd year
- **Final year thesis (Advisor: Dr. Vered Aharonson)**: “Quantizing the Effectiveness of Scientific Speeches on Climate Change” received a mark of 97%. Led to four publications.

EXPERIENCE

Ufonia Feb 2025 – Present
Senior AI Research Engineer *London, United Kingdom*

- Developing LLM-based clinical conversational systems for automated postoperative follow-up.
- Co-authored a research paper accepted at ACL 2025 (ASTRID).
- Benchmarked open-source and multilingual LLMs to guide model selection and inform future research directions.
- Improved evaluation pipelines and model stability, achieving over 95% benchmark accuracy and reduced latency.

The Awareness Company Sep 2022 – July 2023
Junior Data Scientist *Johannesburg, South Africa*

- Worked collaboratively in a cross functional engineering team on an AI-powered data storytelling platform.
- Used data analysis to investigate the robustness of Machine Learning models to maximise their performance and robustness, including a computer vision model (using Azure Computer Vision and Custom Vision Cognitive Services) to identify animals, people, and cars and an energy demand model to predict energy usage in buildings.
- Designed, trained, and deployed custom ML models to solve real-life challenges including a soil moisture predictor and a poaching predictor capable of identifying high-risk areas for poaching in the Kafue National Park, Zambia.
- Utilized Azure OpenAI Service to fine-tune a state of the art LLM-based chatbot with company-specific data for deployment and hosting to end users via an API.

Tesserae Feb 2022 – Aug 2022
Junior Full-Stack Software Engineer *Johannesburg, South Africa*

- Worked in an engineering team of five people on a Full stack AI annotation platform used to provide business intelligence in the FinTech industry.
- Worked on both the frontend (React) and backend (Django) to design, develop, and deploy user-facing features.
- Built an end-to-end testing infrastructure in Python using Cypress and created a bot to enable WhatsApp annotation, deployed using AWS Lambda functions.

LearnTech Dec 2019 – Jan 2020
Software Development Intern *Johannesburg, South Africa*

- Worked as part of a four-person team tasked with developing a gamification application.
- Learned database querying using PHP MySQL and Git for source control.

RESEARCH & COLLABORATIONS

BTheChange: Erasmus+ KA2 Higher Education Project | *NLP, Sentiment Analysis, Educational AI*

- Collaborated with international researchers on an Erasmus+ KA2 project applying AI and machine learning to improve climate change education.
- Worked on the development of interactive learning tools and large-scale data mining pipelines to study how audiences engage with climate communication.
- Applied Natural Language Processing (sentiment and emotion analysis) to evaluate public perception and identify factors that improve educational impact.
- Contributed to publications and presentations at international conferences, advancing the understanding of AI-driven educational strategies.

PROJECTS

Masters Thesis: Inverse Reinforcement Learning For Large Language Model Interpretability | *LLM, RLHF, IRL, TRL, HuggingFace*

- Used the TRL library to take a groundtruth reward model and fine-tune open source Pythia LLMs (70M and 410M parameters) using RLHF to reduce the toxicity of the models.
- Developed a novel max-margin Inverse Reinforcement Learning algorithm applied to LLMs to extract the reward model from the fine-tuned LLMs. Found that IRL can effectively extract reward models that closely approximate the original RLHF objectives. The accuracy of the algorithm reached 80.40%.
- Analysed the results and showed that normal correlation metrics were insufficient for assessing viability for this application. Other techniques including policy comparison were done to show IRL proficiency.
- Using the IRL reward model to fine-tune an LLM results in a lower toxicity than the original RLHF process. Therefore, not only does the work enhance interpretability but also alignment with human values.

Bachelors Thesis: Quantifying the Effectiveness of Speeches on Climate Change | *NLP, NLTK, Scikit-learn*

- Used Natural Language Processing to develop a method capable of quantifying the effectiveness of speeches on climate change by analysing user comments.
- Leveraged NLTK for text preprocessing and feature engineering, followed by developing an ensemble of machine learning models in scikit-learn, achieving 82.35% classification accuracy.
- Analyzed key factors contributing to the video effectiveness, including participant profiles and video formats.

TECHNICAL SKILLS

Languages: Python, MATLAB, JavaScript, R, C++, HTML
Technologies: TensorFlow, PyTorch, Keras, Scikit-learn, HuggingFace, NLTK, Pandas, NumPY, Scipy, LangChain, Matplotlib, Seaborn, React.js, Django, Node.js
Developer Tools: Git, High Performance Computing, SQL (SQLServer, MySQL, PostreSQL, NoSQL), Docker
Cloud Computing: AWS (S3, Lambda, DynamoDB, Elastic Beanstalk, SageMaker), Azure (AI studio, OpenAI Service, Computer Vision and Custom Vision Cognitive Services, CosmosDB)
Concepts: Artificial Intelligence, Machine Learning, Deep Learning, Neural Networks, NLP, Reinforcement Learning, RLHF, LLM, IRL, transformers, Imitation Learning, API, Agile Methodology, Cloud Computing, OOP, HPC

PUBLICATIONS

Insights from the Inverse: Reconstructing LLM Training Goals Through Inverse Reinforcement Learning
Jared Joselowitz, Ritam Majumdar, Arjun Jagota, Matthieu Bou, Nyal Patel, Satyapriya Krishna, Sonali Parbhoo
Conference of Language Modeling (COLM 2025)

ASTRID - An Automated and Scalable TRIaD for the Evaluation of RAG-based Clinical Question Answering Systems
Yajie Vera He, Mohita Chowdhury, Jared Joselowitz, Aisling Higham, Ernest Lim
Findings of the Association for Computational Linguistics: ACL 2025

Video Selection for Enjoyable Learning
Slawomir Nowaczyk, Jared Joselowitz, Vered Aharonson
14th Canada International Conference on Education (CICE-2024)

On Presenters and Commenters in YouTube Climate Change Videos
Vered Aharonson, Jared Joselowitz
11th European Conference on Social Media (ECSM 2024)

Video Features Predicting Engagement in Climate Change Education
Vasiliki Christodoulou, Vaggelis Saprikis, Louiza Kythreotou, Monogios Christodoulos, Ece Calikus, Jared Joselowitz
E3S Web of Conferences Journal (2023)

Collaborative Learning in YouTube: Under Which Conditions Can Learning Happen or Fail to Happen?
Kalypso Iordanou, Vered Aharonson, Vasiliki Christodoulou, Christos Karpasitis, Jared Joselowitz, Byron Lilford, Maura De-Vos, Smrithi Muraleedharan
15th International Conference on Computer-Supported Collaborative Learning (CSCL 2022)

ACHIEVEMENTS

Entelect Prize (1st place in my 4th year of study)	2021
Isazi Consulting Prize (Top 5 students in my 3rd year of study)	2020
Certificates of Merit (Signals and Systems IIA, Physics II, Data Structures and Algorithms, Mathematics I)	2018-2020
Golden Key Society	2018
Dean’s List	2018
University Entrance Scholarship	2018

** References available on request*