Maxim Podgore

(310)-283-2055 | maximjulianpodgore@gmail.com | linkedin.com/in/maxim-podgore | github.com/MaximPodgore

EDUCATION

University of California San Diego - 3.82 GPA   Expected March 2026
Bachelor of Science (B.S) in Computer Science

RELEVANT WORK EXPERIENCE

Lead Machine Learning Engineer   January 2024 - Present
UCSD Information Technology Services
- Developed an AI contract review agent leveraging LiteLLM, LLaMA, and Qwen-Embedding for prompt tuning via user feedback. The agent uses a novel Inverse Reinforcement Learning (IRL) pipeline to suggest edits to legal documents and learn from user feedback, driving a 10% increase in contract settlements per month.
- Led a team of 4 student engineers to design the continual-learning aspect of Contract Reviewer after reviewing goals with stakeholders, resulting in a contract reviewer system that increased long-term lawyer satisfaction by 50%

Machine Learning Engineer Intern   July 2025 - September 2025
SingularityNET
- Resolved knowledge conflicts in a Graph RAG system by integrating timestamps and confidence-based sampling into Neo4j-based graph construction, improving consistency and reliability of retrieved information by 30%
- Refactored a cutting-edge benchmark to extend to graph-RAG systems using qwikidata and pandas, testing knowledge recall and ripple effects caused by introducing counterfactual knowledge triplets to the system

Lead Full Stack Engineer   June 2025 - Present
Groundwork Books
- Led the production of an e-commerce site with Next.js, Tailwind and Firebase Auth for user login, integrating Square for live inventory, cart, hosted checkout, and order tracking across 4,000+ SKUs, hosted on Vercel.
- Implemented Redis-backed caching with component-scoped keys and batched inventory calls, and integrated Pinecone semantic search, reducing Square API pressure to obtain sub-200 ms query latency and 90% decrease

Artificial Intelligence Researcher   May 2024 - May 2025
Stanford Novel Computing System's Lab
- Built a C++ static analysis tool for the dReal SMT solver to address floating-point soundness issues, implementing a graph-based LLVM AST solver to validate rounding mode constraints, which detected 200+ problematic lines.

PROJECTS

AI-powered Course Planner | *React, Vite, Uvicorn, FastAPI, Chroma, Selenium, LiteLLM, TailwindCSS*
- Designed and implemented a high-performance Uvicorn + FastAPI server hosting a custom RAG endpoint via LiteLLM and a Chroma vector store, achieving 60% lower latency than standard RAG stacks.
- Developed a RAG pipeline to target requirements and rank courses by fit+rating, reducing planning time by 35%.

RedShift LLM Jailbreak | *Python, PyTorch, CUDA, FastChat, Transformers, HuggingFace*
- Extended the Distract LLMs for Automatic Jailbreak Attack paper by integrating chain-of-thought prompting and FastChat model support, expanding the attack space and increasing coverage by 67%.

TECHNICAL SKILLS

Programming Languages: Python, Java, C, C++, JavaScript/TypeScript, HTML, CSS, SQL
Libraries & Frameworks: Tensorflow, PyTorch, React.js, Next.js, Node.js, Express.js, Tailwind, FastAPI, HuggingFace, OpenCL, Vite, Puppeteer, Jest, Flutter, SciKit, FastChat, Pandas, NumPy, tqdm, Husky, Wandb
Tools & Technologies: Git, Bash, Postman, Docker, AWS (EC2, S3, Elasticache, Lambda), Lucidchart, Atlassian Suite (Jira, Confluence), Redis, Grafana, Figma, Pinecone, Upstash, Neo4j, MySQL, Azure, LiteLLM, SAP HANA

WORK EXPERIENCE

Mobile Development Intern   July 2023 - December 2023
UCSD Information Technology Services
- Developed the UCSD mobile app and fixed 3+ key issues in AWS Lambda regarding parking data endpoints.

Bioinformatics Researcher   January 2023 - May 2024
UCSD School of Medicine, Dr. Allen Ryan's Lab
- Utilized various gene analysis tools to help find 2 new key regulators for immune response to Otitis Media