

# Du Yiming

+86 15051862878 | [1155186185@link.cuhk.edu.hk](mailto:1155186185@link.cuhk.edu.hk) | PhD Student | Natural Language Processing

## EDUCATION

<b>The Chinese University of Hong Kong</b> <i>PhD : System Engineering and Engineering Management, Minor in <b>Natural Language Processing</b></i>	Hong Kong, China Aug. 2022 – Aug. 2026
<b>Southeast University</b> <i>MEng : Computer Science &amp; Engineering, Minor in <b>Natural Language Processing, TOP 10%</b></i>	Jiangsu, China Aug. 2019 – June 2022
<b>Southeast University</b> <i>BEng : Computer Science &amp; Engineering, Minor in <b>Software Engineering, TOP 30%</b></i>	Jiangsu, China Aug. 2015 – May 2019
<b>Core Course:</b> Artificial Intelligence, Machine Learning, Algorithm Analysis and Design, Natural Language Processing <b>Languages:</b> Python, Java, C++, SQL, HTML/CSS <b>Libraries:</b> Pytorch, NumPy	

## RESEARCH EXPERIENCE

<b>PhD student, Huawei&amp;CUHK Laboratory, The Chinese University of Hong Kong</b> <i>Supervisor: Prof. Kam-Fai Wong</i> <ul style="list-style-type: none"><li>• <b>Dialogue System:</b> Designed human-like memory integration mechanisms leveraging LLMs with RAG architectures.</li></ul>	Aug. 2022 - Now Hong Kong, China
<b>Visiting Researcher, Informatics Forum, The University of Edinburgh</b> <i>Supervisor: Prof. Jeff Z. Pan</i> <ul style="list-style-type: none"><li>• <b>Personal Agent:</b> explore parametric memory and contextual memory in personal assistant.</li></ul>	Jan. 2025 - Now Edinburgh, UK
<b>Research Intern, Sensetime</b> <i>Mentor: Fei Tan</i> <ul style="list-style-type: none"><li>• <b>Model Editing:</b> Inserted industrial knowledge into the llama layer via WISE.</li><li>• <b>Data Synthetic:</b> Generated 3 trillion token Chain-of-Thought dataset improving LLaMA's performance.</li><li>• <b>Multi-turn Dialogue System:</b> Developed turn-weighted noise-robust SFT pipeline for LLMs</li></ul>	Aug. 2024 - Jan. 2025 Shenzhen, China
<b>Visiting Researcher, Noah's Ark Lab, Huawei</b> <i>Mentor: Baojun Wang</i> <ul style="list-style-type: none"><li>• <b>Memory in LLM Dialogue System:</b> Spearheaded the creation of a memory-based LLM dialogue system with RAG framework and FAISS vector databases for optimized performance.</li></ul>	Aug. 2023 - Aug. 2024 Hong Kong, China
<b>Research Assistant, Web Mining Lab, City University of Hong Kong</b> <i>Supervisor: Prof. Liu Xiaofan</i> <ul style="list-style-type: none"><li>• <b>Sentence Classification:</b> Developed DPCIPI model based on DNABERT for viral gene similarity evaluation.</li></ul>	Aug. 2021 - Feb. 2022 Hong Kong, China
<b>Research Experience, Imperial College London</b> <i>Supervisor: Prof. Lucia Specia</i> <ul style="list-style-type: none"><li>• <b>Sentiment Classification:</b> Analyzed news headline sentiment disparities using CNN and BERT on SemEval-2020 Task 1 dataset.</li></ul>	June. 2021 – July. 2021 London, England
<b>Master Student, Atur Lab, Southeast University</b> <i>Supervisor: Prof. Gao Zhiqiang</i> <ul style="list-style-type: none"><li>• <b>Text Generation:</b> Designed middle representation language Pseudo-SQL and Soft Template attention to generate comparative questions.</li><li>• <b>Name Entity Recognition:</b> Recognized named entity with BERT in electronic medical records on CCKS2017 dataset</li></ul>	Mar. 2019 – Aug. 2022 Nanjing, China

## SELECTED PUBLICATIONS

- **Yiming Du\***, Wenyu Huang\*, Danna Zheng\*, Zhaowei Wang, Sebastien Montella, Mirella Lapata, Kam-Fai Wong, Jeff Z. Pan. Rethinking Memory in AI: Taxonomy, Operations, Topics, and Future Directions. *arxiv version* (to submit)
- **Yiming Du\***, Bingbing Wang\*, Yang He, Bin Liang, Baojun Wang, Zhongyang Li, Lin Gui, Jeff Z. Pan, Ruifeng Xu, Kam-Fai Wong. Modeling Long-Term Memory for Multi-Session Task-Oriented Dialogue Systems via Memory-Active Policy. *ARR, Feb 2025*, *ARR, Feb 2025* (under review)
- **Yiming Du**, Yifan Xiang, Bin Liang, Dahua Lin, Kam-Fai Wong, Fei Tan. TWiNS: An Implicit Noise Suppression Approach for Multi-Turn Dialogue Fine-Tuning. *ARR, Feb 2025* (under review)

- Cai Ke, **Yiming Du**, Bin Liang, Xiang Yifan, Lin Gui, Zhongyang Li, Baojun Wang, Yue Yu, Hui Wang, Kam-Fai Wong, Ruifeng Xu. Flexibly Utilize Memory for Long-Term Conversation via a Fragment-then-Compose Framework. *ARR, Feb 2025* (under review)
- Zhengyi Zhao\*, Shubo Zhang\*, **Yiming Du**, Bin Liang, Baojun Wang, Zhongyang Li, Binyang Li, Kam-Fai Wong. Fine-tuning after Prompting: an Explainable Way for Classification. *ARR, Feb 2025* (under review)
- Boyang XUE, Hongru WANG, Rui Wang, Sheng Wang, Zezhong WANG, **Yiming Du**, Kam-Fai Wong. A Comprehensive Study of Multilingual Confidence Estimation on Large Language Models. *ARR, Feb 2025* (under review)
- Bingbing Wang\*, **Yiming Du\***, Bin Liang, Zhixin Bai, Baojun Wang, Ruifeng Xu, Kam-Fai Wong. A New Formula for Sticker Retrieval: Reply with Stickers in Multi-Modal and Multi-Session Conversation. *AAAI 2024* (Accepted, co-first author)
- **Yiming Du**, Hongru Wang, Zhengyi Zhao, Bin Liang, Baojun Wang, Wanjun Zhong, Zezhong Wang, Kam-Fai Wong. PerLTQA: A Personal Long-Term Memory Dataset for Memory Classification, Retrieval, and Fusion in Question Answering. *ACL - SIGHAN 2024* (Best Paper)
- Hongru Wang\*, Baohang Zhou\*, Zhengkun Zhang, **Yiming Du**, David Ho, Kam-Fai Wong. M3SUM: A Novel Unsupervised Language-guided Video Summarization. *ICASSP 2024* (published)
- Hongru Wang, Lingzhi Wang, **Yiming Du**, Liang Chen, Jingyan Zhou, Yufei Wang, Kam-Fai Wong. A Survey of the Evolution of Language Model-Based Dialogue Systems. (*under review*)
- Zhiming Mao, Huimin Wang, **Yiming Du**, Kam-Fai Wong. UniTRec: A Unified Text-to-Text Transformer and Joint Contrastive. *ACL 2023* (published)

## PATENTS

<b>A dialogue memory storage, extraction and electornical device.</b>	Huawei, China
<i>Baojun Wang, Yiming Du, Kam-Fai Wong</i>	<i>2024</i>

## PROJECTS

<b>Human-like Memory Integration for Dialogue System</b>   <i>Python</i>	Jan. 2023 – Jan. 2025
• Led the memory-based dialogue system demo development and organize research group meeting and conduct research related to memory.	
<b>Multi-model summarization system</b>   <i>Java, Python</i>	Oct. 2022 – Oct. 2024
• Led the innovation and development of an advanced multi-modal summarization system, significantly enhancing the integration of audio, image, and text information through Large Language Models (LLM).	
<b>Software Engineering Intern, National Instruments</b>	Oct. 2018 – Jan. 2019
• Refactored Compile Worker Module in LabView Software and redesigned the architecture to be more readable and improve maintainability.	

## HONORS AND AWARDS

<b>ORAP</b>	Hong Kong, China
<i>Overseas Research Attachment Programme Awards</i>	<i>Jan. 2024</i>
<b>Excellent Scholarship</b>	Jiangsu, China
<i>Awarded by Shenzhen Stock Exchange, Southeast University, Huawei, Keyuan Automation separately</i>	<i>2020, 2019, 2018</i>
<b>First Place</b>	Jiangsu, China
<i>Jiangsu Postgraduate Mathematical Modeling Scientific Research Innovation Practice Competition</i>	<i>Apr. 2020</i>
<b>First Place</b>	Shanghai, China
<i>Chinese Undergraduate Computer Design Contest</i>	<i>Aug. 2018</i>
<b>Meritorious Winner</b>	US
<i>Mathematical Competition In Modeling</i>	<i>Apr. 2018</i>

## LEADERSHIP

**MOE Lab, The Chinese University of Hong Kong | Project Leader**

- Led development of a memory-augmented dialogue system, delivering both research outcomes and a dialogue system demo.

**Atur Lab, Southeast University | Group Leader**

- Organize the series of reports on Representation Learning with 5 group members, and I am responsible for designing the report scheme and quality control.

**The School of Computer Science and Engineering, SEU | Forum Organizer**

- Organize the 2021 academic forum at the School of Computer Science and Engineering, Southeast University, with topics including Artificial Intelligence, Image Processing, Software Engineering, Computer Network