

Ziwei Zhao

Research Interests: Trustworthy AI; Multimodal & Agentic Systems; AI for Science
Homepage — Google Scholar — GitHub

ziwei.zhao@tum.de
+49-15560679916
Munich, Germany

EDUCATION

- **Technical University of Munich (TUM)** Munich, Germany
B.Sc. in Information Engineering Oct. 2024 – Jun. 2027 (Expected)

RESEARCH VISITS

- **Renmin University of China, Gaoling School of AI (RUC GSAI)** Beijing, China
Visiting Student | Advisors: Prof. Yanbiao Ma, Prof. Wenxuan Wang Jan. 2025 – Present

SELECTED PUBLICATIONS

- [P2] Haojun Liang, **Ziwei Zhao**, Chenyu Li, Qiyue Zhao, Rongduomijia Mu, Yulan Shen, Yu Gu, Chen Zhang, Kairui Zhang, Yanbiao Ma, Wenxuan Wang*. “ScopeBench: Evaluating Cultural Norm Scope Awareness in LLMs.” **Accepted to EMNLP 2026 Main Conference**. [PDF].
- [P5] **Ziwei Zhao***, Yu Gu*, Haojun Liang*, Chen Zhang, Xizhi Ding. “Skill-as-API: Confidential Multi-Agent Coordination for Agentic Software Engineering.” **Accepted to KDD 2026 Workshop**. [PDF].
- [P4] Jiayi Tuo, Cheng Tang, Zihan Wang, Chenyue Zhou, Yao Li, Yanbiao Ma*, Chao Wang, Wei Dai, Mingxuan Wang, Shitong Qin, **Ziwei Zhao**. “MessToClean: Evidence-Grounded Structure-Preserving Reconstruction for Real-World Degraded Exam Paper Images.” **Accepted to ACL 2026 Main Conference** [PDF].
- [P6] Qing Cheng, Wenwen Wu, **Ziwei Zhao**. “Uncertainty-Calibrated UAV Trajectory Prediction for Beam Management in UAV-Assisted ISAC Scenarios.” **Drones**, 10(6), 434, 2026. [Link].
- [P3] Fei Luo, **Ziwei Zhao**, Mingxuan Wang, Duoyang Li, Zhe Qian, Jiayi Tuo, Chenyue Zhou, Yanbiao Ma*. “Geometric Prior-Guided Federated Prompt Calibration.” arXiv:2512.07208, Dec 2025.

MANUSCRIPTS UNDER REVIEW

- [P1] **Ziwei Zhao**, Yanxi Lu, Yuwei Hu, Shiyang Su, Mingxuan Wang, Chenyue Zhou, Jiayi Chen, Hehan Li, Xiaoshuai Hao, Andi Zhang, Yanbiao Ma*. “Information Density Imbalance in Visual Object Detection.” Under review at **Pattern Recognition**, 2025. [PDF].

RESEARCH EXPERIENCE

- **Information Density Imbalance in Visual Object Detection [P1]** Renmin University of China
First Author | Advisor: Prof. Yanbiao Ma Aug. 2025 – Dec. 2025
 - Defined *categorized information density* as a per-class difficulty signal beyond instance frequency. Pearson correlation with AP reaches $\rho \approx -0.72$ on Pascal VOC, about four times stronger than the instance-count baseline ($\rho \approx -0.171$), exposing a bias source long overlooked in long-tailed detector training.
 - Integrated information density into three class re-balancing objectives (IDG-Seesaw, IDG-FL, IDG-C2AM) as drop-in loss modifications that target high-density categories.
 - Designed a local-to-global covariance merging scheme. The merge keeps global information density *exact* (not approximate) during training, improves rare-class detection across Pascal VOC, COCO-LT, and LVIS v1.0, and reduces per-class AP variance (a direct fairness measure) by up to **45.89%** over EFL on LVIS v1.0.
- **Geometric Prior-Guided Federated Prompt Calibration [P3]** Renmin University of China
Second Author | Advisor: Prof. Yanbiao Ma May 2025 – Aug. 2025
 - Co-designed GGTPC to address local training bias in federated prompt learning. The server reconstructs global geometric priors from per-class statistics without exchanging raw features.
 - Validated GGTPC as a plug-and-play module across 7 mainstream FL algorithms (FedAvg, SCAFFOLD, MOON, FedDyn, FedOPT, FedProto, FedNTD). The geometric prior stabilizes accuracy under extreme client heterogeneity, e.g., **+9.17 pp** over FedAvg on CIFAR-100 at $\beta=0.01$ where baselines collapse, and improves cross-client fairness in heterogeneous settings.

- Implemented the prior reconstruction module and led the heterogeneity ablation study across label-skew, feature-skew, and mixed settings.

• **Benchmarks for LLM Alignment & Document AI [P2, P4]** Renmin University of China
Various roles | Advisors: Prof. Wenxuan Wang, Prof. Yanbiao Ma *May 2025 – Present*

- **ScopeBench [P2, EMNLP 2026 Main, Second Author / Core Contributor]:** Cultural-scope safety benchmark treating norm scope as a first-class safety dimension. Built a dual-anchored annotation protocol across 49 countries and 9 cultural clusters (3,323 norms). Defined three scope error types (Inflation, Deflation, Displacement) and evaluated 21 frontier LLMs across 7 model families.
- **MessToClean [P4, ACL 2026 Main, Contributor]:** Evidence-grounded reconstruction pipeline for degraded exam paper images; benchmark construction and OCR-robustness ablations.

• **Uncertainty-Calibrated UAV Trajectory Prediction [P6]** Technical University of Munich
Contributor | Published in Drones (MDPI), 2026 *Jan. 2026 – Apr. 2026*

- Contributed to a multimodal trajectory prediction framework that fuses visual context (ResNet-18) and motion history (LSTM) via cross-attention and gated fusion to output calibrated probabilistic forecasts. Split conformal calibration converts raw Gaussian uncertainty into statistically reliable risk bounds, improving Coverage@95 by 7 pp.
- The calibrated uncertainty serves as a risk signal for downstream control: a codebook-constrained beam-selection policy increases coverage by 24% and reduces catastrophic failure by 43% in high-risk scenarios, demonstrating a closed perception-calibration-control loop transferable to embodied decision-making under uncertainty.

RESEARCH SYSTEMS & OPEN-SOURCE PROJECTS

• **Skill-as-API: Confidential Multi-Agent Coordination [P5]** GitHub | PyPI
Author | Accepted to KDD 2026 Workshop *Jan. 2026 – Apr. 2026*

- Designed the *Skill-as-API* protocol: a coordination layer for agentic SE where a skill’s public view is limited to name, typed schemas, and trust tier; the body (code + system prompt) is closure-captured and never crosses the wire. Four defense layers address IP extraction, prompt theft via injection, and trust drift.
- Released an open-source Python SDK over XMTP with 1.8–2.9s cross-continent reconnect latency. Featured in *awesome-agentic-patterns* (~4.7K stars) as the *Black-Box Skill Invocation* pattern.

• **HeartBridge: MLLM System for Autism Care** Baidu ERNIE Hackathon (Global 3rd Place)
Tech Lead | Multimodal LLM + Risk Prediction *Oct. 2025 – Dec. 2025*

- Delivered an end-to-end multimodal LLM workflow for family support and clinical resource routing across text, image, and sensor inputs; 95% matching success on therapist recommendation.
- Led next-day risk prediction modeling for autistic children by mapping prior-night sleep features to next-day emotion fluctuation and aggressive behavior.

AWARDS & HONORS

- **Global 3rd Place**, Baidu ERNIE Open-Source Hackathon (Dec. 2025) | Tech Lead of HeartBridge project.
- **Global UGRAD Scholarship**, U.S. Department of State, Bureau of Educational and Cultural Affairs. [PDF]
- **The First Prize**, “Hailiang Cup” National Geoscience Olympic Competition for Secondary School Students (2021–2022). [PDF]

TECHNICAL SKILLS

- **Programming:** Python, Java, C/C++, SQL
- **ML Frameworks:** PyTorch, HuggingFace, MMDetection, vLLM
- **Systems & Tools:** Linux, Docker, FastAPI, PostgreSQL, Git, LaTeX

LANGUAGES

Mandarin (native) | English (**TOEFL iBT: 89**) | German (working proficiency)