张博轩

15623251290 zhangboxuan1005@gmail.com https://zbox1005.github.io
性别：男 出生年月：2000.07 民族：汉族 政治面貌：共产党员

教育背景

罗格斯大学（美国） **博士（即将入学）** 2025.09 –
• 学院：计算机科学系 专业：计算机科学 研究方向：可信机器学习，AI 医疗，大语言模型(LLM)

武汉大学 **硕士** 2022.09 – 2024.06
• 学院：计算机学院 专业：人工智能 研究方向：遥感影像目标检测，可信机器学习

武汉大学 **本科** 2018.09 – 2022.06
• 学院：计算机学院 专业：计算机科学与技术 GPA：3.6 / 4.0

论文发表

（*表示同等贡献）

- **CoT-UQ: Improving Response-wise Uncertainty Quantification in LLMs with Chain-of-Thought**
  Boxuan Zhang, Ruqi Zhang
  预印本，在投.
- **Shakespearean Sparks: The Dance of Hallucination and Creativity in LLMs' Decoding Layers**
  Zicong He*, **Boxuan Zhang***, Lu Cheng
  预印本，在投.
- **What if the Input is Expanded in OOD Detection?**
  Boxuan Zhang*, Jianing Zhu*, Zengmao Wang, Tongliang Liu, Bo Du, Bo Han
  The 38th Conference on Neural Information Processing Systems (NeurIPS), 2024, 已发表.
- **Boosting Semi-Supervised Object Detection in Remote Sensing Images with Active Teaching**
  Boxuan Zhang, Zengmao Wang, Bo Du
  IEEE Geoscience and Remote Sensing Letters (GRSL), 2024 , 已发表.

科研经历

伊利诺伊大学芝加哥分校（美国） 计算机科学系 科研助理 2024.11 – 2025.02
指导教师：Prof. Lu Cheng 研究课题：大语言模型中的幻觉与创造力（指导低年级实习生）
- **方法：**提出一种针对 **LLM 创造力的狭义定义**，并提出 **HCL** 评估框架以**量化** LLM 解码过程中不同层的幻觉和创造力。
- **实验：**结果表明，LLM 中**较早的层**会产生更具创造性的输出，而**较深的层**则优先考虑事实的准确性。基于此，我们确定了兼顾这两方面的**最佳解码层**，并引入了**分层提前退出**策略，以在不牺牲输出质量的前提下提高计算效率。

普渡大学（美国） 计算机科学系 暑期科研实习 2024.06 – 2025.02
指导教师：Prof. Ruqi Zhang 研究课题：大语言模型中的不确定性量化
- **动机：**提供**推理路径**能够让模型结合**额外的上下文**进行**置信度校准**，从而对最终答案做出更可信的正确性评估。
- **方法：**提出将 LLM 的推理路径通过**思维链**整合到**不确定性量化**的流程中（**CoT-UQ**），来量化**每个回答**的不确定性。
- **实验：**结果表明，**CoT-UQ** 明显优于现有的不确定性量化方法，平均提高了约 **5.9 个百分点**的 AUROC 值。

香港浸会大学 计算机科学系 科研实习 2023.11 – 2024.05
指导教师：Prof. Bo Han 研究课题：分布外数据检测
- **动机：**现有方法基于单一输入挖掘判别信息，其**表征维度受限**且**无法识别**与 ID 样本特征相似而难以区分的 OOD 样本。
- **方法：**提出**置信度平均评分法（CoVer）**，平均破坏和原始输入的 OOD 得分以捕捉 OOD 和 ID 数据在其中的**动态差异**。
- **实验：**结果表明，**CoVer** 在 ImageNet OOD 检测基准上取得更好的性能，同时**兼容**来自 DNN 或 VLM 架构的检测方法。

武汉大学 计算机学院 研究助理 2022.09 – 2024.06
指导教师：Prof. Zengmao Wang, Prof. Bo Du 研究课题：遥感影像目标检测
- **方法：**提出利用**主动学习**增强**遥感图像**中的**半监督目标检测（SSOD-AT）**，以减轻遥感场景中对**有限标注图像**的依赖。
- **实验：**结果表明，**SSOD-AT** 只需**有限的标注样本**就能达到**全监督**的检测精度，整个主动学习中相较于 SOTA 提升 **2%**。

技能

专业技能：熟练使用 Python；熟悉 PyTorch 深度学习框架；熟悉可信机器学习、（多模态）大语言模型、计算机视觉等领域，熟悉 Transformer 等深度学习模型；了解 Prompt Engineering、RAG、大模型微调、预训练。

Zhang Boxuan

Github zhangboxuan1005@gmail.com
Hompage (+86) 15623251290

RESEARCH INTEREST

Machine Learning	Trustworthy Machine Learning, Active Learning
Large Language Models	Uncertainty Quantification, Reasoning, Trustworthiness (Reliability, Safety)
Applications	Object Detection and Segmentation, Healthcare, Remote Sensing

EDUCATION

Rutgers University — **New Jersey, USA**
Incoming Ph.D. student in Computer Science — *Aug. 2025 - June 2030 (Expected)*
– Advisor: Prof. Ryan Tang

Wuhan University — **Wuhan, CHN**
M.E. in Electronic Information (Artificial Intelligence) — *Sept. 2022 - June 2024*
– Overall GPA: 91.81/100 (**Top 10%**)
– Advisors: Prof. Zengmao Wang and Prof. Bo Du

Wuhan University — **Wuhan, CHN**
B.E. in Computer Science and Technology — *Sept. 2018 - June 2022*
– Overall GPA: 3.6/4.0

PUBLICATIONS

(* Indicates Equal Contribution)

- **CoT-UQ: Improving Response-wise Uncertainty Quantification in LLMs with Chain-of-Thought**
  Boxuan Zhang and Ruqi Zhang
  Submitted to ACL ARR 2025 February.
- **Shakespearean Sparks: The Dance of Hallucination and Creativity in LLMs' Decoding Layers**
  Zicong He*, **Boxuan Zhang***, and Lu Cheng
  Submitted to ACL ARR 2025 February.
- **What If the Input is Expanded in OOD Detection?**
  Boxuan Zhang*, Jianing Zhu*, Zengmao Wang, Tongliang Liu, Bo Du, and Bo Han
  In *Advances in Neural Information Processing Systems* (***NeurIPS***), 2024.
- **Boosting Semi-Supervised Object Detection in Remote Sensing Images with Active Teaching**
  Boxuan Zhang, Zengmao Wang, and Bo Du
  In *IEEE Geoscience and Remote Sensing Letters* (***GRSL***), 2024.

RESEARCH EXPERIENCE

Department of Computer Science, University of Illinois Chicago — **Chicago, USA**
Research Assistant, Advisor: Prof. Lu Cheng — *Nov. 2024 - Now*

- Research on *examining the relationship between hallucination and creativity in LLMs.*
- Given the philosophical nature of creativity, we propose a narrow definition tailored to LLMs and introduce an evaluation framework, HCL, which quantifies **H**allucination and **C**reativity across different **L**ayers of LLMs during decoding.
- Co-mentored a junior research intern, providing guidance on experiment design, coding implementation, and paper writing. Submitted one paper to ACL ARR 2025 (February), currently under review.
- The code is publicly available at: github.com/ZicongHe2002/HCL-Spark

Department of Computer Science, Purdue University **West Lafayette, USA**
Summer Research Intern, Advisor: Prof. Ruqi Zhang *June. 2024 - Feb. 2025*

- Research on *Uncertainty Quantification in Large Language Models.*
- Propose to quantify response-wise uncertainty by integrating LLMs' inherent reasoning capabilities through Chain-of-Thought (CoT) into the UQ process. The proposed CoT-UQ significantly outperforms existing UQ methods, achieving an average improvement of 5.9% AUROC.
- Submitted one paper to ACL ARR 2025 (February), currently under review.
- The code is publicly available at: github.com/ZBox1005/CoT-UQ

TMLR Group, Hong Kong Baptist University **Hong Kong, CHN**
Research Intern, Advisor: Prof. Bo Han *Nov. 2023 - May. 2024*

- Research on *Out-of-Distribution (OOD) Detection for Trustworthy Machine Learning.*
- Propose a novel perspective to employ different common corruptions on the input space to expand the representation dimension for OOD detection. With the expectation among multiple input dimension, our method performs a better ID-OOD separability.
- Submit one paper as co-first author to NeurIPS 2024 and get accepted.
- The code is publicly available at: github.com/tmlr-group/CoVer

Department of Civil Engineering, Wuhan University **Wuhan, CHN**
Research Intern, Advisor: Prof. Xiaoping Zhang *Aug. 2023 - Oct. 2023*

- Research on *Machine Learning for Tunnel Boring Machine (TBM) Excavation.*
- Work on rock mass accurate classification based on multi-algorithm cross multi-feature optimization selection and TBM parameter efficient prediction using low-dimensional inputs. This helps to predict geological conditions in advance and the optimal operational parameters under geological variations.
- Complete a technical paper [link] and win the national third prize in the Second TBM Excavation Parameter Data Sharing and Machine Learning Competition.
- The code is publicly available at: github.com/ZBox1005/TBM-Competition

Sensing IntelliGence and MAchine learning(SIGMA) Lab, Wuhan University **Wuhan, CHN**
Research Assistant, Advisors: Prof. Zengmao Wang and Prof. Bo Du *Nov. 2022 - Aug. 2023*

- Research on *Active Learning for Semi-Supervised Object Detection in Remote Sensing Images.*
- Propose to boost semi-supervised object detection with active teaching (SSOD-AT) in remote sensing images. SSOD-AT can achieve high detection accuracy only with limited labeled samples, which helps to alleviate the dependency on limited labeled images in remote sensing scenarios.
- Submit one paper to IEEE Geoscience and Remote Sensing Letters (GRSL) and get accepted.
- The code is publicly available at: github.com/ZBox1005/SSOD-AT

AWARDS AND HONORS

National Third Prize, The Second TBM Excavation Machine Learning Competition *Oct. 2023*

SERVICES

Undergraduate Student Mentor, Wuhan University *Sept. 2022 - June. 2023*

- Facilitated freshmen's transition to university life at School of Computer Science.

TECHNICAL SKILLS

Programming Languages Python (Pytorch), C/C++, Linux, Git, LaTeX
English (professional working proficiency), Chinese (native proficiency)