

# Wen Xing

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

---

## SUMMARY

**Former Meta Integrity Tech Lead and Staff ML engineer, now conducting active research in AI Safety.** Throughout my career, I've prioritized making technology transparent and safe—from global-scale social media harm mitigation to LLM monitorability research.

## PUBLICATIONS

- **Frontier Models Can Take Actions at Low Probabilities** A. Serrano\*, W. Xing\*, et al. · ICML 2026 (under review) · Co-first author
- **Can Reasoning Models Obfuscate Reasoning? Stress-testing CoT Monitorability** A. Zolkowski\*, W. Xing\*, et al. · arXiv 2510.19851 · Accepted at NeurIPS 2025 FoRLM Workshop · Co-first author
  - 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** W. Xing\*, P. Moodley\* · Research blog post · Co-first author

## 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 2 researchers on a project I proposed: *Disentangling Instruction-Following from Strategic Obfuscation in Chain-of-Thought Reasoning*

**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 a WSJ series and a NYTimes article.
  - 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

- **Core Engineering and Leadership**

- Architected and continuously improved Instant Articles mobile infrastructure, focusing on product performance, code reusability, and developer experience (Instant Articles was then adopted by over 10,000 publishers worldwide [\\*](#)).
- Built an iOS team in the News org from the ground up, supported multiple product teams inside the News org including Local News, News Credibility, Breaking News, and Instant Articles. Established standards of experimentation process across the News org.
- **Product Leadership**
  - Co-founded Social News team. Built and grew a team to focus on social interactions, including reactions and conversations around news, and sharing news.
  - Collaborated with design, product management, data science, and research to convince News leadership that social interaction around news is worth investing in
  - Planned roadmap for Social News zero-to-one stage, and led experiments that validated the value of this investment. As a result, we doubled the size of the team.
- **Mentorship and Recruiting**
  - Mentored junior engineers on the News team. Participated in company and team recruiting. Taught iOS classes as part of Facebook iOS Academy.

## 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

Competitive Powerlifting, Outdoor Rock Climbing, California Impressionist Painting