James Shiffer

jamesnshiffer@gmail.com | linkedin.com/in/jamesnshiffer | jshiffer.xyz

Education

Sep 2020 – Jun 2026 (expected) **University of California, Los Angeles**
Los Angeles, CA

- Pursuing **Master of Science, Computer Science** (GPA: 3.9/4.0)
- Awarded **Bachelor of Science, Computer Science and Engineering** in June 2024 (GPA: 3.8/4.0)
- **University of California Regents Scholar:** awarded to the top 1.5% of incoming UCLA students.
- **National Merit Scholar:** awarded to the top 0.5% of eligible college-bound students in the U.S.
- President of the Linux Users Group at UCLA

Work Experience

Jan 2025 – Present **Graduate Research Assistant**
Misinformation, AI & Responsible Society (MARS) Lab, UCLA

- ***X-Teaming:*** First author on paper submitted to COLM 2025. Developed robust and generalizable black-box multi-turn LLM jailbreaking techniques that achieve state-of-the-art performance against modern commercial and open-weight models like GPT-4o, Claude, Llama, and Gemini.
- ***AI Debate:*** Fifth author on paper submitted to NeurIPS 2025. Studied the efficacy of AI-human collaboration at evaluating factuality claims in consultancy and multi-agent debate settings.
- Upcoming projects: Scientific Claim Validator, Reward Modeling for Scalable Oversight of Multimodal LLMs.

Oct 2020 – Present **Software Development Team Lead**
Electron Losses & Fields Investigation (ELFIN), UCLA
ELFIN is a student-run lab funded by the NSF and NASA to develop satellites for studying space weather.

- ***CEPHEIDS Mission:*** Upgrading FPGA logic in the Energetic Particle Detector Analog-Digital Converter (D1) and Instrument Data Processing Unit (IDPU). Developing embedded software for ADSP-21489 chip to process FFTs of High Frequency Loop Magnetometer (HFLoop) data at extremely high throughput. These next-generation components improve the detection of solar flares and geomagnetic storms.
- ***Attitude Determination and Control Systems (ADCS):*** Led a team of four students to build Attitools, a Python FastAPI service for managing attitude simulation jobs. Oversaw the complete workflow where satellite operators can curate calculated attitudes through a web application.
- ***ELFIN Mission:*** Maintained core Python libraries for ground support equipment. Added support for data collection during periods when the ELFIN satellites overlapped. Optimized the data pipeline, resulting in an eightfold increase in the amount of data downloaded.

Jun 2024 – Sep 2024 **Flight Software Intern**
Millennium Space Systems, El Segundo, CA

- Developed a Vorago microcontroller bootloader for peripherals and subsystems to be included in future generations of Millennium satellites. Added Golay error detection & correction (EDAC), allowing the bootloader to fully recover from boot image errors of up to 12.5%.
- Improved test coverage, continuous integration, and streamlined build processes across multiple flight software repositories.

Jun 2023 – Sep 2023 **Flight Software Applications Intern**
NASA Jet Propulsion Laboratory, Pasadena, CA

- Contributed to the Vision Compute Element Flight Software (VCEFSW) of the Sample Retrieval Lander for the upcoming Mars Sample Return mission. Added commands to leverage new hardware features, including toggling SUROM write protection and managing data products on NAND flash.
- Improved existing Python tools used for integration and automated hardware-in-the-loop (HIL) testing. Refactored test scripts from the Mars 2020 mission to remove references to obsolete surface (rover) modules. Achieved nearly 50% unit test coverage for the Computer Vision Accelerator

Card Driver (CVACDRV) module in C language flight code, where no unit tests had existed before.

Jun 2021 – Sep 2022 **Software Engineering Intern**
Apton Biosystems (acquired by Pacific Biosciences), Pleasanton, CA

- Singlehandedly accelerated base calling for DNA sequencing by running a boosted trees ML model on FPGAs instead of GPUs. Integrated FPGA workers into the existing C# data pipeline, resulting in a fourfold increase in processing speed to over 40 million samples per sec.
- Designed and built a web analytics portal using JavaScript (Vue) to automatically query, correlate, and graphically present DNA sequencing data, thus eliminating a previously manual and labor-intensive process and allowing scientists to quickly analyze their experimental results.

Projects

May 2025 **Highway Analytics and Navigation Extractor (HANE)**

- A tool for analyzing POV videos of motorcycle rides using computer vision algorithms.
- Written in Python with OpenCV. Uses YOLO11 object detection model to identify regions of interest. Supports traditional optical character recognition (OCR) as well as modern vision language model (VLM) approaches to read dashboard data.
- Produces plots of motorcycle speed over time. Logs potentially important events such as other vehicles and traffic signals.

Mar 2024 **MikuAI**

- A Discord chatbot that participates in group chats as a funny, sarcastic companion. Uses a version of the Llama 3 language model fine-tuned on my own conversation data.
- Uses AI voice cloning to mimic the voice of Hatsune Miku from Crypton Future Media.

Jun 2023 **Kagamine Len Sleeve**

- Designed and built a wearable live audio spectrum visualizer using a FLORA microcontroller, microphone amplifier, and custom-built LED matrix. Programmed in Processing, with an FFT library written in AVR assembly.
- Fine-tuned parameters to reduce background noise and assign relative weights to frequency bins, resulting in a more natural-looking spectrum.

2018 – 2023 **Atmos**
atmos.warplight.dev

- Developed a new social media platform inspired by Google+, using PHP (Laravel) to build the backend and JavaScript (SvelteKit) to build a single-page application (SPA) frontend.
- Scaled to accommodate over 1,900 users, 50,000 posts/comments, and 16,000 media uploads.
- Delivered features such as responsive design, internationalization support, post recommendation engine, image and video uploads, lazy-loading of images, infinite scrolling, privacy customization, communities (user groups) and collections (post categorization), server-side caching, and push notifications to improve UX and retention for desktop and mobile users.
- Made admin tools for banning users, deleting content, accessing site analytics, and more.

Publications

[1] Rahman, S., Issaka, S., Suvarna, A., Liu, G., **Shiffer, J.**, Lee, J., Parvez, M. R., Palangi, H., Feng, S., Peng, N., Choi, Y., Michael, J., Liwei Jiang, & Gabriel, S. (2025). AI Debate Aids Assessment of Controversial Claims. *arXiv.org.* **(preprint pending; under review at NeurIPS 2025)**

[2] Rahman, S., Jiang, L., **Shiffer, J.**, Liu, G., Issaka, S., Parvez, M. R., Palangi, H., Chang, K., Choi, Y., & Gabriel, S. (2025). X-Teaming: Multi-Turn Jailbreaks and Defenses with Adaptive Multi-Agents. *arXiv.org*. https://arxiv.org/abs/2504.13203 **(co-first authorship; under review at COLM 2025)**

[3] Tsai, E., Palla, A., Norris, A., King, J., Russell, C., Ye, S., Wu, J., Mao, J., Jha, S., Young, C., Wing, G., Lian, K., Szeto, A., **Shiffer, J.**, Sankar, R., Tota, K., Liu, A., Lee, D., Patil, U., & Angelopoulos, V. (2024). Remote sensing of electron precipitation mechanisms enabled by ELFIN mission operations and ADCS. *Advances in Space Research*. https://doi.org/10.1016/j.asr.2024.07.008