

# 周文涌

(+86)177-1036-0647 · wenyongz@connect.hku.hk · 视觉模型, 大语言模型 · 高效系统设计

## 求职意向

寻找与生成式人工智能、大语言/多模态模型相关的研究员或投资分析岗位。

## 教育背景

香港大学 (University of Hong Kong), 博士 香港, 中国, 2021.09 – 2025.08

电气与电子工程系, 导师: Prof. Ngai Wong 和 Prof. Can Li

毕业论文: 面向硬件的高效与鲁棒隐式神经表示方法研究

西北大学 (Northwestern University), 硕士 埃文斯顿, 美国, 2019.09 – 2021.06

电子与计算机工程系, 导师: Prof. Sedia Memik

毕业论文: 基于可微分神经架构搜索的在内存计算芯片上的软件-硬件协同设计

天津大学 (Tianjin Univeristy), 学士 天津, 中国, 2015.09 – 2019.07

电子科学与技术系, 保送研究生, 期间赴 UC Berkeley 交换

毕业论文: 三维空间中的 Voronoi 剖分的统计特性分析 (天津市优秀本科毕业论文)

## 实习经历

知存科技, 存算科学家 北京, 中国, 2024.11 – 2025.08

### 大语言模型低精度推理和训练

- 设计了梯度量化方法, 通过动态区间裁剪和误差补偿机制缓解梯度量化带来的数值不稳定问题, 确保模型训练的有效性和收敛速度
- 开发了基于存算一体硬件的低精度梯度更新算法, 优化了梯度累积和反向传播过程中的数值表示, 提升了大语言模型的收敛性能
- 设计了针对硬件噪声和量化误差的鲁棒性训练方法, 增强了大语言模型在模拟环境和实际硬件上的泛化能力

京东集团, 算法实习生 北京, 中国, 2024.01 – 2024.06

### 基于扩散模型的试衣模型搭建

- 设计并实现了基于条件扩散模型的虚拟试衣系统, 结合人体结构特征和服装样式信息, 实现了个性化服饰拟合与高质量图像生成
- 优化了扩散模型的训练流程, 引入多尺度图像重建损失、感知损失及对抗性损失, 显著提升了生成图像的清晰度、纹理细节和视觉真实感
- 构建并标注了大规模服饰数据集, 涵盖多种人体姿态、服装样式和背景场景, 用于增强模型的泛化能力和对多样化需求的适配性能

字节跳动集团, 算法实习生 上海, 中国, 2021.04 – 2021.08

### Tiktok 中商家广告投流的冷启动策略

- 研究商家广告投流的冷启动策略, 通过利用商家历史行为数据和广告内容特征以提升冷启动效果
- 引入基于用户兴趣分布的相似性匹配算法, 提高了冷启动阶段的点击率 (CTR) 和转化率 (CVR)
- 构建了离线评估框架, 模拟冷启动场景下的广告投放效果

## 代表性论文

- Wenyong Zhou, et al., "Distribution-Aware Hadamard Quantization for Hardware-Efficient Implicit Neural Representations," *IEEE International Conference on Multimedia & Expo*, 2025.
- Wenyong Zhou, et al., "MINR: Efficient Implicit Neural Representations for Multi-Image Encoding," *International Conference on Acoustics, Speech, and Signal Processing*, 2025.
- Wenyong Zhou, et al., "Enhancing Robustness of Implicit Neural Representations Against Weight Perturbations," *International Conference on Acoustics, Speech, and Signal Processing*, 2025.
- Wenyong Zhou, et al., "Binary Weight Multi-Bit Activation Quantization for Compute-in-Memory CNN Accelerators," *IEEE Trans. Computer-Aided Design of Integrated Circuits and Systems*.
- Wenyong Zhou, et al., "Towards RRAM-based Transformer-based Vision Models with Noise-aware Knowledge Distillation," *Design Automation and Test in Europe*, 2025.