窦一蒲

15533612980 | douyipu@gmail.com | douyipu.github.io
江苏省南京市江宁区东南大学九龙湖校区

教育背景

东南大学（硕士） 网络空间安全学院 2023/09 - 2026/06

- 研究领域：大语言模型安全，包括越狱、提示词注入等

东南大学（学士） 网络空间安全学院 2019/09 - 2023/06

- 人工智能（研讨、全英）95/100，软件安全与恶意代码分析（企业课程）98/100
- 使用 sklearn 实现多种机器学习算法（如决策树、随机森林等）对恶意文件的 IOC 进行分类和对比分析，评估不同模型在恶意代码检测中的效果

技能

- **编程语言**: Python
  - 精通面向对象编程、函数式编程，熟练使用标准库和常用第三方库（如 NumPy, Pandas）
  - 有丰富的数据处理、分析和可视化经验，能够编写高效的代码，熟悉性能优化技巧
- **深度学习框架**: PyTorch
  - 对 PyTorch 内部机制有深入理解，曾为 PyTorch GitHub 仓库贡献问题解决方案
  - 熟悉张量操作，理解自动微分、反向传播等核心概念，深入实践了 Andrej Karpathy 的 micrograd 和 makemore 项目，深入理解 N-gram，MLP 模型机制

项目经验

浦源大模型挑战赛（安全可信赛道） 2024

- 深入研究了大语言模型的安全性问题，掌握了多种攻击和评估方法，如提示词后缀、语言混杂、设置回答开头、设置受害者身份等技术
- 在比赛过程中，系统地测试和分析了不同大模型（GPT 4o、Claude 3.5 Sonnet、文心一言）面对复杂攻击场景时的安全性表现，根据比赛提供的 50 个恶意提示词，在 GPT 4o 上取得了 86% 的平均成功率（每个问题测试 5 次），积累了宝贵的实践经验

PyTorch 开源贡献 2024

- 解决 GitHub issue #126625：深入分析了 view() 和 reshape() 方法在处理非连续张量时的行为差异
- 提供了详细的技术解释和解决方案，帮助其他开发者理解问题根源

Zeek 网络协议解析器开发 2023

- 使用 Spicy 框架独立开发 MQTT 和 CoAP 协议解析器：GitHub: douyipu/spicy-coap
- 实现网络协议的多层次解析，包括报文分析、字段提取、日志生成和事件处理

实习经历

护网裁判 国家互联网应急中心江苏分中心 2023/07 - 2023/08

- “网安 2023”南京行动暨“蓝剑护网 2023”专项行动，审核漏洞和风险数量 500+
- 复现漏洞，包括 SQL 注入、文件上传等漏洞，根据 CVSS 标准对漏洞进行评级

安全工程师 上海煜日有限公司 2021/03 - 2021/06

- 使用开源图像数据集和对抗样本生成算法，完善对抗样本数据库

奖项

网络安全学院学生创新资助计划 2022
山石网科奖学金 2021