Kieran Fraser, PhD

Staff Machine Learning Engineer. Production ML, fraud detection and adversarial AI safety. Dublin, Ireland — open to remote / EU-based roles.

kfraser9292@gmail.com · linkedin.com/in/kieran-fraser-phd-608a54a7 · github.com/kieranfraser · Google Scholar · kieranfraser.ie

Summary

Staff Machine Learning Engineer with a PhD in Computer Science and eight-plus years spanning production machine learning, fraud and trust & safety systems, and adversarial AI research. Currently build and deploy fraud-detection models protecting 50M+ PayPal users; previously a Research Engineer on IBM's Security and Privacy of AI team building open-source LLM red-teaming and guardrail tools (HEART, ART, ARES, Granite Guardian), and a PhD researcher designing more considerate, context-aware notification and engagement systems, commercialised as the startup EmPushy. Comfortable across the full arc from research to production: shipping models at scale, filing patents, and translating ML work into customer- and business-facing outcomes.

Core skills

Machine learning & AI — machine learning, deep learning, PyTorch, TensorFlow, scikit-learn, large language models (LLMs), generative AI, retrieval-augmented generation (RAG), vector databases, LLM fine-tuning (LoRA/PEFT), prompt engineering, computer vision (image & video classification, object detection, segmentation), graph neural networks, adversarial robustness, AI red-teaming & guardrails, synthetic data generation, agentic AI, fraud detection, MLOps

Engineering — Python, Java, C++, JavaScript, SQL & NoSQL, Docker, Node.js, React, HTML5/CSS/SCSS

Cloud & infrastructure — AWS (SageMaker, Lambda, S3, DynamoDB, API Gateway, ECR, Batch, EC2, IAM, Cognito, Athena, Glue, Amplify), Google Cloud Platform, Microsoft Azure

Leadership & delivery — technical mentorship, cross-functional collaboration, patent development, project management, customer & trial-partner engagement, academic research & publication

Experience

Staff Machine Learning Engineer, PayPal

NOV 2025 – PRESENT · DUBLIN

- Deployed a flagship identity and onboarding fraud-detection foundation model serving 50M+ PayPal users, improving on prior versions across detection accuracy, efficiency and explainability.
- Shaped architecture and strategy for a redesign of PayPal's Global Fraud Prevention platform, partnering with cross-domain teams.
- Lead inventor on a filed patent applying model-uncertainty measures to adaptive step-up authentication; founded a cross-team patent group within the Fraud & Data Science Centre of Excellence to help colleagues navigate the IP process.
- Mentor two engineering interns.

Research Engineer, IBM Research

AUG 2022 – NOV 2025 · DUBLIN

- Contributed to HEART, ART and ARES — IBM's open-source toolkits for adversarial-robustness evaluation and automated LLM red-teaming — including 104 commits to HEART per its public contributor record.
- Extended adversarial-robustness evaluation and defense coverage in ART across computer vision (image and video classification, object detection, segmentation) and graph-based models.
- Co-authored peer-reviewed and arXiv research on LLM guardrails, jailbreak detection and red-teaming methodology (Granite Guardian, MoJE, MAD-MAX, Attack Atlas), plus synthetic data generation and agentic-AI workload management.
- Named inventor on four published patent applications spanning federated learning, retrieval-augmented generation security, visual-content anonymisation and model-capability analysis.
- Represented IBM at IEEE ERAS 2025 (Worcester Polytechnic Institute) and AI4Media's demonstration sessions at the University of Florence.

Research Assistant, ADAPT Centre for Digital Content Technology

SEP 2020 – AUG 2022 · DUBLIN

- Led an Enterprise Ireland-funded Feasibility Study and Commercialisation project (EmPushy) in empathetic push-notifications, recruiting and managing a small team.
- Built and deployed an MVP; analysed 3M+ real-world notifications and presented findings directly to partner CTOs and R&D leads.
- Managed AWS infrastructure, GDPR compliance and IP documentation for the project.

Earlier experience

2013 – 2016

Android Engineer at Lumafit (2016) — migrated a local Android database to a real-time cloud backend. Executive Officer at the Revenue Commissioners (2015) — rebuilt internal web applications with the Spring Framework. Data Analyst at the Aon Centre for Innovation & Analytics (2014) — built a BI dashboard (SQL, Tableau) presented to ACIA's CEO, CBO and CCO. Post-processing Engineer at Septentrio, Belgium (2013) — contributed to in-house testing and CI (Jenkins) for satellite-navigation software.

Education

PhD Computer Science, Trinity College Dublin

SEP 2016 – MAR 2021 · DUBLIN

Thesis: *ETHOS Push* — EmpaTHetic Orchestration of Scrutable Push. A framework for transparent, context-aware, empathetic push-notification delivery.

MAI Computer Engineering, Trinity College Dublin

SEP 2015 – MAY 2016 · DUBLIN

Graduated with Distinction.

BAI Computer Engineering & BA Mathematics, Trinity College Dublin

SEP 2011 – MAY 2015 · DUBLIN

Graduated with Second Class Honours, First Division (2.1).

Selected publications

19 publications in total. Full list on Google Scholar or the Work page.

Granite Guardian: Comprehensive LLM Safeguarding
I. PADHI, M. NAGIREDDY, G. CORNACCHIA, ET AL. · NAACL 2025

Attack Atlas: A Practitioner's Perspective on Challenges and Pitfalls in Red Teaming GenAI
A. RAWAT, S. SCHOEPF, G. ZIZZO, G. CORNACCHIA, K. FRASER, ET AL. · 2024

MoJE: Mixture of Jailbreak Experts, Naive Tabular Classifiers as Guard for Prompt Attacks
G. CORNACCHIA, G. ZIZZO, K. FRASER, ET AL. · AAAI/ACM AIES 2024

ETHOS Push: An Approach toward Empathetic Orchestration of Scrutable Push — PhD thesis, Trinity College Dublin 2021

A Context-Aware, Info-Bead and Fuzzy Inference Approach to Notification Management
K. FRASER, B. YOUSUF, O. CONLAN · UEMCON 2016 · BEST PAPER AWARD

Patents

Named inventor on four published IBM applications and two further applications filed at PayPal (not yet public). Full record on Google Patents.

Securing retrieval-augmented generation IBM

Model emergent capability analysis IBM

Hobbies

CrossFit, golf, tag rugby, art, photography, VR and photogrammetry — more on the Elsewhere page.