Echo Diyun Lu

AI Engineer/Researcher | Recommender Systems, NLP, ML & LLMs

Paris, France | +33 767 295 717 | ludiyun@gmail.com

echo-lu.com | github.com/Echo0117 | linkedin.com/in/diyun-lu

Professional Summary

Machine learning engineer and PhD researcher with 5+ years experience across recommender systems, ranking, NLP/LLMs, causal inference, time-series forecasting, and production ML. Built end-to-end NLP and recommendation workflows spanning data collection, feature engineering, distributed training, deployment, and iterative optimization. Current research focuses on long-term marketing effects, adaptive forecasting, regime switching, and distribution-free uncertainty.

Core Expertise

Recommendation and personalization: Two-tower models, DIN, DIEN, vector search, ranking, user profiling, click-through-rate and retention optimization.
LLM and NLP: Transformers, BERT, RoBERTa, ALBERT, InstructGPT, fine-tuning, NER, semantic matching, multi-label classification, topic modeling.
Research and modeling: Causal inference, time-series forecasting, regime-switching models, conformal prediction, probabilistic generative models, deep learning.
ML systems: Python, PyTorch, TensorFlow, distributed training, CUDA, ONNX, TensorRT, Docker, Kubernetes, AWS, REST APIs, gRPC, CI/CD.

Publication

Lu, E. D., Findling, C. S. M., Clausel, M., Leite, A., Gong, W., & Kersaudy, P. (2025). *Adaptive Regime-Switching Forecasts with Distribution-Free Uncertainty: Deep Switching State-Space Models Meet Conformal Prediction.* NeurIPS 2025 Workshop on Recent Advances in Time Series Foundation Models (BERT2S).

Work Experience

Pernod Ricard **Nov 2023 – Present**
AI Researcher *Paris, France | Full-time*

- Research long-term marketing effects using causal inference, time-series forecasting, regime-switching models, and uncertainty quantification.
- Design adaptive forecasting and evaluation approaches for non-stationary business data.

Go Playfully Inc. **Sep 2023 – Jan 2025**
AI Engineer *California, USA | Remote*

- Built an LLM-based recommender system for personalized activities and events.

Snoop Media Co., Ltd. **Mar 2023 – Sep 2023**
AI Engineer Intern *Paris, France*

- Built a recommender system for streaming media using NLP, LLM, and personalization methods.

TideSwing Technology Co., Ltd. **Jul 2020 – Aug 2022**
AI Engineer *Beijing, China | Full-time*

- Owned end-to-end NLP and recommender-system workflows from data collection and labeling through training, testing, deployment, and iterative optimization.
- Built user and text representations with two-tower models and developed retrieval and ranking pipelines using vector search, DIN, DIEN, tree models, and contrastive learning.
- Performed feature filtering, data preprocessing, model training, and iterative system optimization to improve click-through rates and retention rates.
- Distributed-trained an LLM on user-generated content and fine-tuned pretrained models for classification, NER, and semantic matching.
- Accelerated model training and inference with CUDA, ONNX, and TensorRT.
- Built user-profile systems covering label design, label extraction, graph-database storage and querying, regular updates, and integration with recommendation systems.
- Delivered NLP tasks including named-entity recognition, multi-label classification, topic modeling, word segmentation, and new-word discovery.

Technologies: Transformers, word embeddings, BiLSTM, LDA, CRF, Word2Vec, POS tagging, XGBoost, BERT, RoBERTa, ALBERT, InstructGPT, transfer learning, knowledge graphs, graph databases.

Neusoft Corporation, Advanced Product Development **Apr 2018 – Jul 2018**
NLP Engineer Intern *Shenyang, China*

- Worked on research and development of intelligent dialogue robots using NLP.

UEC GROUP LTD **Nov 2017 – Feb 2018**
Big Data Engineer Intern *Beijing, China*

Additional Research Experience

University of Cambridge, Wellcome Sanger Institute **Aug 2019 – Dec 2019**
AI Researcher / Computational Analyst *Cambridge, UK | Full-time*

- Processed genomic and gene-expression datasets and compared machine-learning models, including linear and Lasso regression, to predict gene essentiality.

University College London - Master's Research **Sep 2018 – Sep 2019**
Combining Imaging Biomarkers and AI to Shed Light on Alzheimer's Disease *London, UK*

- Combined machine-learning and NLP techniques on imaging biomarkers and medical records to investigate Alzheimer's disease and support earlier diagnosis.

Techniques: LDA topic modeling, Python data processing and analysis.

Education

PhD, Artificial Intelligence **Mar 2024 – Present**
CNRS (French National Centre for Scientific Research), Inria, Universite de Lorraine *France*

MSc, Artificial Intelligence **Sep 2022 – Sep 2023**
Universite Paris-Saclay *France*

Relevant modules: Deep Learning for NLP; Probabilistic Generative Models; AI Algorithms for Data Science; Computer Vision; GPU Programming; Graphical Models; Signal Processing; Reinforcement Learning; Multilingual NLP; Advanced Optimization.

MSc, Biomedical Engineering (AI Route) **Sep 2018 – Sep 2019**
University College London *United Kingdom*

Relevant modules: Programming Foundations for Medical Image Analysis; Information Processing in Medical Imaging; Computational Modelling for Biomedical Imaging.

BSc, Biomedical Engineering (AI Route) **Sep 2014 – Jul 2018**
Northeastern University *China*

Honours and Awards: Scholarship of Excellent Student; Progressive scholarship

Exchange Student, Biomedical Engineering **Sep 2016 – Feb 2017**
University of Silesia *Poland*

Honours and Awards: Overseas study scholarship

Skills

Programming and systems:
Python, Kotlin, MATLAB, Linux, MySQL, graph databases, knowledge graphs, GPU computing, PySpark, Flask, Docker, RESTful APIs, gRPC, Git, Kubernetes, AWS, CI/CD.

AI and machine learning:
PyTorch, TensorFlow, pandas, NumPy, machine learning, deep learning, computer vision, NLP, LLMs, recommender systems, scikit-learn, XGBoost.

Languages:
Chinese Mandarin (native); English (full professional proficiency); French (intermediate).