Thodsaporn Chay-intr

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

AI/ML Engineer and NLP Researcher with over 6 years of experience in developing AI applications across both academic and commercial settings. Expert in leveraging LLMs for chatbot/search development, text classification, and synthetic data generation, achieving high performance metrics. Proficient in OCR, ASR, TTS, multimodal learning, and graph techniques. Recognized for leading strategic AI initiatives, managing multidisciplinary teams, and delivering successful project outcomes.

Work Experience

iApp Technology Co., Ltd., Thailand Jan 2024 – Aug 2024
AI/ML Engineer and Head of AI
AI innovator in Thailand, created OpenThaiGPT and developed AI applications for private and government sectors across Thailand and Asia.

- Developed an LLM-based chatbot/search using LLMs with llama.cpp, TensorRT-LLM, and RAG, achieving 97.67% QA accuracy and 87.53% perfect recall with response times under 6.5 seconds.
- Synthesized training samples for text classification with LLMs, boosting accuracy above 90% and improving by up to 27%.
- Managed and led a multidisciplinary team to develop AI projects from inception to deployment, including OCR, ASR, TTS, LLMs, and image processing, doubling team capacity and efficiency.
- Led R&D initiatives to transform cutting-edge research into practical AI products.

Artificial Intelligence Association of Thailand, Thailand Jan 2018 – Aug 2023
ML Researcher and Lecturer (Ad Hoc and Project-based)
Non-profit aiming to advance AI in Thailand through consultations, conferences, courses, and events nationwide.

- Delivered Python courses on data analysis and ML to hundreds of AI enthusiasts, covering tools such as pandas and scikit-learn, and algorithms including SVM and NNs.
- Consulted with industry professionals and scholars on ML concepts, guiding plans and assisting in publishing over 20 international research papers and projects.

Tokyo Institute of Technology, Japan Sep 2019 – Mar 2020
Research Assistant
A top-100 by QS-ranked university, known for diverse research and strong emphasis on innovation and technology.

- Collaborated with a multidisciplinary team to develop modules for NLU/NLG units for Japanese conversational dialogs.
- Implemented a Seq2Seq model using BiLSTM with cross-attention in PyTorch for more natural text generation, validated through human evaluations.

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

- Constructed the first Thai Treebank with over 5,000 entries with a team of linguists and developers, enhancing resources, research, and applications in the Thai NLP community.
- Developed a syntactic annotation tool in native (Java) and web (React, Python) applications, deployed on GCP, supporting ongoing resource developments.

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)*

Key Skills

Technical Skills

- **Programming Languages**: Python, C/C++, Rust, Java
- **ML Toolkits**: PyTorch/Lightning, TensorFlow, Hugging Face, PyG, OpenCV, Scikit-learn, Spacy, NLTK, llama.cpp, TensorRT-LLM
- **Tools & Technology**: Linux, Hadoop/Spark, SQL, NoSQL(MongoDB, Neo4j), Docker, Elasticsearch GCP, AWS, Git

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

Selected Projects

LLM-based Conversational AI System for General Banking Queries Jul 2024

- Contributed to POCs and development of an LLM-based chatbot/search system for a tech innovation division of a leading bank in Thailand using LLaMa (2, 3, and 3.1), OpenThaiGPT, llama.cpp, TensorRT-LLM, and RAG.
- Achieved 97.67% QA accuracy, 87.53% perfect recall, and maintained response times under 6.5 seconds.
- Introduced intent classification with over 92% accuracy as a guardrail to filter the input/output of the system, ensuring the responses align with banking policies.

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

- Proposed a novel text classification fine-tuning approach incorporating backpropagation with extreme learning machine, reducing fine-tuning time while retaining classification accuracy and F1-score.
- Attained faster fine-tuning time by up to 74.8% with comparable scores over recent state-of-the-art models on MELD, IEMOCAP, IMDb, and AG News datasets.

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

- Proposed a method to fine-tune Large Language-and-Vision Assistant (LLaVA) as a classifier for classifying multimodal sentiment labels by designing a prompt to consider unimodal and multimodal labels and generating predicted labels.
- Outperformed state-of-the-art baselines by up to 7.31% in accuracy and by 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 for 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 captions fusion, outperforming baselines by 5.2% in sBLEU and ROUGE-L scores.
- Accomplished a sBLEU score of 79% and a ROUGE-L score of 90% for a small captioning dataset comprising 3,168 samples.

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

- Proposed a sequence labelling method that integrates multi-granularity linguistic units, Lattices, GNNs, PTMs, and Attention Mechanism to generate and refine text representations for word segmentation. in PyTorch with PyG.
- Achieved state-of-the-art performance (97.7% to 99.4% of F1-score) across Asian languages: Japanese, Chinese, and Thai.

Multimodal Sentiment Analysis Using Multiple Labels from Different Modalities Mar 2023

- Collaborated with students to design and implement a sentiment analysis model for social network data, leveraging text, image, and multimodal labels using CLIP, BERT, and RoBERTa. Yielded up to 2% improvement in F1-score over recent models.
- Attained F1-scores of 74.1% for MVSA-single and 62.0% 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)

- A novel fine-tuning approach that incorporates backpropagation followed by extreme learning machine for text classification.
- Achieved comparable performance state-of-the-art works while reducing fine-tuning time by up to 74.8% across well-known text classification datasets (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

- A sequence labelling method that integrates multi-granularity linguistic units, Lattices, GNNs, PTMs, and Attention Mechanism
- Achieved state-of-the-art performance (97.7% to 99.4% of F1-score) across Asian languages: Japanese, Chinese, and Thai.

Character-based Thai Word Segmentation with Multiple Attentions Jun 2023
Journal of Natural Language Processing, Volume 30, Issue 2

- A sequence labeling model that utilizes BiLSTM, PTMs and multiple attention mechanisms across multiple linguistic units, including characters, character clusters, subwords, and words.
- Achieved state-of-the-art performance on well-known Thai datasets, on par with LATTE: Lattice ATTentive Encoding for Character-based Word Segmentation.

Character-based Thai Word Segmentation with Multiple Attentions Sep 2021
International Conference on Recent Advances in Natural Language Processing (RANLP 2021)

- A preliminary sequence labeling model that utilizes only BiLSTM and multiple attention mechanisms across multiple linguistic units, including characters, character clusters, subwords, and words
- Achieved state-of-the-art performance on a well-known Thai dataset.