Xuming (Mac) Huang

xuming@cs.wisc.edu | +1 (608) 286-4006 | https://xuming.ai

EDUCATION

University of Wisconsin–Madison **Madison, WI**
B.S. with Honors in Computer Sciences (Programming Abstractions, Operating Systems) *Jan 2025 – Present*
B.S. in Computer Engineering (Digital Systems, Computer Architecture)

GPA: 4.0/4.0

Stanford University **Stanford, CA**
Visiting Scholar (Computer Organization & Systems, Design & Analysis of Algorithms) *Jun 2025 – Aug 2025*

GPA: 4.0/4.0

University of Shanghai for Science and Technology **Shanghai, China**
Computer Sciences (Machine Learning, Artificial Intelligence) *Sep 2023 – Dec 2024*

Major GPA: 4.5/4.5

PROFESSIONAL SKILLS

Languages: English (Fluent), Mandarin (Native), Japanese (Awful)
Programming: Python, C/C++, Verilog, Assembly, Java, JavaScript, Swift
Tools: Git, Linux, FastAPI, Docker
Research: LLMs, Computer Systems/Architectures

WORK EXPERIENCE

Microsoft **Remote**
LLM Research Intern *Dec 2024 – Jan 2025*

- Authored comprehensive SOTA survey analyzing 15+ NER and multimodal sentiment models, creating comparative framework adopted by the research team for model selection

Apple **Remote**
NLP Algorithm Intern *Oct 2024 – Nov 2024*

- Engineered real-time multilingual translation system using optimized Transformer architecture, supporting Chinese and English with <100ms latency for 95% of requests

Cool AI **Shanghai**
Technical R&D, Product Dev & Ops Intern *Jul 2024 – Sep 2024*

- Architected scalable FastAPI backend for LLM integration, handling 1,000+ concurrent requests and reducing API response time from 3s to 800ms
- Deployed Prompted Agents for AI-Hub platform with serving 20+ enterprise clients, generating $10K in new revenue within first month of launch

RESEARCH EXPERIENCE

LinuxGuard: AI-Powered Kernel Security Analysis **Madison, WI**
Research Assistant (Supervised by Prof. Remzi Arpaci-Dusseau and Vinay Banakar) *Jan 2025 – Present*

- Designed AI pipeline processing 50,000+ Linux kernel commits to automatically generate static analyzers, reducing manual vulnerability detection time from weeks to hours
- Applied ML clustering (K-means, TF-IDF) on 10,000+ code patterns to extract vulnerability anti-patterns, resulting in detection framework adopted by 3 major Linux distributions

Domain Adaptation for Agricultural Image Analysis **Shanghai**
Research Assistant (Supervised by Prof. Xing Hu) *Oct 2024 – May 2025*

- Benchmarked 6 state-of-the-art DA methods on plant disease (56% accuracy improvement), yield prediction (0.85 R^2), and land extraction tasks (55.76% IoU), identifying optimal strategies for cross-sensor and cross-regional agricultural applications