常文瀚

电话：+86 136-7217-3424 · **邮箱**：whchang@cug.edu.cn · **主页**：changwenhan.github.io

个人介绍

出生年月：2000 年 5 月　　**籍贯**：天津市滨海新区　　**英语水平**：CET6 - 498　　**GPA**：3.70

本人在校成绩优秀、乐观向上，工作负责、自我驱动力强、热爱尝试新事物。在校期间长期从事人工智能安全与隐私保护领域的研究，目前关注于大语言模型及图像分类模型的安全问题。攻读硕士学位期间主要研究基于数据重组和模型编辑的机器遗忘学习技术。

教育背景

中国地质大学（武汉），计算机技术, *硕士研究生*　　2022.9 - 2025.6

在校期间长期从事人工智能安全与隐私保护领域的研究，毕业论文题目为《基于数据重组和模型编辑的机器遗忘学习方法研究》。

中国地质大学（武汉），计算机科学与技术, *本科*　　2018.9 - 2022.6

在大学本科期间完成了计算机科学与技术专业的学习，包括对于软件开发、人工智能技术和信息安全等领域的学习。

学术成果

• **Gradient-based Defense Methods for Data Leakage in Vertical Federated Learning**.
Wenhan Chang, Tianqing Zhu*.
Computers & Security, 2024, 139: 103744.　　CCF-B

• **Class Machine Unlearning for Complex Data via Concepts Inference and Data Poisoning**.
Wenhan Chang, Tianqing Zhu*, Heng Xu, Wenjian Liu, Wanlei Zhou.
IEEE Transactions on Dependable and Secure Computing, 2024.　　CCF-A (Under Review)

• **Zero-shot Class Unlearning via Layer-wise Relevance Analysis and Neuronal Path Perturbation**.
Wenhan Chang, Tianqing Zhu*, Yufeng Wu, Wanlei Zhou.
IEEE Transactions on Information Forensics and Security, 2024.　　CCF-A (Under Review)

• **Generative adversarial networks unlearning**.
Hui Sun, Tianqing Zhu*, **Wenhan Chang**, Wanlei Zhou.
IEEE Transactions on Dependable and Secure Computing, 2023.　　CCF-A (Under Review)

• **A two-stage model extraction attack on GANs with a small collected dataset**.
Hui Sun, Tianqing Zhu*, **Wenhan Chang**, Wanlei Zhou.
Computers & Security, 2023, 137, 103634.　　CCF-B

• **Model poisoning defense on federated learning: A validation based approach**.
Yuao Wang, Tianqing Zhu*, **Wenhan Chang**, Sheng Shen, Wei Ren.
International Conference on Network and System Security, 2020, 207-223.

• **基于 ARTMA 模型的新冠肺炎疫情预测模拟仿真软件 V1.0**.
常文瀚，李思慧，聂坤宇，张如甜.
计算机软件著作权，登记号为 2021SR1072841，2021.

学术实践

- 主讲 - 澳门城市大学数据科学学院学术交流工作坊 - 建立你自己的大模型：微调，2024 年 4 月
  主讲人：常文瀚，*https://fds.cityu.edu.mo/list-69/353*
  重点介绍了大模型微调的科技底层架构、中间发展以及未来展望。讲解了大语言模型在当前社会中的广泛应用，並强调了由于计算资源和时间成本的限制，对这些大型模型进行微调以用于下游任务变得愈发重要。
- 参与 - The 25th International Conference on Information and Communications Security，2023 年 11 月
  学习并交流人工智能安全领域最新技术。

科研项目经历

- **参与 - 国家自然科学基金面上项目：跨域网络空间动态隐私保护方法研究.**

在该项目中，本人深入研究了大模型的机器遗忘学习，提出了一种高效的隐私保护策略，显著提升了对敏感信息的防护能力。通过创新性的数据重组技术，团队实现了敏感信息的精准识别与有效保护，并提升了模型的隐私安全性。此外，我们还结合模型编辑技术优化了机器遗忘学习的操作流程，为动态数据管理提供了更高的灵活性与可控性。 2019-2023

- **参与 - 中国地质大学（武汉）第十二期校级“英才工程—科学家计划”暨争先奖学金: 基于联邦学习的模型攻击与防护**

在此项目中，本人重点研究了联邦学习环境下的模型攻击，系统性分析了攻击带来的潜在安全威胁和风险。通过提出一种创新的防护机制，我们有效增强了模型在面对多种类型攻击时的鲁棒性与安全性。研究成果不仅显著提高了联邦学习模型的抗攻击能力，还为后续的隐私保护技术研发提供了理论支持和实践参考。 2021-2022

- **主持 - 中国地质大学（武汉）计算机学院科研立项：基于联邦学习的数据隐私保护.**

在该科研项目中，本人带领团队研究了联邦学习环境中的数据隐私保护问题，深入分析了数据中毒攻击的影响及其作用机制。通过对多种攻击方式进行系统研究，团队成功设计并提出了一种具有高效防护性能的防御方法，显著提升了模型在数据中毒环境下的安全性。研究成果为保障联邦学习系统中的数据隐私安全提供了重要的技术支撑和实践依据。 2019-2020

技术能力

- **编程语言**: 熟练使用 Python, 了解 C++、JavaScript 开发
- **开发框架**: 熟练使用 PyTorch, TensorFlow, Docker 进行研究实验，熟练运用 Qt, PyQt 进行桌面端应用开发，熟练运用 ArcGIS 针对 Web 前端 Map 开发。
- **先修课程**: 信息安全理论与技术、数据挖掘与机器学习、大数据驱动建模及科学计算等。