Thodsaporn Chay-intr

AI/ML Engineer and NLP Specialist
Bangkok, Thailand • +66-84-568-1166 • t.chayintr@gmail.com • linkedin.com/in/tchayintr • github.com/tchayintr • https://boated.zip

AI/ML Engineer with 6+ years of experience specializing in NLP and developing scalable end-to-end AI solutions. Specialized in PyTorch, foundation models, and DevOps/MLOps pipelines. Expertise in word tokenization, multimodal AI, ML algorithms and techniques. Proven ability to lead multidisciplinary teams to deliver AI initiatives with measurable outcomes.

Key Skills

Technical Skills

- **Programming Languages**: Python, C/C++, Rust, Shell script
- **ML Toolkits**: PyTorch/Lightning, TensorFlow, HuggingFace, PyG, OpenCV, Scikit-learn, Spacy, NLTK, TensorRT-LLM, llama.cpp
- **Tools & Technology**: Linux, Hadoop/Spark, SQL, NoSQL, Docker, Kubernetes, Elasticsearch GCP, AWS, Git

Languages: Thai (Native), English (Advanced), Japanese (Intermediate)

Work Experience

iApp Technology Co., Ltd., Thailand Jan 2024 – Jan 2025
AI/ML Engineer and Head of AI
Leading AI solutions provider in Thailand, key contributor to OpenThaiGPT, delivering tailored AI solutions for diverse clients.

- Led AI projects (text, vision, and audio), transforming practical research into production and doubling team efficiency.
- Managed AI server infrastructure with containers and orchestration, ensuring scalable and reliable ML performance.
- Developed efficient LLM-based agents with RAG and TensorRT-LLM, achieving 97.67% QA accuracy, 87.53% recall for Thai.
- Analyzed linguistic aspects of international languages (Thai, Chinese, Japanese) to deliver NLP projects from concept to release.

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

- Advised professionals and scholars on ML, contributing to the planning and publication of over 20 research papers.
- Delivered courses on ML/NLP, focusing on algorithms, techniques, and tools (e.g., SVM, parsing, and scikit-learn).

Tokyo Institute of Technology, Japan Sep 2019 – Mar 2020
Research Assistant

- Collaborated with a multidisciplinary team to develop NLU/NLG modules for Japanese conversational dialogues.
- Built a BiLSTM Seq2Seq model with cross-attention in PyTorch for natural text generation, validated via human evaluation.

iApp Technology Co., Ltd., Thailand Mar 2017 – Feb 2018
ML Engineer and Researcher

- Led the development of the first Thai Treebank (5,000+ entries) with linguists and developers to advance Thai NLP.
- Developed a syntactic annotation tool for native and web applications, deployed on GCP to support resource development.

Education

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)*

Highlight Projects

ChindaLLM: LLM-powered Chatbot Platform for Advanced Business Automation Sep 2024

- Led a multidisciplinary team to create a chatbot platform powered by multimodal LLMs with a custom RAG engine.
- Fine-tuned multimodal LLMs to meet client requirements and developed a graph-based RAG for enhanced retrieval.

LLM-based Conversational AI System for Banking Queries Jul 2024

- Contributed to developing an LLM-based agent for general banking queries using TensorRT-LLM and customized RAG.
- Synthesized data with LLMs to build intent classification guardrail, boosting accuracy by 27% to 92% for banking compliance.
- Achieved 97.67% QA accuracy, 87.53% recall, and maintained response times under 6.5 seconds.

LATTE: Lattice ATTentive Encoding for Character-based Word Segmentation (Journal of NLP) Jun 2023

- Proposed a method using candidate lattices, GNNs, and attention to refine character representations for word tokenization.
- Integrated Tries with Aho-Corasick to extract candidate characters and words for lattice construction in linear time.
- Achieved SOTA F1-score (97.7% to 99.4%) across Asian languages: Japanese, Chinese, and Thai.

Character-based Thai Word Segmentation with Multiple Attentions (RANLP 2021/Journal of NLP) Sep 2021/Jun 2023

- Proposed a PTM-based word segmentation model with attention across linguistic units (characters, character clusters, subwords, and words), achieving SOTA performance on well-known Thai datasets.
- Developed a subword tokenizer using SentencePiece and a character-cluster tokenizer optimized for Thai linguistic characteristics.

Selected Projects

SpeechFlow: AI-powered Application for Thai-English Transcription, Summarization, and Translation Dec 2024

- Contributed to integrating AI services into an application for seamless Thai-English transcription, summarization, and translation.
- Led the deployment of the ASR Pro engine on server infrastructure, scaling to support thousands of users.

LLM-based Chatbot for Elderly Comfort and Consultation Oct 2024

- Fine-tuned an open-sourced LLM using SFT, DPO, KTO to build a RAG-based chatbot for elderly conversations and support.
- Designed LLM agents for various tasks, including data synthesis and automatic evaluation.

Dual-Stage Face Anti-Spoofing for Active and Passive Liveness Detection Oct 2024

- Led the development of a FAS model with active liveness detection and passive spoofing prevention stages.
- Achieved Level 1 Presentation Attack Detection certification from iBeta with 0% APCER and BPCER below 3%.

ASR Pro: Advanced Context-aware ASR for Thai Aug 2024

- Developed an approach to enhance ASR contextual awareness by integrating LLMs into a fine-tuned ASR model.
- Reduced WER by 3.12% and improved inference speed by 1.3x than top commercial competitors.

Fine-tuning Thai-English TTS Models with Phoneme-level Representations Aug 2024

- Fine-tuned Thai-English TTS models using phoneme-level tokenization, achieving more natural speech than previous models.
- Contributed Thai-English support to a public TTS repository, extending its functionality with fine-tuned models.

Extreme Fine-tuning: A Novel and Fast Fine-tuning Approach for Text Classification (EACL 2024) Mar 2024

- Proposed a fine-tuning approach combining backpropagation with Extreme Learning Machine (ELM) for efficient text classification.
- Reduced fine-tuning time by up to 74.8% with SOTA-level performance on MELD, IEMOCAP, IMDb, and AG News.

LLaVAC: Fine-tuning LLaVA as a Multimodal Sentiment Classifier Jan 2024

- Proposed a method to fine-tune LLaVA for classifying multimodal sentiment labels, incorporating unimodal and multimodal inputs.
- Outperformed SOTA baselines by up to 7.31% in accuracy and 8.76% in weighted-F1 in the MVSA-Single dataset.

A Unification-based Knowledge Graph Construction for Thai Profile Generation from Online Resources Sep 2023

- Constructed a knowledge graph of Thai researchers, using 6+ million entries crawled from online research databases.
- Designed a semi-supervised method with multi-task learning to extract entities/relations, improving F1-score by 8% over baseline.

Simple2In1: A Simple Method for Fusing Two Sequences from Different Captioning Systems into One Sequence Sep 2023

- Developed a T5-based generative model for Thai caption fusion, paraphrasing and merging sequences into one.
- Outperformed baselines by 5.2%, achieving 79% sBLEU and 90% ROUGE-L on a small captioning dataset of 3,168 samples.

Multimodal Sentiment Analysis Using Multiple Labels from Different Modalities Mar 2023

- Developed a sentiment analysis model with CLIP, BERT, and RoBERTa, leveraging text, image, and multimodal labels
- Achieved up to 2% higher F1-scores than previous models, with 74.1% on MVSA-single and 62.0% on MVSA-multiple datasets.

Detecting Fraud Job Recruitment Using Features Reflecting from Real-world Knowledge of Fraud Mar 2022

- Developed a method to classify fake job recruitments using a set of novel features designed to reflect fraudster behaviors.
- Yielded accuracy of 97.64% for Employment Scam Aegean Dataset (EMSCAD).

Public Budget Usage Monitoring System (Bronze Medal - The 47th International Exhibition of Inventions Geneva) Feb 2019

- Cooperated with an interdisciplinary team to develop a monitoring system that utilizes Scrapy to crawl large-scale unstructured data from government sites, such as procurement and budget portals, for corruption detection in text data. Deployed by two organizations.
- Developed a text classification method in TensorFlow, with rule-based enhancements, for corruption detection, validated by experts.

Selected Relevant Publications

Extreme Fine-tuning: A Novel and Fast Fine-tuning Approach for Text Classification Mar 2024
18th Conference of the European Chapter of the Association for Computational Linguistics (EACL 2024)

- Proposed a fine-tuning approach combining backpropagation with Extreme Learning Machine (ELM) for efficient text classification.
- Reduced fine-tuning time by up to 74.8% with SOTA-level performance on MELD, IEMOCAP, IMDb, and AG News.

LATTE: Lattice ATTentive Encoding for Character-based Word Segmentation Jun 2023
Journal of Natural Language Processing, Volume 30, Issue 2

- Proposed a method using candidate lattices, GNNs, and attention to refine character representations for word tokenization.
- Integrated Tries with Aho-Corasick to extract candidate characters and words for lattice construction in linear time.
- Achieved SOTA F1-score (97.7% to 99.4%) across Asian languages: Japanese, Chinese, and Thai.

Character-based Thai Word Segmentation with Multiple Attentions Sep 2021/Jun 2023
International Conference on Recent Advances in Natural Language Processing (RANLP 2021) and Journal of Natural Language Processing, Volume 30, Issue 2

- Proposed a PTM-based word segmentation model with attention across linguistic units (characters, character clusters, subwords, and words), achieving SOTA performance on well-known Thai datasets.
- Developed a subword tokenizer using SentencePiece and a character-cluster tokenizer optimized for Thai linguistic characteristics.

Selected Certificates

Generative AI with Large Language Models (LLMs) Aug 2023
DeepLearning.AI, AWS (Coursera)

Introduction to Big Data with Spark and Hadoop Aug 2023
IBM (Coursera)

Introduction to Deep Learning & Neural Networks with Keras Jul 2023
IBM (Coursera)

Machine Learning with Python Apr 2023
IBM (Coursera)

Academic Services, Leaderships, and Mentoring

Conference Publication Chairs

- iSAI-NLP-AIoT 2024, 2021
- KICSS 2021
- KSE 2020
- STUD 2019

Conference Organizing Members

- iSAI-NLP-AIoT 2024, 2021, 2020, 2019, 2018, 2017
- KSE 2021
- KICSS 2018
- IDT 2019
- ISAAC 2017
- PRICAI 2016

Conference Reviewers

- iSAI-NLP-AIoT 2024, 2021, 2018

Consultants

- Super AI Engineer Thailand 2024, 2023, 2022

Journal Association Editor

- Journal of Intelligent Informatics and Smart Technology (JIIST)