Yuxuan (Vincent) Ma

809 S Damen Ave, Chicago, IL | 217-904-2163 | mayuxuan135@gmail.com | LinkedIn | GitHub | Portfolio

EDUCATION

University of Illinois Urbana-Champaign, *Urbana, IL* *Aug. 2024 - Dec.2025*
Master of Computer Science
- **GPA: 4.00/4.00**

University of Illinois Urbana-Champaign, *Urbana, IL* *Aug.2021 - May 2025*
B.S. in Electrical Engineering, Minor in Computer Science
- **GPA: 3.65/4.00, Core GPA: 3.857/4.00**

INTERNSHIP EXPERIENCE

Horus Intelligence Solutions **Urbana, IL**
Software Engineering Intern *Jun. 2022 - Aug. 2022*
- Optimized data collection and analysis pipelines for large-scale data to improve the company's ability to deliver actionable insights to clients in a fast-paced and data-driven pattern.
- Re-designed and implemented scalable and efficient web scraping tools (Python), significantly accelerating the data collection process and improving the inefficiencies in data collection due to unstructured web data and inconsistencies in scraping results.
- Enhanced collected data quality with optimized validation pipelines, ensuring that inconsistencies were identified and resolved in the early stages of the data processing workflow.
- Delivered project on time, improved overall data processing efficiency by 10%, reduced scraping error by 30%, and received very positive customer feedback.

PROJECTS

Improve Code Reasoning & Test Gen. in DeepSeekCoder (Python), *Urbana, IL* *Aug.2024 - Dec.2024*
- Improved code reasoning and test generation accuracy by systematically tailoring prompts.
- Classified common problematic prompts with misinterpretation of task structure, incorrect function naming, and poor handling of complex constructs like nested loops or implicit input validations. These overly general or ambiguous prompts limit the accuracy of the model.
- Designed crafted prompts with explicit task instructions, modular structure, and labeled sections for code and expected outputs. Employed iterative experimentation to refine prompts, improving comprehension, and reducing errors in output generation.
- Achieved a 50% improvement in prediction accuracy and a 20% increase in test coverage compared to vanilla prompts. Demonstrated the potential of structured prompt engineering to enhance the reasoning capabilities of AI models in software development contexts.

SteamGenie, *Urbana, IL* *Oct.2024 - Dec. 2024*
An Interactive Steam Game Recommendation App (React/Express.js, JavaScript, MongoDB)
Project Leader
- Created SteamGenie to streamline game discovery for gamers, integrating a conversational LLM for personalized recommendations.
- Overcame challenges in integrating natural language queries with traditional browsing filters by designing dynamic feedback loops to adapt suggestions in real time and fine-tuning the LLM for better query interpretation.
- Used a RESTful API in Express.js for efficient data handling, handled user interactions with React, and utilized MongoDB to store and retrieve game metadata.
- Optimized performance bottlenecks by indexing the MongoDB database, batching data requests, implementing lazy loading for game metadata, and using caching strategies to reduce redundant queries.
- Delivered a user-focused platform with both text-based and interactive filtering options, illustrating versatile full-stack development skills and an emphasis on intuitive design. *Website Link: https://steamgenie.vercel.app/*

Autonomous Multi-Step Robotic Task Execution System (C, MATLAB, Simulink), *Urbana, IL* *Jan. 2024 - May. 2024*
Robot Dynamics and Control Laboratory
- Designed and implemented the Autonomous Multi-Step Robotic Task Execution System for a CRS robot arm, enabling precise navigation, obstacle avoidance, and task-specific operations.
- Created to tackle real-world robotic challenges by developing a system capable of performing complex multi-step tasks, such as peg insertion, weight application, and precise path following, within a competitive time frame.
- Faced difficulties in trajectory optimization, maintaining stability during force-sensitive tasks (e.g., controlled egg pressing), and ensuring precise execution of depth-specific maneuvers in a zig-zag path while avoiding obstacles.
- Implemented task-space PD control with friction compensation, utilizing the Jacobian transpose for accurate force and torque calculations. Developed trajectory planning algorithms in MATLAB and employed real-time feedback to dynamically adjust PID control gains, ensuring smooth and precise execution.
- Delivered a robust, high-performing robotic system capable of executing all assigned tasks efficiently. Achieved significant precision in navigating obstacle courses, inserting pegs, and applying controlled forces, showcasing advanced skills in robotic control, modeling, and real-time systems engineering.

Dynamic 'Flappy Bird' on FPGA (SystemVerilog & C++), Digital Systems Laboratory*, Urbana, IL* *Aug. 2023 - Dec. 2023*

- Created a hardware-based version of 'Flappy Bird' to explore how classic games can leverage real-time performance on FPGA rather than software emulation alone.
- Encountered frame rendering delays and memory constraints, which required careful module design. Integrated a custom VGA controller, color mapper, and pillar dynamics to maintain smooth gameplay at variable speeds.
- Addressed real-time challenges by splitting the system into parallel hardware modules (e.g., bird movement, collision detection, scoreboard tracking), ensuring consistent frame rates and responsive controls.
- Implemented adjustable game speed, background schemes, and an invincible mode to enrich user interaction and a scoring function and start/end screens for a polished, arcade-like experience.
- Delivered a fully functional game demonstrating comprehensive skills in FPGA programming, digital design, and user interface development—highlighting the potential of hardware-based solutions for interactive applications.

Bitcoin Blockchain Analysis, *Urbana, IL* *Mar. 2025 - Mar. 2025*

- Leveraged Blockchain.info APIs in Java to retrieve and examine block data, including block size, previous block hash, and transaction counts, ensuring accurate extraction of key blockchain metrics.
- Implemented detailed transaction analysis by identifying transactions with the most inputs and outputs, extracting Bitcoin addresses, and determining the largest single transaction in terms of Satoshi received.
- Analyzed coin base transactions and computed the total Satoshi generated within a block, deepening the understanding of Bitcoin's monetary policy and block reward mechanism.
- Designed and executed an algorithm to cluster Bitcoin addresses based on joint control heuristics, effectively grouping addresses likely managed by the same user and generating a comprehensive user graph for further network analysis.

Reproducibility Research for Research Software (Docker, Python, YAML), *Urbana, IL* *Aug. 2024 – Dec. 2024*

- Developed reproducible workflows to replicate research artifacts from top-tier computer science publications, addressing challenges like software decay, incomplete documentation, and unavailable dependencies.
- Created Dockerfiles and YAML metadata files for 26 research papers, targeting artifacts with no prior reproducibility attempts.
- Solved build issues systematically by debugging incomplete instructions, outdated dependencies, and sourcing historical software versions using archive.org.
- Implemented functional validation tests ensuring consistency and accuracy relative to original research claims, demonstrating strong skills in dependency management and software archaeology.
- Proposed practical solutions for enhancing software reproducibility, standardizing workflows utilizing Docker and Ubuntu, significantly mitigating software decay risks.

Regional Video Games Sales Analysis (Python, Logistic Regression, LASSO), *Urbana, IL* *Jan. 2022 – May 2022*

- Conducted comprehensive analysis of global video game sales, examining the influence of genres, platforms, and regional preferences on market performance.
- Addressed significant data inconsistencies and multicollinearity issues through extensive preprocessing, dataset filtering, and feature selection.
- Applied LASSO regression for feature selection and logistic regression for sales prediction and genre classification, validating model performance using hypothesis testing and ROC curve analysis.
- Delivered actionable insights into regional market trends, highlighting genre popularity differences between North America and Japan, achieving an AUC of 0.64 for role-playing game classification.
- Demonstrated advanced proficiency in data cleaning, statistical inference, and predictive modeling.

Kernel-Level Optimization for ML Operators (C++, CUDA, OpenMP), *Urbana, IL* *Jan. 2025 – Feb. 2025*

- Achieved significant performance improvements by manually optimizing kernels for matrix multiplication and 2D convolution across CPU, GPU, and AWS Tranium accelerators, reflecting key ML workloads.
- Enhanced CPU performance by applying strategic loop ordering, tiling, and OpenMP parallelization with vectorization, achieving up to 14.9× speedup for 1000-sized matrices compared to the naive implementation.
- Dramatically improved GPU efficiency using CUDA, with parallelization and shared-memory tiling strategies yielding a peak speedup of 45,216× for matrix size 1000 over the naive baseline; maintained runtimes below 700 ms for matrices as large as 10,000×10,000×10,000.
- Optimized convolution kernels on AWS Tranium by effective tiling, weight preprocessing, and masked memory operations, achieving a matrix floating-point utilization (MFU) of approximately 65% for bfloat16 precision.
- Conducted thorough scalability studies highlighting optimized GPU kernels' near-linear performance, contrasting CPU kernels' significant runtime increases (e.g., around 80,000 ms for 10,000-sized matrices), underscoring the advantages of hardware-specific optimizations.

Tensor Graph Rewriting in XLA Compiler (C++, Bazel, XLA HLO), *Urbana, IL* *Feb. 2025 – Mar. 2025*

- Developed advanced tensor graph rewrite rules within Google's XLA compiler, significantly enhancing tensor computation efficiency for machine learning workloads.
- Implemented algebraic simplifications including distributing element-wise addition over scalar division, and fusion of matrix multiplications and element-wise operations based on TASO and DNNFusion research.
- Conducted extensive unit testing using Google Test, ensuring correctness of algebraic simplifications and demonstrating

accurate graph transformations.
- Achieved successful validation of rewrite rules through XLA's HLO computational graph tests, confirming semantic correctness and consistency across tensor transformations.
- Improved the algebraic simplifier by adding comprehensive rewrite rules, contributing to tensor compiler optimizations that underpin modern deep learning frameworks.

CryptoToolkit (Python), *Urbana, IL* *Feb.2025 - Mar. 2025*
https://github.com/YuxuanMa-sys/Cryptography
- Developed a comprehensive cryptography toolkit demonstrating various encryption techniques and cryptographic primitives.
- Implemented symmetric encryption algorithms such as AES and ChaCha20, and asymmetric encryption using RSA.
- Added functionalities for digital signatures, secure key derivation, password hashing, file encryption/decryption, and message authentication using HMAC.
- Ensured user-friendly command-line interface for seamless encryption and decryption processes, enhancing data security practices.