ZHAOKUN WANG

AI Engineer | LLM Agents & Enterprise Software

wang@cl.uni-heidelberg.de | +49 152 09723758 | Heidelberg, Deutschland
www.github.com/BufferHund | www.iamzhaokun.com/de/
www.linkedin.com/in/zhaokun-wang-8802452a9

STUDIUM

M.A. Computerlinguistik, Schwerpunkt KI

Universität Heidelberg

Sep. 2024 – Okt. 2026 | Heidelberg, Deutschland

- Aktuelle Durchschnittsnote: **1,5**. Schwerpunkte: Agentic AI, multimodale Sprachmodelle, VLM-basiertes Grounding, wissensgestützte Systeme und Mensch–KI-Interaktion.
- Ausgewählte Lehrveranstaltungen: Multimodal Language Models, Computer Vision, Knowledge Graphs, Temporal Reasoning und AI-Assisted Software Engineering.

B.Eng. Informatik

Central South University of Forestry and Technology

Sep. 2019 – Jun. 2023 | Hunan, China

- Abschlussnote: **1,8**. Fundierte Grundlagen in Algorithmen, Datenstrukturen, Datenbanken und Software Engineering.

BERUFSERFAHRUNG

AI Research Intern / Masterand

Volkswagen AG & Group Innovation

Jan. 2026 – heute | Wolfsburg, Deutschland

- Arbeit mit **Computer Vision und Vision-Language Models** für multimodale KI-Anwendungen, insbesondere in den Bereichen visuelles Verständnis, Grounding und interaktive KI-Systeme.
- Masterarbeit im Bereich **AI Agent Safety und Robustness** mit Schwerpunkt auf der systematischen Evaluation von Agentenverhalten, Interaktion und Zuverlässigkeit.
- Mitarbeit an der Entwicklung eines **Agent Harness** für multimodale KI-Systeme und experimentelle Agent-Workflows.

Software Engineer

Shenzhen S.C New Energy Technology Corporation

Jul. 2023 – Jun. 2024 | Shenzhen, China

- Entwicklung und Wartung von MES-Funktionen mit TypeScript und C# zur Unterstützung von Produktionsabläufen und zur Verarbeitung von Fertigungsdaten.
- Überführung fachlicher Anforderungen aus Produktion und operativen Bereichen in zuverlässige Softwarelösungen für Reporting, Workflow-Konfiguration und Prozessunterstützung.
- Enge Zusammenarbeit mit Produktmanagement, QA und operativen Teams im agilen Entwicklungsprozess, einschließlich Testing, Fehleranalyse und Deployment-Support.

PROFIL

Masterstudent der Computerlinguistik mit Informatik-Hintergrund und Erfahrung in **LLM Agents, multimodaler KI und Software Engineering**. Entwicklung zuverlässiger KI-Anwendungen, die Modelle, APIs, Tools und Enterprise-Workflows miteinander verbinden.

PROJEKTE

Resource-Efficient Qwen Distillation

- Entwicklung einer Teacher–Student-Distillation-Pipeline für **Qwen2.5-0.5B** mit LoRA und 8-Bit-Quantisierung.
- Erreichte **75% der Teacher-Genauigkeit bei 6× geringerem Rechenaufwand**.

Hierarchical Language Model with Cache

- Neuimplementierung eines hierarchischen zeichenbasierten Sprachmodells in **PyTorch** mit vektorisierter Berechnung und kontinuierlichem Cache-Management.

Mehr unter https://iamzhaokun.com/de/research/.

KENNTNISSE

AI Agents & LLM Systems
LLMs, RAG, MCP, Tool Calling, Prompt Engineering, Context & Memory Management

Machine Learning
PyTorch, VLMs, LoRA, Quantization, Knowledge Distillation, Model Evaluation

Programmierung & Backend
Python, Java, TypeScript, C#, SQL, FastAPI, REST APIs, .NET

Cloud & Engineering
Azure, Docker, Linux, Git, CI/CD, VS Code, Cursor

SPRACHEN

Englisch (fließend)
Deutsch (Grundkenntnisse)
Chinesisch (Muttersprache)

VERFÜGBARKEIT

Verfügbar ab Oktober 2026