Tal Erez

talerez525@gmail.com • (818) 518-8195 • https://www.tal-erez.com/ • linkedin.com/in/tal-erez/ • github.com/notthattal

Education

Duke University April 2025
Master of Engineering — Artificial Intelligence | 4.0 GPA

University of California, San Diego March 2020
Bachelor of Science — Applied Mathematics | 3.65 GPA

Experience

Arthur AI, New York, NY | *Machine Learning Engineer* May 2025 – Present

- Designed an agentic system which allows users to automate the evaluation of traces and run prompt experiments prior to production deployment
- Introduced a sliding window chunking strategy for the prompt injection model, reducing false positives by approximately 70%
- Implemented batching for the toxicity model and expanded the blacklist, improving accuracy by ~10% and tripling token throughput while maintaining a latency of under 3 seconds
- Worked directly with clients to ensure ML solutions aligned with their business goals

Algorithmic Research Group, Durham, NC | *Contract Machine Learning Research Scientist* November 2024 – April 2025

- Expanded the main agent's coding capabilities enabling it to create deep learning models, hyperparameter tune its results and the implement both regularization and optimization techniques.
- Migrated the main agent's architecture to a state machine setup in order to create a more fault-tolerant system which can retry execution failures and evaluate model performance during execution.
- Connected the main agent to the research agent to provide the main agent with a detailed plan of execution based on a literature review and to the judge agent to provide a thorough evaluation of model performance at the end of execution.

Amazon, Remote | *Contract Software Development Engineer* July 2022 – January 2023

- Migrated databases for the Related Accounts Presentation Service (RAPS), the service used to connect one merchant account to another across regions worldwide, preventing loss of data before deprecation of the previous storage service.
- Created alarms in AWS to monitor errors, fatal logs and CPU utilization thresholds for the RAPS service. This resulted in faster response times to service failures.
- Built a filtering method in Ruby to retrieve a merchant's compliance status within a designated timeframe for the internal website used to conduct seller investigations. This new approach eliminated the need to parse through a seller's full history and reduced investigation times.

Shiver Entertainment Inc., Miami, FL | *Software Engineer* July 2021 – July 2022

- Contributed to the development of Hogwarts Legacy in collaboration with Warner Bros. and Avalanche Studios for the PS4, XB1 and Nintendo Switch consoles using Unreal Engine. Selling over 24 million copies globally, Hogwarts Legacy became the #1 best-selling video game of 2023.
- Converted the codebase from Unicode to UTF-8, saving 250 MB of physical used memory as reported by automation tests.
- Altered the multi-thread framework of the game to efficiently pin threads to specific cores in order to reduce idle time on the Nintendo Switch platform. This improved the average frame rate by 10 ms per frame.
- Implemented an LOD system for game visual effects which reduced memory usage by an average of 100 MB and utilized Unreal Engine Python scripting to create an automated way of implementing the new system for all platforms.

Kalloc Studios Inc., Carlsbad, CA | *Software Engineer* October 2020 – July 2021

Research

Mirage AI: Conducted research in generating imperceptible audio perturbations that safeguard artists' work from copyright infringement and unauthorized data scraping, aiming to mitigate the risk of AI mimicry.

Multiple Sclerosis Classification: Designed an early-onset predictive model for multiple sclerosis using data from functional electrical stimulation braces. Data is collected from IMUs for gait event detection and spatial orientation reconstruction, and EMGs.