Qinyuan Wu

(+49) 17685279012 | qwu@mpi-sws.org | qinyuanwu0710.github.io

Research Interests

My research explores the internal mechanics of deep learning models to build more reliable AI systems, with a focus on how Large Language Models (LLMs) encode, recall, and generalize knowledge.

Education

CS@Max Planck & Saarland University Oct 2022 – Present
Ph.D. in Computer Science
Advisor: Prof. Krishna P. Gummadi (MPI-SWS) & Prof. Muhammad Bilal Zafar (Ruhr-Universität Bochum)
University of Electronic Science and Technology of China (UESTC) Sep 2018 – Jun 2022
B.Sc. in Mathematics-Physics Fundamental Science
Yingcai Honors College (top 2% selected)

Work Experience

Max Planck Institute for Software Systems (MPI-SWS) Oct 2022 – Present
Doctoral Researcher
Institute of Fundamental and Frontier Science (IFFS), UESTC Sep 2020 – Jun 2022
Research Intern, Information Fusion and Intelligent Systems Group
Worked on probabilistic and information-theoretic models to improve decision reliability under uncertainty.
Big Data Research Center, UESTC Sep 2019 – Jun 2020
Research Intern, CompleX Group
Analyzed socioeconomic data to model the impact of transportation networks on regional development.

Publications

Conferences

C4. **The Algorithmic Self-Portrait: Deconstructing Memory in ChatGPT.**
The ACM Web Conference (WWW 2026), accept rate: 20.1%
The International Association for Safe & Ethical AI second annual conference (IASEAI 26), non-archival, accept rate: 18.75%
Abhisek Dash*, Soumi Das*, Elisabeth Kirsten*, **Qinyuan Wu***, Sai Keerthana Karnam, Krishna P. Gummadi, Thorsten Holz, Muhammad Bilal Zafar and Savvas Zannettou
* Equal contribution

C3. **In Agents We Trust, but Who Do Agents Trust? Latent Source Preferences Steer LLM Generations.**
The Fourteenth International Conference on Learning Representations (ICLR 2026), accept rate: 28%
The International Association for Safe & Ethical AI second annual conference (IASEAI 26), non-archival, accept rate: 18.75%
Early version in ICML 2025 Workshop: Reliable and Responsible Foundation Models .
Mohammad Aflah Khan, Mahsa Amani, Soumi Das, Bishwamittra Ghosh, **Qinyuan Wu**, Krishna P. Gummadi, Manish Gupta, Abhilasha Ravichander.

C2. **Rote Learning Considered Useful: Generalizing over Memorized Knowledge in LLMs.**
The Fourteenth International Conference on Learning Representations (ICLR 2026), accept rate: 28%
The International Association for Safe & Ethical AI second annual conference (IASEAI 26), non-archival, accept rate: 18.75%
Early version in ICML 2025 Workshop: Impact of Memorization on Trustworthy Foundation Models.
Qinyuan Wu, Soumi Das, Mahsa Amani, Bishwamittra Ghosh, Mohammad Aflah Khan, Krishna P. Gummadi, Muhammad Bilal Zafar.

C1. **Towards Reliable Latent Knowledge Estimation in LLMs: Zero-Prompt Many-Shot Based Factual Knowledge Extraction.**
ACM International Conference on Web Search and Data Mining (WSDM) 2025, accept rate: 17.3%.
Qinyuan Wu, Mohammad Aflah Khan, Soumi Das, Vedant Nanda, Bishwamittra Ghosh, Camila Kolling, Till Speicher, Laurent Bindschaedler, Krishna P. Gummadi, Evimaria Terzi.

Workshops

W3. **Revisiting Privacy, Utility, and Efficiency Trade-offs when Fine-Tuning Large Language Models.**
The International Association for Safe & Ethical AI second annual conference (IASEAI 26), accept rate: 18.75%
Soumi Das, Camila Kolling, Mohammad Aflah Khan, Mahsa Amani, Bishwamittra Ghosh, **Qinyuan Wu**, Till Speicher, Krishna P. Gummadi.

W2. **Rethinking Memorization Measures in LLMs: Recollection vs. Counterfactual vs. Contextual Memorization.**
ICML 2025 Workshop: Impact of Memorization on Trustworthy Foundation Models (in submission as full paper).
Bishwamittra Ghosh, Soumi Das, **Qinyuan Wu**, Mohammad Aflah Khan, Krishna P. Gummadi, Evimaria Terzi, Deepak Garg.

W1. **Testing Memory Capabilities in Large Language Models with the Sequential Ordered Recall Task.**
LatinX in AI @ NeurIPS 2024.
Mathis Pink, Vy A. Vo, **Qinyuan Wu**, Jianing Mu, Javier S. Turek, Uri Hasson, Kenneth A. Norman, Sebastian Michelmann, Alexander Huth, Mariya Toneva.

Preprints

P3. **Position: Episodic Memory is the Missing Piece for Long-Term LLM Agents.**
arXiv preprint, 2025.
Mathis Pink, **Qinyuan Wu**, Vy Ai Vo, Javier Turek, Jianing Mu, Alexander Huth, Mariya Toneva.

P2. **Assessing Episodic Memory in LLMs with Sequence Order Recall Tasks.**
arXiv preprint, 2024.
Mathis Pink, Vy A. Vo, **Qinyuan Wu**, Jianing Mu, Javier S. Turek, Uri Hasson, Kenneth A. Norman, Sebastian Michelmann, Alexander Huth, Mariya Toneva.

P1. **Understanding Memorisation in LLMs: Dynamics, Influencing Factors, and Implications.**
arXiv preprint, 2024.
Till Speicher, Mohammad Aflah Khan, **Qinyuan Wu**, Vedant Nanda, Soumi Das, Bishwamittra Ghosh, Krishna P. Gummadi, Evimaria Terzi.

Before Ph.D.

J1. **Exponential negation of a probability distribution.** *Soft Computing*, 2022. **Qinyuan Wu**, Yong Deng, Neal Xiong.

Talks

Data Science for Humanity Group, MPI-SP (Germany) 2025
Rote Learning Considered Useful: Generalizing over Memorized Knowledge in LLMs
AI and Society Group, Ruhr Universität Bochum (Germany) 2024
Reliable Knowledge Estimation of LLMs

Teaching Experience

Efficient Training of Large Language Models (TA) Winter 2025–26, Saarland University
Systems for Large (Language) Models (TA) Winter 2023–24, Saarland University

Technical Skills

Model Training & Serving: Trained and deployed LLMs with open-source distributed frameworks; optimized inference using vLLM and SGLang; fine-tuned models (LoRA) and integrated HuggingFace pipelines on Slurm-managed multi-GPU clusters; performed data preprocessing (Pandas, NumPy, Polars).
Programming & Tools: Python, C, Bash, Git, Linux.
Languages: English (fluent), Chinese (native), German (beginner)

Awards

First-class Scholarship (Top 10%) 2020–2021
Innovation & Entrepreneurship Project: Excellent (Top 10%, State Level, 20K RMB) 2020–2021
Innovation & Entrepreneurship Project: Excellent (Top 10%, Univ. Level) 2019–2020
Second-class Scholarship (Top 20%) 2019–2020

References

Krishna P. Gummadi
gummadi@mpi-sws.org
MPI-SWS, scientific dirctor

Muhammad Bilal Zafar
bilal.zafar@rub.de
Ruhr University Bochum, professor

Mariya Toneva
mtoneva@mpi-sws.org
MPI-SWS, tenure-track faculty