Parinthapat Pengpun

Pittsburgh, PA | ppengpun@andrew.cmu.edu | GitHub | Google Scholar | Hugging Face | LinkedIn

Education

Carnegie Mellon University, School of Computer Science — Aug 2025 – present
B.S. in Artificial Intelligence — Pittsburgh, PA

Bangkok Christian International School — 2025
High School Diploma, **Valedictorian** — Bangkok, Thailand
SAT: 1550/1600 (2024) | IELTS Academic: 8.5/9.0 (2024)

Publications

* Equal contribution. Conference, workshop, and preprint status indicated for each work.

[1] **Think Before You Link: Rarity, Reasoning, and Retrieval in Multilingual Entity Linking**
Parinthapat Pengpun, Simran Khanuja, Graham Neubig.
EMNLP 2026 Main Conference, accepted. arXiv:2609.10745.

[2] **When Better Teachers Don't Make Better Students: Revisiting Knowledge Distillation for CLIP Models in VQA**
Pume Tuchinda, **Parinthapat Pengpun**, Romrawin Chumpu, Patomporn Payoungkhamdee, Sarana Nutanong, Peerat Limkonchotiwat.
MRAC @ ACM Multimedia 2026, accepted workshop paper. arXiv:2511.17886.

[3] **On Creating an English-Thai Code-switched Machine Translation in Medical Domain**
Parinthapat Pengpun*, Krittamate Tiankanon*, Amrest Chinkamol, Jiramet Kinchagawat, Pitchaya Chairuengjitjaras, Pasit Supholkhan, Pubordee Aussavavirojekul, Chiraphat Boonnag, Kanyakorn Veerakanjana, Hirunkul Phimsiri, Boonthicha Sae-jia, Nattawach Sataudom, Piyalitt Ittichaiwong, Peerat Limkonchotiwat.
Findings of EMNLP 2024, pp. 6055–6073.

[4] **Seed-Free Synthetic Data Generation Framework for Instruction-Tuning LLMs: A Case Study in Thai**
Parinthapat Pengpun, Can Udomcharoenchaikit, Weerayut Buaphet, Peerat Limkonchotiwat.
ACL 2024 Student Research Workshop, pp. 445–464.

[5] **SICAR at RRG2024: GPU Poor's Guide to Radiology Report Generation**
Kiartnarin Udomlapsakul*, **Parinthapat Pengpun***, Tossaporn Saengja, Kanyakorn Veerakanjana, Krittamate Tiankanon, Pitikorn Khlaisamniang, Pasit Supholkhan, Amrest Chinkamol, Pubordee Aussavavirojekul, Hirunkul Phimsiri, Tara Sripo, Chiraphat Boonnag, Trongtum Tongdee, Thanongchai Siriapisith, Pairash Saiviroonporn, Jiramet Kinchagawat, Piyalitt Ittichaiwong.
23rd Workshop on Biomedical Natural Language Processing (BioNLP @ ACL 2024), pp. 635–644.

[6] **MAHASAMUT: AI-Powered Thai Tourism using Multimodal Agents**
Nongnuch Ketui, **Parinthapat Pengpun**, Konthee Boonmeeprakob, Pitikorn Khlaisamniang, Thepchai Supnithi.
LaCATODA @ PRICAI 2024. CEUR Workshop Proceedings, vol. 3862, pp. 56–65.

[7] **Typhoon 2: A Family of Open Text and Multimodal Thai Large Language Models**
Kunat Pipatanakul, Potsawee Manakul, Natapong Nitarach, Warit Sirichotedumrong, Surapon Nonesung, Teetouch Jaknamon, **Parinthapat Pengpun**, Pittawat Taveekitworachai, Adisai Na-Thalang, Sittipong Sripaisarnmongkol, Krisanapong Jirayoot, Kasima Tharnpipitchai.
Technical report, arXiv:2412.13702, 2024.

[8] **Can General-Purpose Large Language Models Generalize to English-Thai Machine Translation?**
Jirat Chiaranaipanich, Naiyarat Hanmatheekuna, Jitkapat Sawatphol, Krittamate Tiankanon, Jiramet Kinchagawat, Amrest Chinkamol, **Parinthapat Pengpun**, Piyalitt Ittichaiwong, Peerat Limkonchotiwat.
GenBench @ EMNLP 2024, non-archival. arXiv:2410.17145.

[9] **Cross-Lingual Data Augmentation For Thai Question-Answering**
Parinthapat Pengpun, Can Udomcharoenchaikit, Weerayut Buaphet, Peerat Limkonchotiwat.
GenBench @ EMNLP 2023, pp. 193–203.

[10] **Elucidating the Relationship Between Climate Change and Poverty using Graph Neural Networks, Ensemble Models, and Remote Sensing Data**
Parinthapat Pengpun, Alessandro Salatiello.
NeurIPS 2023 Workshop on Tackling Climate Change with Machine Learning, Papers Track.

Research & Professional Experience

NeuLab, Carnegie Mellon University *Aug 2025 – Present*
Undergraduate Researcher, Language Technologies Institute Pittsburgh, PA
Advisor: Prof. Graham Neubig. Research on teaching LLMs and agents complex tasks with less compute and supervision (synthetic data and environments, curricula, reward design).

Multilingual Retrieval & Entity Linking *Aug 2025 – May 2026*
First-author *Think Before You Link*, accepted to the **EMNLP 2026** Main Conference [1]; with Simran Khanuja and Prof. Graham Neubig.

- Built training-free VLM agents that interleave Wikipedia search and reasoning to link entities from text and images across Hindi, Indonesian, Japanese, Tamil, and Vietnamese.
- Built FAISS and BM25 retrieval over **7.4M Wikipedia articles** with 1,024-dimensional E5 embeddings; memory-mapped streaming and chunked encoding fit exact search over the index in **32 GB**.
- The study introduced structural measures of entity rarity based on knowledge-graph coverage and connectivity, complementing popularity-based measures and exposing different failure modes on rare entities.
- Controlled comparisons separated reasoning, retrieval, and their combination: reasoning alone did not substantially improve rare-entity accuracy, while retrieval without reasoning could reduce overall accuracy. Their combination performed best in the study.

SCB10X — Typhoon Team *Mar 2024 – Jan 2025*
Multimodal Junior Scientist (part-time) Bangkok, Thailand

- Led and built **Typhoon Vision (8B)**, Thailand's first vision-language model, from data pipelines through H100/A100 training to open release, with support from the Typhoon team.
- Generated **3M+ synthetic training samples** using vLLM/SGLang batch-inference distillation, quality filtering, and cross-lingual augmentation for Thai multimodal learning.
- Co-authored the Typhoon 2 technical report. The synthetic-data pipeline was later reused and extended by the team for Typhoon OCR.

Siam Commercial Bank *Jun – Jul 2026*
Data Science Intern Bangkok, Thailand

- Built an end-to-end Databricks/Spark X-learner for investment acquisition, estimating **16% more conversions at the same send budget** versus the bank's existing ranking in offline evaluation (Jan–Feb 2026; 2.39M evaluation rows).
- Trained LightGBM on randomized-arm data with Optuna tuning and doubly robust evaluation; recovered corrupted monitoring scores to enable per-customer causal measurement.
- Delivered resumable training pipelines, offline tests, and a documented handoff.

Cognitus Consulting — LambdaX *2024 – 2025*
AI/ML Engineer Remote, US

- Built self-verifying enterprise LLM agents with planner, executor, and verifier components to generate workflows across **100+ custom node types** in Flows, the company's in-house workflow platform.
- Built an MCP server connecting Cursor to the platform for agent-driven workflow creation and editing; worked on synthetic labeling, vision-language extraction from government-contract PDFs, and SAP integrations.

Vidyasirimedhi Institute of Science and Technology — NRL Lab *May – Aug 2024*
Synthetic Data Research Intern Rayong, Thailand
Advisors: Prof. Sarana Nutanong and Dr. Peerat Limkonchotiwat.

- Built a seed-free synthetic-data framework for Thai instruction tuning with **100K+ Wikipedia-grounded samples**; LLM-as-judge filtering yielded a **5K subset that outperformed prior methods trained on 100K samples**.
- Dataset adopted by **WangchanX and AI Singapore**; first-author paper at ACL Student Research Workshop 2024.

Deep Capital Co., Ltd *May 2023 – May 2025*
LLM, Financial Investment AI & Computer Vision Engineer

- Led development of a Thai-language LLM for AI-influencer content and an in-house multimodal model under data and resource constraints; applied model quantization and in-context learning.
- Developed multimodal encoders for visual and numerical stock data and designed transformer-based architectures for trend prediction.
- Led computer-vision development for limited-data settings, including representation learning, unsupervised learning, and model fine-tuning.

Preceptor AI *2024*
NLP Researcher

- Researched English–Thai medical translation using synthetic data, uncertainty estimation, and active learning, including code-switched output that preserves medical terminology in English within Thai text.
- Co-first-authored *On Creating an English-Thai Code-switched Machine Translation in Medical Domain*, published in **Findings of EMNLP 2024** [3].
- The research combined medical-keyword masking with translation to construct approximately **64K synthetic translation pairs**; investigated augmentation, quality filtering, and fine-tuning of **NLLB 3.3B** models.
- The team study compared **52 translation-system configurations** using automatic metrics and medical-professional assessment, finding a mismatch between standard MT metrics and clinicians' preferences for preserving medical terms. Code, models, and the test set were released publicly.

Honors & Awards

13th Heidelberg Laureate Forum — Thailand's Sole CS Representative 2026
Selected as one of 200 young researchers in mathematics and computer science worldwide, and the only computer scientist among Thailand's three national representatives. Selected through Thailand's inaugural national nomination by the National Science and Technology Development Agency (NSTDA). The September 13–18 forum in Heidelberg brings young researchers together with Turing Award and Fields Medal laureates for research discussions and mentorship. NSTDA/POSN sponsored travel; the HLF Foundation covered accommodation and meals.

HackIllinois — Three Project Awards 2026
OpenReality: First Place, Best AI Inference (Modal track); MLH Best Use of Google Gemini API; Most Creative.

180DC × BNY Mellon Case Competition — First Runner-Up 2025

AI Thailand Hackathon — First Runner-Up Jun 2024
Tourism-agent project; competition entry listed as Sawasdee.

SCB10X Bangkok AI Hack — First Place, Model Layer Track Nov 2023
$3,000 team prize; $300 VISTEC Special Bounty; AWS Bounty (credits).

Super AI Engineer Season 3, AIAT — Gold Medal Aug 2023
One of seven Gold Medalists in the six-month AI training and competition program. Hackathon wins included:

- **TPMAP Poverty Prediction**: Integrated satellite imagery with tabular data; developed and tuned ensemble models.
- **Human Activity Recognition**: Developed transformer and hybrid vision models for ultra-wideband signals, with ensembling and pseudo-labeling.
- **Unsupervised Logo Image Search**: Developed image embeddings and optimized a CLIP model using ONNX Runtime.
- **Liver Lesion Detection**: Winning ultrasound-image detection challenge entry.

AI Builders 2 — First Place & Best NLP 2022
Thai question-answer generation project using cross-lingual augmentation; subsequent GenBench @ EMNLP 2023 paper [9].

Scholarships

SCBX Young Tech Scholar 2025
Full-ride scholarship; sole recipient from Thailand in the 2025 cohort.

AWS AI/ML Scholarship 2022 – 2023

Bertelsmann Data Science Scholarship 2023

Projects

OpenReality — Autonomous Agents for Live 3D Mapping HackIllinois 2026

- Built with a four-person team: a phone-browser interface streams a room into a live cloud-based 3D map, without a native app or specialized hardware.
- Personally built the autonomous AI agent that interprets a user's goal, such as a firefighter sweeping a building, and determines what to search for in the 3D space while mapping continues.
- **Awards:** First Place, Best AI Inference (Modal track); MLH Best Use of Google Gemini API; Most Creative Project Overall.

Parameter Golf — Language-Model Training Experiments Mar 2026

- Built an automated research loop that generated and screened training hypotheses: **45 SLURM screening experiments** across 9 tested code revisions, followed by **8×H100 validation on Modal** ($545 total).
- Surpassed the then-standing score (1.1428 → **1.1309 validation bits per byte**, three-seed mean) in public, unmerged PR #493; introduced LeakyReLU^2 to the challenge submissions, later adopted by merged entries.

MAHASAMUT — Multimodal Tourism Agents Jun 2024

- Developed a tourism platform combining a PhuKhao-LLM-based multi-agent system with personalized content and itinerary generation, including discovery of local vendors and less-visited destinations.
- Integrated Thai-dialect speech recognition, text-to-speech, visual question answering, and OCR for Thai-English text trans-

lation and multimodal travel assistance.

- **First Runner-Up, AI Thailand Hackathon 2024** (competition entry listed as Sawasdee); co-authored the LaCATODA @ PRICAI 2024 workshop paper [6].

Thai Vision-Language Prototype — Bangkok AI Hack Nov 2023

- Built a Thai vision-language prototype in under one week using LLaVA, addressing scarce Thai multimodal training data.
- Developed a synthetic-data pipeline using Thai-English translation and language-model quality filtering to preserve meaning during data generation.
- **First Place, Model Layer track**; VISTEC Special Bounty and AWS Bounty. This earlier prototype preceded the Typhoon Vision work described under SCB10X.

Haru + forge — Personal AI Assistant & Coding-Agent Loop 2026

- Built a personal assistant around an append-only event log, 17 rebuildable projections, and local/hosted LLM integration; built forge, a coding-agent loop that extends the assistant against a fixed design specification.
- Measured and fixed multi-writer SQLite ingestion loss from 17% to 0; developed regression tests for ingestion.

GPU-Accelerated PageRank on Wikipedia Dec 2025

Implemented sparse power iteration over 8.3M pages and 163M edges; benchmarked SciPy CPU against CuPy/CUDA GPU execution, achieving a 36.6× speedup and a final runtime under 2 seconds.

3D Hand-Tracking Robotic Arm Jan 2026

Built with one collaborator in one week, combining YOLO hand detection and monocular depth estimation using Apple MPS.

cmuGPT / ScottyLabs 2025 – 2026

Worked on a retrieval-augmented campus LLM and MCP servers.

Earlier Software Projects 2021 – 2022

Solo-built **BCIS App** (2022) with React Native, Django, and AWS; **NEXUS App** (2021) with React Native, FastAPI, WebSockets, and Redis, supporting a live leaderboard with hundreds of concurrent users.

Teaching & Speaking

Teaching Assistant, Carnegie Mellon University Fall 2026

15-213/15-513: Introduction to Computer Systems

- Teach recitations and hold office hours; help students work through course concepts and laboratory assignments.
- Support grading and laboratory work for the undergraduate and graduate course offerings.

Teaching Assistant — AI Builders

Returned to the program as a teaching assistant, mentoring students on their research projects after participating in AI Builders 2.

Workshop Instructor & Lecturer — Artificial Intelligence Association of Thailand 2022 – 2023

- **JCSSE 2023 — Computer Vision Workshop & Image-Generation Talk**: Taught object detection, image classification, and image generation. The accompanying conference talk covered variational autoencoders, generative adversarial networks, and Stable Diffusion.
- **iSAI-NLP 2023 — Advanced Techniques in Prompt Engineering for LLMs**: Covered chain-of-thought, self-consistency, generated-knowledge prompting, automatic prompt engineering, and LLM agents.
- **ChatGPT/OpenThaiGPT for Education** Sep 1, 2023
  Lecture and workshop on prompting techniques, curriculum creation, and research workflows using language models.
- **Chiang Mai University — Data Science Project Contest** Aug 13, 2023
  Advances in AI for Data Science and Applications in Daily Life. Covered preprocessing, imputation, decision trees, and tabular transformers.
- **RMUTL Nan — NLP with Python: Knowledge Graph** Apr 27, 2023
  Taught regular expressions, part-of-speech tagging, and knowledge-graph techniques in an online workshop.

Additional speaking engagements: **AWS Tech Exchange** (data collection); **National Coding Day: Code Era** (AI Builders).

Technical Skills & Languages

ML: PyTorch, Hugging Face, vLLM, SGLang, FAISS; synthetic data, distillation, multimodal training.
Systems: NVIDIA H100/A100 and AMD Instinct training, CUDA/CuPy, Modal, Docker, FastAPI, AWS, GCP, Redis.
Data & Evaluation: Databricks, Spark, LightGBM, Optuna, causal inference, uplift modeling, doubly robust evaluation.
Programming: Python, C, SQL, TypeScript/JavaScript. **Languages:** Thai (native), English (fluent).