Hande Dong

donghd66@gmail.com | (+86) 17737085538 | https://donghande.github.io/ | Google Scholar | Github

Short Bio

- **Full-Process Development of LLMs**: Expertise in pre-training, fine-tuning, agent integration, Retrieval-Augmented Generation (RAG), and efficient deployment, with extensive hands-on experience.
- **Academic Excellence**: Over 1,000 citations, following the latest advancements in LLM research community.
- **Product Innovation**: Successfully delivered impactful products at Tencent, gaining valuable insights into product innovation and development to drive future success.
- **Balance Between Technology and Product**: Proficient in bridging the gap between technology and product needs from the engineering perspective.
- **Collaboration and Communication**: Extensive experience in teamwork, cross-departmental coordination, and customer engagement to deliver solutions.
- **Career Objectives**: In the short term, focus on delivering valuable products powered by LLMs; in the long term, promote AI's deep integration across industries to achieve both technological and social benefits.

Professional Experience

Tencent - CSIG - Developer Product Center, Senior AI Researcher Aug 2023 – Present

(1) Tencent Cloud AI Code Assistant (Core Responsibility, 80%)

- **Responsibilities**: Owned the development of code large language models, focusing on key modules such as completion, dialogue, and agents. Enhanced model performance, significantly improving the accuracy of code LLMs to enhance developers' productivity.
- **Achievements**: (1) Played a pivotal role in the product's journey from concept to launch. (2) Expanded internal adoption from trial use to covering 80% of internal users. (3) Achieved daily active users (DAUs) in the tens of thousands and secured over 10 key accounts for private deployments.

(2) AI Technology Support for Tencent Design Products (Core Responsibility, 20%)

- **Responsibilities**: Developed AI technologies for "Text-to-UI" and "Image-to-UI" modules using multi-modal LLMs and vision models, alongside agent integration.
- **Achievements**: Refined functionalities to enhance user experience and validated the potential of generative AI in UI design, setting the direction for future AI-based design innovations.

Performance Evaluations: Consistently rated "Outstanding" in performance reviews.

International Digital Economy Academy, Algorithm Engineer Jul 2022 – Aug 2023

- Responsibilities: Researched pre-training and LLM applications in code understanding and generation, reproducing state-of-the-art models and exploring their academic and practical potential.
- Achievements: Enhanced code search precision by 5% using advanced methods. Published 2 high-quality papers. Provided crucial support for theoretical and practical advancements in the team.

Education

University of Science and Technology of China, M.S. in Electrical Engineering and Information Science, **Advisor: Prof. Xiangnan He** Sep 2019 – Jun 2022

- **Research Areas**: Graph Neural Networks, Data Mining, Recommendation Systems, Information Retrieval.
- **Achievements**: (1) Published 3 top-tier conference/journal papers. (2) A survey on biases in recommendation systems and the AutoDebias method have collectively garnered over 1,000 citations, demonstrating significant influence within the research community.

University of Science and Technology of China, B.S. in Applied Physics Aug 2015 – Jun 2019

- **Honors**: Graduated with provincial and university-level excellence awards.

Projects

Post-Training Data Flywheel for Model Optimization

- **Background & Objective**: With tens of thousands of DAUs, user feedback (e.g., bad cases) and usage logs provided a rich source for model improvement. Key challenges included: (1) addressing frequent user-reported issues, (2) utilizing user data to drive model optimization, and (3) ensuring an efficient and scalable process.

- **Solution**: (1) Collaborated with the data team to build a centralized data platform to store user logs, product reports, and feedback. (2) Developed a post-training framework using SFT, DPO, and rejection sampling to iteratively improve the model. (3) Introduced multi-source labeling strategies, including temporal log analysis, advanced model annotations, and human-assisted data labeling.
- **Outcome**: Successfully resolved high-frequency user issues, transforming user data into reusable model assets. This initiative established a scalable feedback loop, driving long-term optimization.

User-Centric Code Completion Data Construction

- **Background & Objective**: Traditional code completion models trained using FIM (Fill-in-the-Middle) often rely on random segmentation, which does not align with user-triggered behaviors, leading to many bad cases.
- **Solution**: Simulated user-trigger patterns to construct realistic FIM training data, categorizing user actions into line completion, partial-line completion, inline completion, and block completion. Balanced data generation across these categories to better reflect real-world usage patterns.
- **Outcome**: Improved completion acceptance rate by 8 percentage points (from 15% to 23%).

Data Cleaning and Repository-Level Pre-Training for Code Models

- **Background & Objective**: Early team efforts in code LLMs development faced challenges due to inadequate data cleaning and construction techniques. Addressing these gaps was critical to building a robust foundation for future exploration of our product.
- **Solution**: Conducted large-scale experiments to develop over ten effective data cleaning strategies, including text-level cleaning and static code analysis. Introduced an innovative repository-level data construction method, surpassing traditional file-level approaches, to enhance contextual understanding.
- **Outcome**: These data processing techniques became core team assets of our product. Repository-level training improved code generation acceptance rates by 3 percentage points (9% $\rightarrow$ 12%).

Inference Optimization for Low-Cost, Low-Latency Models

- **Background & Objective**: Ensuring low-cost, low-latency inference while maintaining model accuracy was critical for competitive differentiation.
- **Solution**: Following the SOTA techniques such as prefix caching, chunked prefill, and improved KV cache CPU offloading to reduce latency and costs. Introduced an intelligent routing strategy to allocate requests based on model size, balancing efficiency and accuracy. Designed new inference methods (SP and FIMX) to optimize suffix processing in FIM.
- **Outcome**: Reduced inference latency by 30%, increased generation rates by 5%, and enabled scalability for larger model exploration, significantly improving user experience.

Client Support and Technical Consulting for Key Accounts

- **Background & Objective**: To facilitate the private deployment and adoption of Tencent Cloud AI solutions for key enterprise clients, technical consulting was essential in bridging technical gaps and ensuring client success.
- **Solution**: Collaborated with client technical teams to provide in-depth technical support and workshops. Proposed customization tailored to client-specific datasets, successfully implementing them across multiple enterprise scenarios. Co-developed post-training mechanisms with client teams.
- **Outcome**: Established strong client relationships and delivered high-impact solutions, contributing to product growth and securing key accounts across multiple industries.

Publications

- Improving Code Search with Hard Negative Sampling Based on Fine-tuning, APSEC 2024, **Hande Dong**, Jiayi Lin, Yanlin Wang, Yichong Leng, et al.
- Survey of Code Search Based on Deep Learning, TSEM 2023, Yutao Xie, Jiayi Lin, **Hande Dong**, Lei Zhang, et al.
- Bias and Debias in Recommender System: A Survey and Future Directions, TOIS 2023, Jiawei Chen, **Hande Dong**, Xiang Wang, Fuli Feng, Meng Wang, Xiangnan He.
- AutoDebias: Learning to Debias for Recommendation, SIGIR 2021, Authors: Jiawei Chen*, **Hande Dong***, Yang Qiu, Xiangnan He, Xin Xin, Liang Chen, Guli Lin, Keping Yang. (Co-first author)
- On the Equivalence of Decoupled Graph Convolution Network and Label Propagation, WWW 2021, **Hande Dong**, Jiawei Chen, Fuli Feng, Xiangnan He, et al.

Note: **My papers have been cited over 1,000 times** according to Google Scholar. Full publication list available at Google Scholar.