

Wenhan Zhou

AI Systems & Evaluation Engineer

AGENT SYSTEMS · EVALUATION DESIGN · APPLIED RESEARCH

📍 Uppsala, Sweden (available in Stockholm regularly) · Email · [in](#) LinkedIn · [GitHub](#) · [Website](#)

SUMMARY

I build AI systems that help people solve complex, open-ended problems more reliably. I design agent workflows that investigate, have their work independently reviewed, and improve from real-world feedback. I applied this approach to cybersecurity research, reaching Microsoft's MSRC Q2 2026 researcher leaderboard within two months. My background spans agent systems, machine learning research, and product development.

EXPERIENCE

Independent AI Systems & Evaluation Researcher

2026–Present

Building agent systems that turn expert feedback into reusable evaluation criteria.

- Placed **#143 on the Microsoft Security Response Center (MSRC) Q2 2026 Points Leaderboard** after beginning cybersecurity research two months earlier; used the system to investigate Microsoft open-source software and submit 20+ security reports for review. [Leaderboard Evidence](#)
- Built a Git-versioned agent workflow where a Markdown specification defines research goals, agent roles, evidence requirements, candidate states, severity thresholds, and stopping rules. [Research-loop architecture](#)
- Extended investigations from frequent human steering to several hours of autonomous work by separating coordinator, generator, and reviewer roles and keeping reviewers in isolated contexts. [Research-graph architecture](#)
- Converted MSRC rejection feedback and previously implicit program constraints into screening rules that helped filter weak candidates before submission.
- Designed and operated agents that built disposable test environments from cloned repositories using scripts, containers, virtual machines, synthetic data, computer use, and cloud deployments.
- Documented the pattern as a reusable method for preserving expert feedback across future agent runs so investigations do not start from scratch.

Technical Lead / Interim CTO — Spucce Venture AB

2026 Q2

Turned a broad venture-studio idea into a concrete product and execution plan.

- Defined the first product: a company memory and workflow system that could later support a larger venture platform.
- Created the MVP architecture, product sequence, team responsibilities, fallback plans, and three-month roadmap.
- Designed a source-grounded memory approach that kept raw evidence, generated summaries, and provenance separate so people could inspect where an answer came from.
- Rejected fully automated memory rewriting after identifying a risk that silent errors could propagate to every downstream agent.
- Introduced [documentation-driven development](#) so intent, constraints, ownership, and lessons remained version-controlled, helping keep AI-assisted implementation aligned with product goals.
- Produced investor-facing technical strategy and helped a four-person team agree on what to build first under pre-seed constraints.

Solo Founder — Decision Brief

2026 Q1

Designed and built an AI decision-support product from idea to deployment. [Live app](#)

- Built and deployed the product in three weeks, covering product strategy, application architecture, implementation, SQLite storage, Stripe payments, and operational workflows.

- Separated unpredictable LLM output from reliable application logic and explicit failure handling so model failures were easier to inspect and contain.
- Observed five users interact with the product and simplified onboarding by removing steps and reducing cognitive load rather than adding features.
- Diagnosed distribution and trust as the binding constraints and stopped further engineering when additional product work would not solve them.

ML Researcher — Uppsala University

2023–2025

First author of a peer-reviewed paper on neural surrogate models for nuclear engineering. [Publication](#)

- Took over a loosely scoped Temporal Convolutional Network project and independently built the data parser, preprocessing pipeline, training loop, model implementations, evaluation code, and research plots.
- Introduced Fourier Neural Operators as an alternative and co-developed the Temporal Frequency Network (TFN), combining complementary time-domain and frequency-domain modeling.
- Reduced validation error by about 36% versus comparable TCN and FNO baselines; achieved 12.7 ms inference per 1,000 fuel rods and 1.70% mean relative error across seven safety-related outputs.
- Tested whether the improvement was robust using five-fold cross-validation, ablation studies, compute-matched comparisons, and failure analysis.
- Developed a custom loss for near-zero outputs and documented failure cases caused by artifacts in the simulator data.
- Presented the work at TopFuel Grenoble 2024; the research direction was later continued by the project supervisor at Westinghouse.

EDUCATION

Human Cognition & AI — Uppsala University

2025–2026

- Studied cognitive load, learning theory, qualitative interviewing, and human-computer interaction.
- Applied hypothesis-driven learning and adaptive questioning to AI-assisted research, product design, and technical discovery.

MSc Data Science — Uppsala University

2023–2025

Thesis: [Uncertainty Quantification for Receiver Operating Characteristic Curves](#)

- Recognized that ROC curves can be treated as conditional cumulative distribution functions.
- Developed finite-sample, distribution-free confidence bands that cover the entire ROC curve rather than only individual points.
- Compared bootstrap, Beta-Binomial, and Monte Carlo approaches using theory, simulation, and ECG data.

BSc Physics — Uppsala University

2020–2023

Thesis: [Deep Neural Networks as Surrogate Models for Fuel Performance Codes](#)

- Built the first version of the TFN surrogate model and an encoder-decoder check for unreliable out-of-distribution predictions.
- Originally scoped as a master's thesis; awarded the Sigvard Eklund Prize for Best Nuclear Engineering Thesis (30,000 SEK).

SKILLS

AI systems: multi-agent orchestration · LLM evaluation · LLM-as-a-judge · human-in-the-loop systems · long-running agents · computer use

Engineering: Python · PyTorch · Git · Markdown · Docker · Linux · SQLite · cloud deployment

Research: uncertainty quantification · deep learning · experimental design · ablation studies · failure analysis

LANGUAGES

Swedish · English · Chinese