

Julian Quick

I build **evaluations and failure monitors** for AI and autonomous systems. Ten years turning ambiguous notions of trustworthiness into defensible metrics, from gigawatt turbine fleets to frontier language models.

27 / 7

LLMs / provider families audited for silent error suppression

0.92

AUROC, policy-failure prediction on corruptions unseen in training

+72%

validation scenario coverage via risk-targeted evaluation

-60%

false alarms, anomaly monitors running in production

781

citations · h-index 14 · 26 peer-reviewed publications

● LLM EVALUATION & RED-TEAM PROJECTS

eager-baker: benchmarking agentic scope calibration. Pre-registered benchmark isolating scope over-eagerness from task accuracy: models spread 20-72% on over-eagerness vs ~6 pts on accuracy; every prompting intervention suppressed capability rather than calibrating behavior.

Silent Killers: auditing error swallowing in LLM-generated code. Audited 27 frontier and open-weight models across 7 provider families: 50-100% of generated exception handlers silently suppress errors on scientific-computing tasks. Shipped static-detection CLI (PyPI, pre-commit).

Visual safety monitor for robot manipulation: generalization to unseen failure modes. 33K-parameter monitor predicts vision-language-action policy failure from camera corruption; leave-one-out AUROC 0.92 on corruption types held out of training. [AIAA SciTech](#), [under review](#)

Turnstile: mechanistically contrasting harm and compliance. Multi-turn adversarial self-play (DPO, 4,500 conversations): victim models detect jailbreaks yet comply 26% of the time; layer-16 causal steering shifts compliance bidirectionally (+11 / -13 pp) without increasing judged harmful uplift.

Adversarial self-play for hardening safety-critical control. RL adversary learns temporally correlated sensor attacks; self-play curriculum recovers worst-case performance from -39% to +7.9% with no degradation in clean environments. [TORQUE 2026](#), [peer-reviewed](#)

● EXPERIENCE

Visiting Researcher · Lawrence Berkeley National Laboratory, Berkeley CA

01/2026 – present

- Built an agentic capability-evaluation harness: LLM agents iteratively modify optimization code, scored on held-out scenarios. Findings: scaffolding explains more variance than model scale, and frontier models discover non-monotonic strategies absent from the published literature.

Researcher · Wind & Energy Systems, Technical University of Denmark (visiting appointment above)

05/2022 – present

- Deployed XGBoost / LSTM anomaly-detection monitors for autonomous wind-turbine fleet safety, cutting production false-alarm rate 60% while preserving early-failure warning.
- Led V&V strategy across 5 independent simulation platforms via a shared schema; built a predictive validation tool that increased high-risk scenario coverage 72%. Informs the IEC 61400-60 standard.
- Shipped stochastic optimization enabling 1,200+ unit fleet design, 95% faster, in production via TOPFARM; maintain CI/CD and test automation for open-source design tools used by 30+ companies.
- Uncertainty-quantification lead, IEA Wind JAM task; led short-term power forecasting project for TotalEnergies R&D.

Graduate Researcher · Turbulence & Energy Systems Laboratory, CU Boulder

08/2017 – 01/2022

- Built multi-fidelity uncertainty-propagation and optimization frameworks cutting V&V simulation cost 70-90%; designed uncertainty-aware control reducing critical mechanical failure risk 47%; deployed in field operations, cited 100+ times.

Earlier · NREL · NCAR, Boulder CO

2014 – 2017

- NREL: Python interface to DAKOTA C++ API with multi-node parallelism for stochastic V&V at HPC scale. NCAR: live SQL monitoring system alerting research-aircraft technicians in real time when instrument data left valid range.

● TOOLBOX · EDUCATION · COMMUNITY

TOOLBOX

Python (10+ yrs) · PyTorch & HF (DPO fine-tuning) · XGBoost / LSTM · multi-node HPC & parallel pipelines · CI/CD, test automation (>90% coverage) · SQL · statistics, experimental design & UQ

EDUCATION

Ph.D. Mechanical Engineering, CU Boulder, 2022 (uncertainty quantification for CFD systems) · **BlueDot Impact**, 2026: Technical AI Safety Project; AGI Strategy

COMMUNITY

Technical lead, IEA Wind Task 55 (3 GW reference systems, adopted on 5+ platforms) · invited seminars incl. TotalEnergies R&D series · reviewer for 6 journals incl. IEEE Trans. Control Systems