

Wen Xing

She/Her | San Francisco Bay Area | wen.xing.us@gmail.com | (650) 666-5695 | <https://wenx.io>

SUMMARY

AI Safety Researcher and former Meta Integrity Tech Lead / Staff ML engineer. I've prioritized making technology transparent and safe throughout my career, from global-scale social media harm mitigation to AI monitorability research.

PUBLICATIONS

- **Frontier Models Can Take Actions at Low Probabilities.** Co-first author. Accepted at ICML 2026 main conference and ICLR 2026 Trustworthy AI Workshop.
- **Can Reasoning Models Obfuscate Reasoning? Stress-testing CoT Monitorability.** Co-first author. NeurIPS 2025 FoRLM Workshop. arXiv 2510.19851.
 - Findings used by Anthropic in Claude Opus 4.6 safety evaluations and cited by OpenAI's Monitoring Monitorability paper
- **Vulnerability in Trusted Monitoring and Mitigations.** Co-first author. Research blog post.

RELEVANT EXPERIENCE

ML Alignment and Theory Scholars (MATS) June 2025 - present

- Conducting empirical research with Deepmind mentors Erik Jenner and David Lindner.

SPAR Research Mentor Feb 2026 - present

- Mentoring 3 undergrad researchers on a reasoning project I designed and proposed

Independent AI Research Engineer 2024 - present

- Conducted funded AI Control research in AI Safety Camp, published a research blog post
- ARENA at Recurse Center

Meta - Integrity Team — ML Tech Lead 2019 - 2021

- **Safety ML Ranking Interventions.** Investigated unprecedented social media ML ranking safety issues, architected and launched ML interventions that stabilized the Facebook ecosystem during critical global events. Some of this work was highlighted in WSJ and Time.
- **Convinced** reluctant product organizations to adopt changes.
- **Analysis and Experimentations.** Designed and conducted ML ranking experiments and risk analysis, and presented critical insights weekly to directors from several organizations. These insights led to interventions.
- **Emergency Monitoring:** Built up and co-led a 15 person emergency monitoring team across four organizations to assess and handle unprecedented risks during critical times. This is a first at Facebook and in the industry.

Meta - News Product Team — Tech Lead 2017 - 2019

- Architected Instant Articles mobile infrastructure (adopted by 10,000+ publishers [*](#)), built the iOS team in the News org from the ground up, and established experimentation standards across multiple product teams (Local News, News Credibility, Breaking News).

- Co-founded the Social News team, leading zero-to-one experiments that validated social interaction around news and doubled the team's size. Collaborated cross-functionally with design, PM, data science, and research to secure leadership investment.
- Mentored junior engineers and taught iOS development through Facebook iOS Academy

Meta - App Infrastructure Team — Software Engineer IC 2015 - 2017

- Core contributor to Meta's industry-leading full-stack mobile framework (React-inspired), supporting billions of daily user interactions.

ADDITIONAL EXPERIENCE

Zitara Technologies — Head of Platform Software 2022 - 2023

- Led a product team to launch the company's first web product
- Built up the software department from 4 to 10 members

Zitara Technologies — Staff Software Engineer 2021 - 2022

- Built a scalable model research and validation Infrastructure for battery models on AWS (EC2, S3, Python, Pandas, SQLAlchemy).
- Accelerated SoC/SoH modeling via distributed simulation and automated data workflows.
- Collaborated with researchers to prototype tools, visualize model performance, and ensure reproducible experiments.
- Led codebase design and coordination across research and engineering teams.

EDUCATION

Rice University — Bachelor of Science in Computer Science

Applied Math Focus

PROGRAMMING AND NATURAL LANGUAGES

Python, PHP, Objective-C, Javascript, C++, C; English, Mandarin Chinese

FRAMEWORKS & TOOLS

PyTorch, AWS, EC2, S3, Terraform, Flask, SQLAlchemy, Boto3, Pandas, Numpy, Hive, Jupyter, Einops, Git, Mercurial, Github Actions, React, Bootstrap, etc

HOBBIES AND INTERESTS

Landscape Painting, Outdoor Rock Climbing, Competitive Powerlifting