

何炜程 | WEICHENG HE

可信人工智能 | 人工智能安全 | 计算机视觉

法国巴黎 | weicheng.he@outlook.com | +33 0664856929 | GitHub: github.com/hugoQAQ

个人简介

计算机视觉与可信人工智能方向博士研究生，研究部署中的视觉模型何时失效、如何实时发现风险，以及如何降低错误输出。已在 YOLO、Faster R-CNN 和 Vision Transformer 等不同架构上验证通用监控与缓解方法，将模型幻觉造成的误检降低 90% 以上。此前参与与西门子合作的欧洲研究项目，负责 AI 组件开发、可验证仿真集成及工业用例落地。持有 AWS Certified Machine Learning Engineer - Associate 认证。

工作经历

计算机视觉与可信人工智能博士研究生 | 法国国家科学研究中心 Verimag 研究所 | 2023.05 至 2026.11

- 问题：实时目标检测模型在未知、危险或分布外场景中会产生幻觉与错误检测，且不同架构的失效模式并不一致。
- 方法：设计面向多种目标检测架构的通用监控与缓解方法，并建立可复现的基准、对照实验与逐类误差分析流程。
- 结果：在 YOLO、Faster R-CNN 与 Vision Transformer 上将模型幻觉造成的误检降低 90% 以上；成果发表于 TPAMI、IROS 与 RV。

应用人工智能研发工程师 | 法国国家科学研究中心 Verimag 研究所 | 2021.11 至 2023.04

- 问题：在西门子作为工业合作方的欧洲研究项目中，开发可在真实仿真流程中验证的 AI 驱动自动驾驶系统。
- 方法：开发 AI 组件并接入协同仿真架构，包括自动驾驶场景的实时交通标志识别模型，以及面向医疗信号分析的循环神经网络。
- 结果：研究成果集成至西门子自动驾驶解决方案框架；相关论文录用于 北约 Modelling and Simulation for Autonomous Systems

研究工程师实习生 | 华为法国格勒研究所费马实验室 | 2021.04 至 2021.10

- 问题：将 AI 组件纳入形式化语言仿真框架，并与西门子协同仿真框架连接，使混合系统能够被复现和验证。
- 方法：扩展形式化建模框架以兼容 AI 组件与传统模型组件；开发适用于多种天气条件的限速标志识别模型。
- 结果：限速识别准确率达到 95%，执行速度提升 10 倍；该可验证 AI 用例被纳入西门子协同仿真框架的官方用例。

教育背景

计算机科学博士 | 法国格勒诺布尔阿尔卑斯大学 (UGA) | 2023 至 2026

课题：目标检测算法运行时监控方法的设计与评估。

数据科学硕士 | 法国南锡国立高等矿业学校 (Mines Nancy) | 2020 至 2021

机器学习、深度学习、优化、高维统计与大数据架构。

法国通用工程师学位 (硕士) | 法国南锡国立高等矿业学校 (Mines Nancy) | 2018 至 2020

材料科学硕士 | 中国南京航空航天大学 | 2014 至 2018

专业认证

AWS Certified Machine Learning Engineer - Associate (MLA)

竞赛荣誉

2022.08 | Kaggle NLP 竞赛银牌 - Predicting Effective Arguments

融合句级分类器与 token 级分类器构建集成系统；在 1,557 支队伍中排名第 34，前 3%

2022.03 | Kaggle NLP 竞赛银牌 - Evaluating Student Writing

集成 Longformer、DeBERTa 等 Transformer 模型；在 2,058 支队伍中排名第 51，前 3%

核心能力

可信人工智能与 AI Safety：分布外检测、模型幻觉缓解、运行时监控、形式化验证、基准设计、误差分析

计算机视觉：目标检测、图像分类、实时视觉、鲁棒性评估、YOLO、Faster R-CNN、Vision Transformers、OpenCV

机器学习工程：Python、C/C++、PyTorch、Scikit-Learn、Pandas、NumPy、SQL、Git、Shell、AWS

语言：中文母语；英语熟练；法语熟练

代表性论文

2026 | Revisiting Out-of-Distribution Detection in Real-time Object Detection: From Benchmark Pitfalls to a New Mitigation Paradigm | IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)

2025 | Mitigating Hallucinations in YOLO-based Object Detection Models: A Revisit to Out-of-Distribution Detection | IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

2024 | BAM: Box Abstraction Monitors for Real-time OoD Detection in Object Detection | IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)

2024 | Box-based Monitor Approach for Out-of-Distribution Detection in YOLO: An Exploratory Study | International Conference on Runtime Verification (RV)