

OLLY STYLES

@ ollystyles@gmail.com

📍 London, UK

📧 [Olly Styles](#)

🔗 [olly-styles](#)

in [olly-styles-090437132](#)

🔗 [olly-styles.github.io](#)

AI engineer who ships and evaluates production LLM systems. Author of [WorkBench](#) (COLM 2024; [code](#)). PhD in ML, ex-founder (Techstars '21), 7+ years of ML engineering.

EXPERIENCE

Founding AI Engineer

Artanis *AI alignment platform for keeping model behaviour consistent in production*

📅 2025 – Present

- Own customer deployments end-to-end, from scoping with domain experts through build, production rollout, and iteration.
- Built an order-processing system that uses an LLM to parse 4k customer emails into structured SAP orders with a human-in-the-loop review step; 2k orders validated to date, halving per-order processing time and reducing manual-entry errors.
- Built the labelling and evaluation system behind an LLM essay grader deployed in 100+ schools; domain experts have created 21k labels (and counting) to iteratively refine the grading policy.

Senior ML Engineer

Multiverse *Upskilling platform for AI adoption*

📅 2024 – 2025

- Built [Instant Feedback](#), delivering automated rubric-based feedback on apprentice project submissions for Multiverse's 20,000+ learners; internally validated as on par with human markers over six months of testing.

Cofounder / CTO

Atlas AI *Vision-based AI personal trainer*

📅 2019 – 2024

- Led the founding tech team, building our product from scratch; co-funded £450k from [Techstars Montréal AI](#), [Innovate UK](#), and others.
- Built the end-to-end labelling, training, and deployment pipeline for a video object tracker that generalised zero-shot to gyms not in the training dataset.

AWARDS

- **Best Student Paper Award**, CVPR 2020 Precognition workshop, for work on Multi-Camera Trajectory Forecasting.
- **Best Research Presentation Award**, University of Warwick, 2020, for talk on [Anticipating Pedestrian Movements for Safer Self-driving Vehicles](#).

SKILLS

- **Agent evals**: agent and tool-use benchmark design, deterministically verifiable outcome evaluation, ground-truth and side-effect verification, safety-oriented task design (irreversible actions, data leakage).
- **Data**: dataset creation, LLM annotation and review pipelines, LLM-as-judge grading policies, structured-extraction evaluation.
- **Engineering**: backend and API design, cloud, CI/CD; Python, FastAPI, PostgreSQL, Git, uv.

EDUCATION

PhD in Machine Learning

University of Warwick

📅 2017 – 2021

- Published work on [scaling trajectory forecasting pre-training](#) in 2019 for safer self-driving vehicles, then showed the pre-training generalised zero-shot to an [unseen dataset](#) in 2020.
- Thesis: [Scalable Methods for Single and Multi Camera Trajectory Forecasting](#).
- Taught classes on Machine Learning, Digital Forensics, and Object-Oriented Programming.

Masters in Computer Science

University of Warwick *1st Class*

📅 2013 – 2017

- Theses: emotion classification from speech (CNNs on spectrograms); online-course dropout prediction (LSTMs on time series).

PUBLICATIONS

247 citations, h-index 6 ([Google Scholar](#)).

- **O. Styles** and S. Miller, [WorkBench Revisited: Workplace Agents Two Years On](#), **Preprint** 2026. *Re-evaluates 24 frontier models: task completion 43%→98%, harmful side effects largely eliminated (26%→1.9%); rare irreversible harms persist.*
- **O. Styles** et al., [WorkBench: A Benchmark Dataset for Agents in a Realistic Workplace Setting](#), in **COLM 2024**. *690 tasks with deterministic, outcome-based evaluation of tool-using agents, including their harmful side effects.*

Others

- **O. Styles** et al., [Multi-Camera Trajectory Forecasting with Trajectory Tensors](#), in **PAMI** 2021.
- **O. Styles** et al., [Multi-Camera Trajectory Forecasting: Pedestrian Trajectory Prediction in a Network of Cameras](#), in **CVPR-W** 2020.
- **O. Styles** et al., [Multiple Object Forecasting: Predicting Future Object Locations in Diverse Environments](#), in **WACV** 2020.
- **O. Styles** et al., [Forecasting Pedestrian Trajectory with Machine-Annotated Training Data](#), in **IV** 2019.

OPEN SOURCE & WRITING

- 200+ GitHub stars across MIT-licensed research repos, including [WorkBench](#).
- [Writing](#) on agent evals, alignment, and startups.