# Fraser Greenlee-Scott

**Email**: fraser.greenlee@mac.com **Mobile**: +447989190626 **LinkedIn**: linkedin.com/in/fraser-greenlee-9658a7101/
**Blog**: frasgreen.com **GitHub**: github.com/Fraser-Greenlee

## Work

**Principal Software Engineer (Autonomous Agents ML Research Lead) - Salesforce**, London, UK (July 2025-Present)
Building RL infrastructure for training GUI agents in simulated desktop environments.
- Developed a Ray-based RL training library scaling to 1k+ parallel environments with 126 steps/hour at batch size 1024
- Created a task generator and evaluator yielding +10% absolute success rate improvements on OS-World
- Led a team to train a computer-use agent from non-restricted open-source weights

**Founding Member of Technical Staff (Research Scientist) - Convergence Labs (Acquired by Salesforce)**, London, UK (July 2024-July 2025)
Built the foundational codebase for Proxy, the company's flagship web agent product.
- Designed the browser tool specification and Reflexion-style prompting framework that maximised agent performance
- Created the grounding setup enabling VLMs to interact with web elements before coordinate-based grounding was reliable
- Built the evaluation infrastructure for web agents; detailed trajectory analysis drove iterative improvements to our agents
- Developed the training pipeline for GUI agent models, producing 10k+ synthetic trajectories and processing customer trajectories into an SFT dataset with multi-layer automatic quality evaluation at both trajectory and step level
- Trained Proxy-Lite, a 3B VLM achieving SOTA on WebVoyager at its scale (938 GitHub stars); also open-sourced a lightweight Playwright-based web agent framework
- Halved inference costs by compressing the two-step think-then-act loop into a single reasoning step
- Implemented context management enabling agents to solve long-horizon 100 step tasks
- Led setup of distributed vLLM hosting serving 125k+ users; step-level evaluation heuristics were integrated into production to monitor model performance

**Machine Learning Engineer - Cohere**, London, UK (May 2022-July 2024)
Started the Command project, a generic instruction following LLM which is now Cohere's primary model.
- Produced core instruction following and preference datasets with external and internal annotation teams.
- Led human preference evaluations to benchmark our models against competitors. Ran multiple evaluations weekly, with results presented to the board to demonstrate competitive parity.
- Created an automated human evaluation tool using reward models, which became the company standard with hundreds of auto-evals conducted across teams.
- Created training infrastructure using JAX used to create state-of-the-art reward and vision models.
- Implemented a shared tokenizer used across Cohere's internal repos and resolved issues with the production tokenizer.
- Research - Spearheaded improvements in reward models, achieving SOTA results on several benchmarks including RewardBench where we outperformed GPT-4 and other top LLMs in preference modelling. Co-authored the paper "Improving Reward Models with Synthetic Critiques".

**Machine Learning Engineer - Darktrace**, Cambridge, UK (March 2021-May 2022)
As part of the Developer-Analyst team, I made ML models & services to alert customers on indicators of compromise.
- Led replacement of old heuristic-based classifiers to using transformers, categorising hostnames, file paths, emails & more.

**Software Engineer - Skyscanner**, Edinburgh, UK (Sept 2019-March 2021)

## Education

University of Glasgow - BSc Software Engineering, 1st Class (2015-2019)

## Projects/Awards

**Hugging Face JAX Contest** (June 2021)
Top 15 in Hugging Face Flax-JAX Community Week (2021) trained a T5-based MMD-VAE on TPUv3-8

**Google's 'Show Your Work' Paper Dataset (900+ citations)** (2019-2020)
Developed a dataset of Python state changes, crucial for the groundbreaking "Show Your Work: Scratchpads for Intermediate Computation with Language Models" paper by Google Research.

**Pioneer Award - Winner** (2019) Created a VSCode Extension for coders that got over 27k installs.