

Feng-Chieh Ku

✉ henryku90911@gmail.com 📞 +886-905-275-780

EDUCATION

National Taiwan Normal University (NTNU), Taipei, Taiwan
BS in Electrical Engineering

Sep 2021 – Jun 2025

- **Cumulative GPA:** 4.17/4.3; Ranked 3/63 from Fall 2021 to Jan 2025
- **Coursework:** Machine Learning(A), Neural Network(A+), Computer Vision(A+), Reinforcement Learning (A+), Data Structure(A+), Intelligent Humanoid Robot(A+), Algorithm(A-)
- **Student Project Presentation and Competition Award:** 1st/30, Fall, 2023

University of Michigan, Ann Arbor, MI
MS in Robotics

Expected Sep 2026

PROFESSIONAL EXPERIENCE

Ret[AI]ling Data Co., Ltd., Taipei, Taiwan
AI Engineer Intern

Aug 2024 – Apr 2025

VoyageMind: Multi AI Agent Tourism Assistant

- Building a two-tier StateGraph with MemorySaver checkpointer to persist node states while supervisors dynamically route messages among booking, personal, restaurant, attraction, and hotel agents
- Wrapping each agent in a ToolNode runner with custom Guard tools and RunnableLambda fallbacks to support auto-retries, LLM hesitation detection, and unified error feedback
- Injecting real-time context via Google Maps Directions, CWA weather, and TDX tourism APIs, while enriching responses with TavilySearchResults for updated routes and POI details
- Persisting conversations using a TTL-managed Redis-based ConversationStore, and storing semantic records in MongoDB Atlas Vector Search with OpenAI embeddings
- Deploying the Flask API and Redis Stack via Dockerfile and Docker Compose on a shared bridge network for seamless development-to-production consistency

A government project on building an AI agent for Ximen district's tourism development

- Establishing a RAG system that includes self-reflection: Retrieval Grader, Hallucination Grader, and Answer Grader to reduce ineffective information retrieval, LLM hallucinations, and improve answer accuracy, while using the Tavily API to enable web search capabilities over traditional RAG
- Implementing a RAG evaluation framework (RAGAS) to assess the relevance, correctness, and semantic similarity between answers generated by the RAG model, their context, the questions, and the true answers
- Expanding the Ximen business district dataset by using Google Maps API and web crawling, and storing the data in a MongoDB database using embeddings
- Using data augmentation techniques to address data insufficiency, along with an SBERT-based semantic system to ensure the quality of the augmented data

A government project on fine-tuning models of LLM assistant for Ministry of Digital Affairs

- Combining LoRA and 4-bit quantization to significantly reduce parameter counts and memory requirements for fine-tuning large models while maintaining high efficiency
- Implementing a sliding window approach to create contextually coherent samples and handles structured text like tables and LaTeX formulas to enhance the model's understanding
- Configuring the NCCL distributed environment and enabling mixed-precision computation to improve efficiency and resource utilization in multi-GPU training
- Using LlamaParse to parse long PDF files, extracting text, tables, mathematical formulas, and images, and converting them into Markdown format

Researching InfoSec On-Premises Generative AI rapid deployment solution for automatic high-quality copywriting, content creation, question answering, translation, and customer service

- Developing an enterprise-level RAG system that allows users from different regions to access heterogeneous data sources for achieving cross-regional data retrieval and AI-generated responses
- Establishing full AI model fine-tuning and application pipeline of LLMs for creating enterprise-specific generative AI chatbots that allow businesses to automatically adapt to different intelligent scenarios
- Building a dynamic database management system that automatically updates from sources based on regional needs, supporting continuous model optimization and updates, ensuring LLMs remain efficient and accurate

RESEARCH & PROJECTS at NTNU

Energy-Saving Management Technology for Smart Home Appliances for Panasonic

Jul 2024 – Dec 2024

- Utilized an LSTM model to analyze historical temperature and humidity data with various environmental and equipment parameters, predicting future trends to adjust air conditioning and dehumidifier settings in advance for energy savings
- Integrating fuzzy control and the PMV (Predicted Mean Vote) to enable dynamically adjustment on air conditioner and dehumidifier settings according to real-time conditions, preventing excessive energy use from overly low settings
- Developing a personalized model from the user's historical temperature and humidity preferences, combined with LSTM predictions, to dynamically adjust settings, optimizing comfort and providing energy-saving recommendations

Automated Intelligent Billiard Robot

Aug 2023 – June 2024

- Used xArm-Python-SDK to control initialization, error handling, and movement execution based on calculated positions and angles of 6 Degree-of-Freedom manipulator
- Detected real-time image with Python OpenCV and emphasized perspective transformation to capture billiard table and converted into coordinate system designed; Enhanced object recognition to identify ball types and positions with YOLOv8
- Utilized rule-based control to make striking decisions and achieved an accuracy of over 85%
- Presented machine learning, computer vision, and reinforcement learning technique in International Conference on System Science and Engineering 2024 hosted by National Yang Ming Chiao Tung University
- Granted 2000 USD by Undergraduate Research Scholarship from Ministry of Science and Technology

International Robotics Festival RoboScience Hackathon

Oct 2023 – Dec 2023

- Utilized the PID controller to adjust the robot's movement direction for quick response, stability, and minimizing overshoot and oscillation
- Integrated the proportional and differential control of the heading to correct the robot's trajectory accuracy in real-time by gradually increasing the rotation increment to reach the target
- Adjusted the rotation increment parameter to precisely control the robot's movement path, enabled smooth and continuous movement along designated path
- Won first place out of fourteen teams in the simulation league HuroCup ELSIROS Marathon

Gesture-Controlled Contour Interaction System

Sep 2023 – Nov 2023

- Combined MediaPipe and Python OpenCV and applied with image processing capabilities for achieving real-time advanced hand landmark detection
- Calculated hand angles by using a custom function to determine the orientation and position of fingers based on detected hand landmarks; achieved nearly 100% accuracy
- Implemented Python OpenCV to achieve contour finding for analyzing and detecting the dice's position and angle in real-time
- Manipulated dynamic annotations with OpenCV to recognize hand gesture and overlaid respective dice numbers in projection

TECHNICAL SKILLS

Programming Languages: C/C++, Python, Verilog, HTML

Frameworks/Tools: PyTorch/TensorFlow, YOLO, Hugging Face, Ollama, LangChain/LangGraph