

William Nguyen

Palo Alto, CA 94303, USA

nguyennm1024@gmail.com | [LinkedIn](#) | [GitHub](#) | [Google Scholar](#) | nguyennm1024.github.io | +1 (585) 540-6264

PROFESSIONAL SUMMARY

AI/ML engineer with 6+ years spanning applied computer vision research and production machine learning, currently Senior Applied Scientist at Aitomatic leading production LLM and agent systems. Ships reliable on-prem agents with harnesses and guardrails, post-trains compact models to frontier-comparable quality on target use cases, and builds platforms that cut agent development cycle time by 5×. Domain LLMs adopted by Tokyo Electron and Furuno; first-author publications at CVPR and NAACL; 3 US patent applications in domain-expert agents.

WORK EXPERIENCE

Senior Applied Scientist, Aitomatic, Inc., Palo Alto, CA 08/2024 – present

- **Domain LLMs in production:** Led development of SemiKong and Lllamarine, open-source LLMs for the semiconductor and maritime domains, **adopted by Tokyo Electron and Furuno** respectively (first author, Lllamarine; second author, SemiKong).
- **Deployment impact:** Semiconductor root-cause-analysis deployment cut equipment troubleshooting time by 30%; maritime navigation assistant passed all regulation-compliance checks in complex navigation evaluations.
- **Agents (ProSEA):** Developed ProSEA, an agent framework that learns from failure cases to self-improve; **outperformed competing providers' solutions** on real-world device troubleshooting in customer evaluations, as reported in the ProSEA paper (see Publications, page 2).
- **Physical AI agents:** Built a Building Management System (BMS) agent that runs fully on-prem on workstation hardware, owning the reliability stack end-to-end from agent harness to guardrails; its context-aware energy optimization reduced monthly energy consumption by 10%.
- **On-prem post-training:** Led post-training of compact models (Gemma 4 E4B, Nemotron 4) with SFT, DPO, and GRPO to adapt small LLMs to industrial building-ops use cases, **matching frontier closed-model (GPT-5.4) quality on these tasks** in internal evaluations while running entirely on local hardware.
- **TRAIN platform:** Led technical development of TRAIN, a platform for building domain-specific agents with minimal setup; designed and built its automated data curation and expert-knowledge synthesis loop, cutting agent development cycle time by 5×, knowledge-update time by 2×, and letting the team onboard 20% more customers.
- **Research & IP:** First author of the ProSEA and Lllamarine papers; second author of SemiKong; technical lead and 2nd-listed co-inventor on 2 of 3 US patent applications (page 2).

Research Assistant, University of Rochester, Rochester, NY 06/2022 – 06/2024

- **Multimodal LLM systems:** Proposed object-state captioning and built a compact multimodal LLM for QA, dialogue, and reasoning, retaining ~90% of then-frontier GPT-4V quality on QA, dialogue, and reasoning evaluations at substantially lower model size and latency.
- **Scene text & video:** Improved scene-text spotting by 2–4% absolute over prior SOTA across standard benchmarks; built task-guidance algorithms from egocentric and instructional videos. ([Demo video](#)).

AI Research Engineer, VinAI Research (now Qualcomm AI Research), Hanoi, Vietnam 01/2022 – 06/2022

- **Face recognition:** Built unified models for normal, masked, and extreme-pose faces; improved masked recognition by 18% via synthetic data; distilled lightweight models for on-edge deployment.

AI Research Resident, VinAI Research (now Qualcomm AI Research), Hanoi, Vietnam 12/2019 – 12/2021

- **Scene text systems:** Dictionary-guided recognition surpassed SOTA by 3–5%; built traffic-sign spotting with faster inference and an annotation method that cut labeling cost by 50%.

TECHNICAL SKILLS

Programming	Python, C/C++, Bash, SQL
Agents & Serving	vLLM, llama.cpp, Ollama, LangChain, RAG (LlamaIndex, FAISS), harnesses & guardrails
Post-training / Fine-tuning	SFT, DPO, GRPO, LoRA/QLoRA
ML Frameworks	PyTorch, HuggingFace, TRL, TensorFlow, Scikit-learn
Data & Infra	Pandas, NumPy, PySpark, GCP, Docker, Slurm, Git
Computer Vision	OpenCV, Detectron2, CLIP, LLaVA, Tesseract
Prototyping	Streamlit, Flask, React

OPEN-SOURCE PROJECTS

AgentHub (github.com/agentplug/agenthub): personal open-source Python SDK and CLI for running AI agents, built around isolated per-agent environments that prevent dependency conflicts and a unified interface across OpenAI, Anthropic, and Google APIs plus local models via llama.cpp, Ollama, and LM Studio.

Tapestry (The AI Alliance, github.com/The-AI-Alliance/tapestry): built and merged a socio-cultural alignment pipeline that adapts LLMs to Vietnamese cultural values via synthetic data, LoRA post-training, and model-soup fusion, reducing Inglehart–Welzel cultural distance by ~45% with no model capacity loss.

EDUCATION

University of Rochester, Rochester, NY 08/2022 – 05/2024

M.S. in Computer Science

University of Engineering and Technology, Hanoi, Vietnam 08/2016 – 08/2020

B.S. in Computer Science

PROFESSIONAL SERVICES

Reviewer: WACV 2022, CVPR 2023, CVPR 2024, ACM MM 2024, AAAI 2025, CVPR 2025.

Invited Speaker: VinAI Research Workshop 2021, OSAI4MU Workshop (AAAI'25).

Organizer: Vietnamese Scene Text Recognition Challenge 2021.

SELECTED PUBLICATIONS

Published as William Nguyen and Nguyen Nguyen; patent applications filed under legal name Manh-Nguyen Nguyen.

1. **William Nguyen**, Vinh Luong, Christopher Nguyen, “*ProSEA: Problem Solving via Exploration Agents*”, preprint, 2025.
2. **William Nguyen**, An Phan, Konobu Kimura, Hitoshi Maeno, Mika Tanaka, Quynh Le, William Poucher, Christopher Nguyen, “*Llamarine: Open-source Maritime Industry-specific Large Language Model*”, The 39th Annual Conference of the Japanese Society for Artificial Intelligence, 2025.
3. Christopher Nguyen, **William Nguyen**, Atsushi Suzuki, Daisuke Oku, Hong An Phan, Sang Dinh, Zooey Nguyen, Anh Ha, Shruti Raghavan, Huy Vo, Thang Nguyen, Lan Nguyen, Yoshikuni Hirayama, “*SemiKong: Curating, Training, and Evaluating A Semiconductor Industry-Specific Large Language Model*”, OSAI4MU Workshop at AAAI, 2025.
4. **Nguyen Nguyen**, Jing Bi, Ali Vosoughi, Yapeng Tian, Pooyan Fazili, Chenliang Xu, “*OSCaR: Object States Captioning and State Changes Representation*”, NAACL, 2024.
5. **Nguyen Nguyen**, Thu Nguyen, Vinh Tran, Triet Tran, Thanh Duc Ngo, Thien Nguyen, Minh Hoai, “*Dictionary-guided Scene Text Recognition*”, IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.

Full publication list (9 papers) available on [Google Scholar](#).

PATENT APPLICATIONS

1. Christopher Nguyen, **Manh-Nguyen Nguyen**, et al., “*Delivering Domain-Expert Agents and Models Using Synthetic Knowledge*”, US Patent App. 63/726,322, 2024.
2. Christopher Nguyen, **Manh-Nguyen Nguyen**, et al., “*Delivering Domain-Expert Agents for Improving Problem-Solving*”, US Patent App. 63/721,419, 2024.
3. Christopher Nguyen, The Vinh Luong, Xuan Sang Dinh, Zooey Nhu-Quynh Nguyen, Shruti Raghavan, **Manh Nguyen Nguyen**, et al., “*Domain-Aware Neurosymbolic Agents For Improving Problem-Solving Accuracy And Consistency*”, US Patent App. 63/696,337, 2024.