

Thodsaporn Chay-intr

Hands-On AI/ML Engineer · Chief AI Officer — LLMs, Speech, and Multimodal AI

Bangkok, Thailand · +66-84-568-1166 · t.chayintr@gmail.com · linkedin.com/in/tchayintr · github.com/tchayintr · boated.zip

AI/ML leader with 9+ years in NLP and production AI, shipping domain-tuned Thai LLMs (OpenThaiGPT, ChindaLLM, ChindaMT, OpenThai Legal) with speech and multimodal systems behind 70M+ API calls. Chief AI Officer at iApp Technology, working directly with enterprise, banking, and government clients — including builds with NVIDIA and NEC — from scoping through delivery. Doctorate from Institute of Science Tokyo, with publications at EACL, RANLP, and the Journal of Natural Language Processing.

Work Experience

iApp Technology Co., Ltd., Bangkok, Thailand Jan 2024 – Present
Creator of OpenThaiGPT; named a Top 5 Thai AI startup in 2026, serving 100+ enterprise, banking, and government clients.

Chief AI Officer · Jun 2025 – Present
Previously Head of AI (2024–25)

- Led the AI and R&D departments, 10+ members across AI, software engineering, and design.
- Delivered 10+ client AI projects, from pre-sales scoping through maintenance.
- Grew the AI product lines (LLMs, speech, vision) 3.5x in a year, to 70M+ calls, using the latest methods.
- Set AI strategy with the CEO and scaled self-hosted GPU serving to hundreds of thousands of daily requests.
- Raised the company's enterprise profile through one.iapp.co.th, open.iapp.co.th, and national AI talks.

Artificial Intelligence Association of Thailand, Thailand Jan 2018 – Aug 2023
ML Researcher | Lecturer

- Advised government and public sectors on AI transformation strategies, contributing to 10+ projects and publications.
- Delivered ML/NLP courses on algorithms, techniques, and tools (e.g., SVM, parsing, scikit-learn).

Institute of Science Tokyo (formerly Tokyo Institute of Technology), Japan Sep 2019 – Mar 2020
Research Assistant

- Collaborated on an NLU module for real-world Japanese dialogue, focusing on linguistic attribute extraction.
- Developed a Seq2Seq model with cross-attention in PyTorch for NLG, improving naturalness in Japanese text generation.

iApp Technology Co., Ltd., Thailand Mar 2017 – Feb 2018
AI/ML Engineer | Researcher

- Led the first Thai Treebank, still used in Thailand's national AI competitions.
- Deployed a GCP annotation tool: +3% annotation quality, half the annotation time.

Education

Institute of Science Tokyo (formerly Tokyo Institute of Technology), Tokyo, Japan Apr 2019 – Sep 2023
Doctor of Engineering — Information and Communications Engineering (*NSK Scholarship Foundation*)

Sirindhorn International Institute of Technology, Pathum Thani, Thailand Jul 2015 – Aug 2018
Master of Engineering — Information, Communication and Technology for Embedded Systems (*TAIST Tokyo Tech Scholarship*)

Thammasat University, Pathum Thani, Thailand Jun 2011 – Aug 2015
Bachelor of Science — Computer Science (*Chairman of the Student Representative Council*)

Key Skills

Programming	Python, C/C++, Rust, Shell
ML Frameworks	PyTorch, PyTorch Lightning, Hugging Face, TensorFlow, PyG, scikit-learn, OpenCV, spaCy, NLTK
Fine-tuning	SFT, DPO, KTO, GRPO, LoRA/PEFT, LLaMA-Factory, ms-swift, TRL, DeepSpeed/FSDP, NVIDIA NeMo
Models tuned	Llama, Qwen, Gemma, Mistral, Typhoon, Nemotron, LLaVA, Whisper, BERT
Domains	Legal, Medical, Banking and finance, Government, Broadcast and telecom, Elderly care
Serving	vLLM, TensorRT-LLM, Triton, Quantization, LoRA serving
Speech	ASR, TTS, Voice cloning, Speech-to-speech, Real-time translation
Infrastructure	Linux, Docker, Kubernetes/k3s, Jenkins, Kong, MinIO, Elasticsearch, Kibana, Hadoop/Spark, SQL/NoSQL, GCP, AWS, Git, FastAPI
AI tooling	Claude Code, Codex, MCP
Languages	Thai (Native), English (Advanced), Japanese (Intermediate)

Selected Publications

- **SiamGPT: Quality-First Fine-Tuning for Stable Thai Text Generation.** *arXiv:2512.19455*, 2025.
- **OpenThaiGPT 1.6 and R1: Thai-Centric Open Source and Reasoning Large Language Models.** *arXiv:2504.01789*, 2025.
- **LLaVAC: Fine-tuning LLaVA as a Multimodal Sentiment Classifier.** *arXiv:2502.02938*, 2025.
- **Extreme Fine-tuning: A Novel and Fast Fine-tuning Approach for Text Classification.** *EACL* 2024.
- **LATTE: Lattice ATTentive Encoding for Character-based Word Segmentation.** *Journal of Natural Language Processing* 2023.

Products and Projects

OpenThai 2.0 Legal: Open-weight Thai legal LLM, built with NVIDIA, BDI, and AIEAT - Built OpenThai 2.0 Legal with NVIDIA, BDI, and AIEAT on Nemotron-3-Nano-30B; released open-weight. - Targets Thai legal language, formal and precise, where general-purpose models fall short.	Jul 2026 – Aug 2026
CIBbot: Legal advice and incident reporting, Royal Thai Police (CIB) - Led the public assistant that advises on law and cybercrime, then prepares and routes each citizen's report. - Built Tri-RAG (hybrid search + PageIndex + knowledge graph): 96.6% retrieval, 91.7% QA, 208-item legal eval. - Ran on-premise and air-gapped over 670 documents from 7 bureaus, with a 3.2s p95 first token.	Mar 2026 – Jun 2026
OpenThaiGPT 1.6 & R1: Thai-centric open-source reasoning LLMs - Built OTG 1.6 by task-arithmetic merging; beat same-base Typhoon2-72B on OpenThaiEval, 78.7 vs 77.2. - Built OTG-R1 on DeepSeek-R1-Distill-32B with LIMO training; beat DeepSeek-R1-70B on AIME24-TH, 56.7 vs 33.3.	Mar 2025
Chinda 3: Enterprise knowledge platform — Chinda to ChindaX to Chinda 3 - Led three generations of iApp's enterprise AI platform; ChindaX and Chinda 3 serve 2,000+ users. - Chinda 3 retrieves at 98.7% vs ~70% for chunk-based RAG, using adaptive hierarchical indexing from PageIndex.	Jan 2024 – Present
Conversational AI for Banking: 70B Thai financial assistant for a banking group Built a 70B Thai financial assistant for a major banking group — benchmarked above GPT-3.5, Gemini Pro, and Claude 3 Haiku on Thai exams — and cut answer latency 7.6s to 3.1s on TensorRT-LLM/Triton.	2024
ASR Pro: Context-aware Thai ASR - Developed an approach enhancing ASR contextual awareness by integrating LLMs with prompt tuning into a fine-tuned ASR model. - Cut WER 3.12 points below Google STT on Thai Common Voice 17, running 1.3× faster.	Aug 2024
NBTC Broadcast Compliance Screening: Thai TV ASR with violation classification - Built content-compliance screening for NBTC, Thailand's broadcast and telecom regulator. - Fine-tuned Whisper for Thai TV transcription (CER < 0.05), then BERT to grade programmes safe, borderline, or violating at 95%+ F1.	Aug 2024
Internal AI Serving Platform: Hybrid k3s across cloud and colo GPU datacenters - Co-designed a two-datacenter k3s cluster, 26 GPUs on 10 servers, running every AI service. - Cut AI cloud spend 46% alongside the migration.	Jul 2026 – Aug 2026
Dual-Stage Face Anti-Spoofing: Active and passive liveness detection - Led development of a face anti-spoofing model with active liveness detection and passive spoof prevention. - Achieved iBeta Level 1 Presentation Attack Detection certification with 0% APCER and BPCER below 3%.	Oct 2024
Passive Liveness and Deepfake Detection: For Bangkok Bank, with NEC Developed passive liveness detection with NEC for Bangkok Bank, screening face spoofing and deepfakes.	Oct 2024 – Present
ASRx: Real-time speech-to-text with instant translation Led a real-time EN/TH/ZH speech-to-text and translation system, deployed live at large conferences.	Jun 2025
LLM Chatbot for Elderly Care: RAG-based companion and consultation assistant Aligned an open-source LLM with SFT/DPO/KTO into a RAG companion for elderly users, with LLM agents handling data synthesis and auto-evaluation.	Oct 2024

Additional work

Open models	ThaiLLM (advising BDI on national LLM serving) · ChindaMT (4B Thai-English MT, 62.3% LC win rate) · one.iapp and open.iapp (commercial APIs + open weights) · ETDA national OCR standard (co-author)
LLM & speech	ChindaLLM (open-source Thai reasoning LLM) · SpeechFlow (Thai-English transcription, 10K+ users) · Thai-English TTS (phoneme-level fine-tuning) · ChindaTTS (voice cloning, sub-second streaming) · ChindaGo (on-device iOS assistant, any open model)
Vision	Multimodal sentiment analysis (CLIP + BERT)
Research	Thai researcher knowledge graph · Simple2In1 (caption fusion for Thai) · First Thai Treebank (still used in national competitions)