2025-06-13周报

马嘉润

1.扩充训练集样本

筛选有效数据源和博客，继续扩充大模型训练集的样本。目前更新和爬取功能在正常运行，之前有一部分数据源的博客爬取没有正常保存，现在排查了问题，可以正常全量爬和每日更新。

目前训练集样本量扩充了2035。

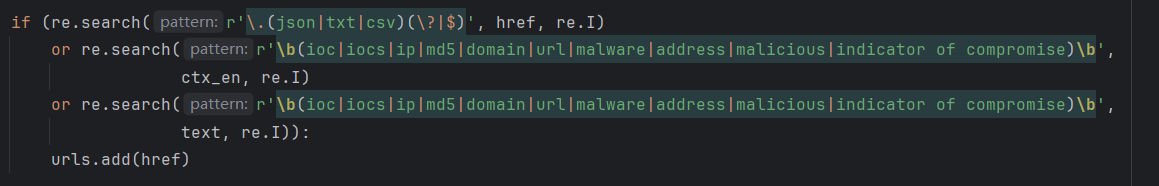
后续进展：

先将筛选出来的训练集样本的正文数据进行大模型的问答，爬虫的全量爬和每日更新都在运行，后续在补充进训练集里面。

曾家豪-本周周报

1. 完善提取文章中有关ioc的url

根据上周老师您给的相关关键字，扩充了相关的关键字。



逻辑为如果本链接是以json、csv、txt为后缀，或者链接的文本或上下文包含ioc关键字，则直接加入候选url。



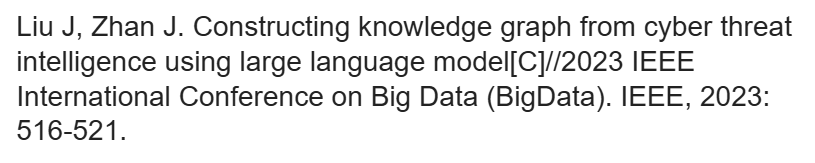
然后如果链接包含github字样或以json、csv、txt为后缀则直接获取该内容嵌入到原本链接的位置。如果不是，则需要判断返回的内容中是否包含ioc相关字样



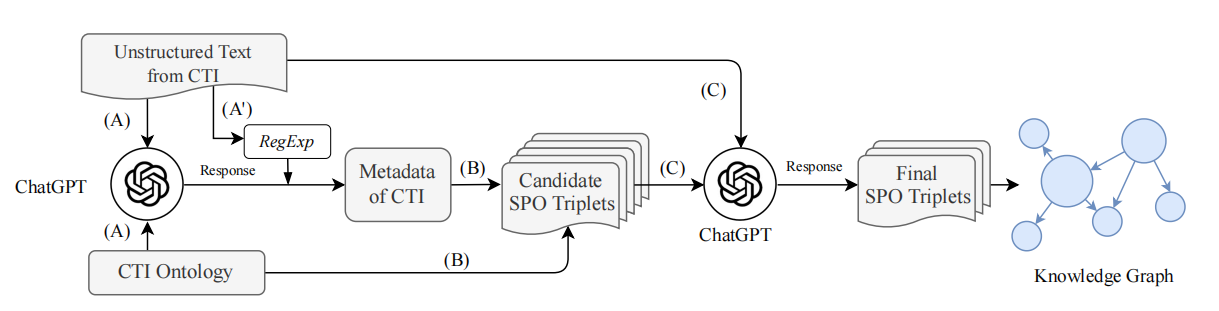
1. 1w7数据的爬取，上周相关的爬取代码出现了问题，这周做了排查，当前正在进行数据的爬取。

冯世玲-本周周报

1、



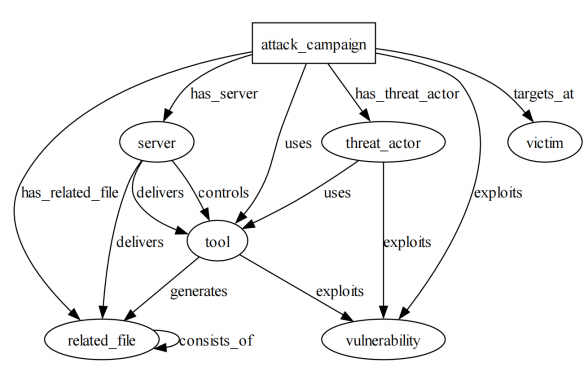
B会

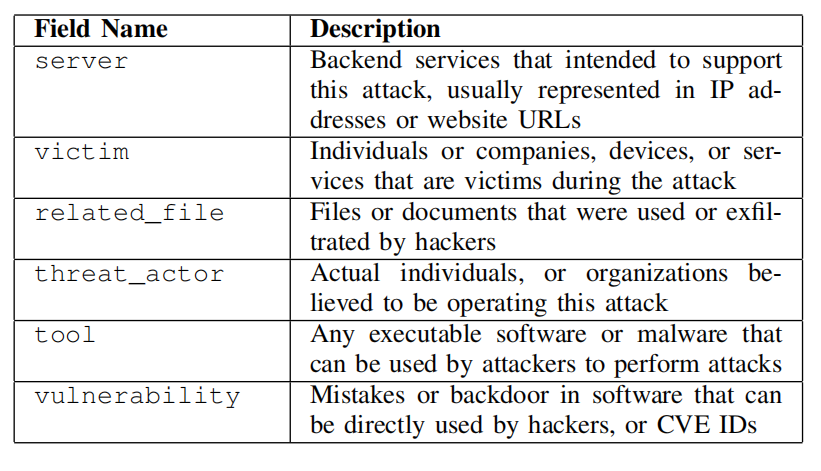


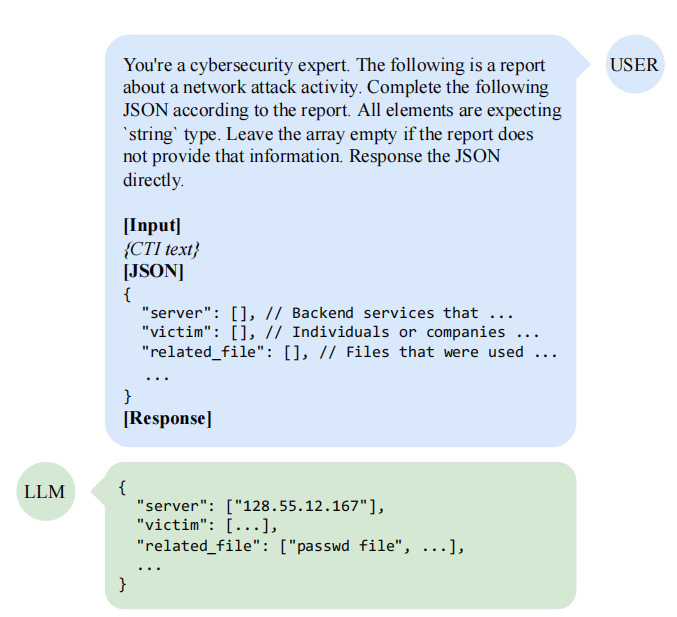
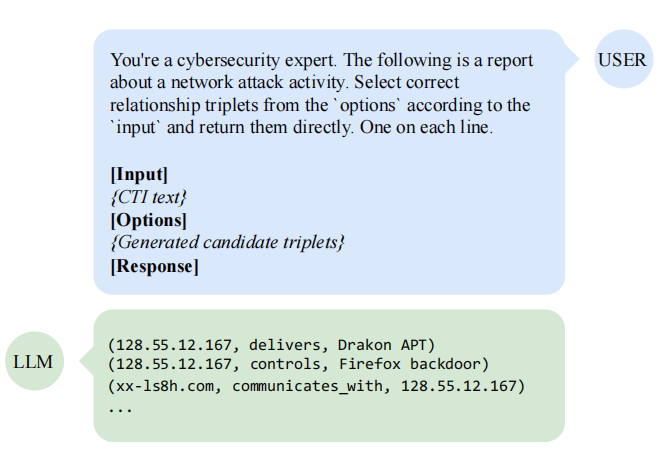
1. 指示ChatGPT使用为CTI设计的预定义本体从CTI中提取元数据。

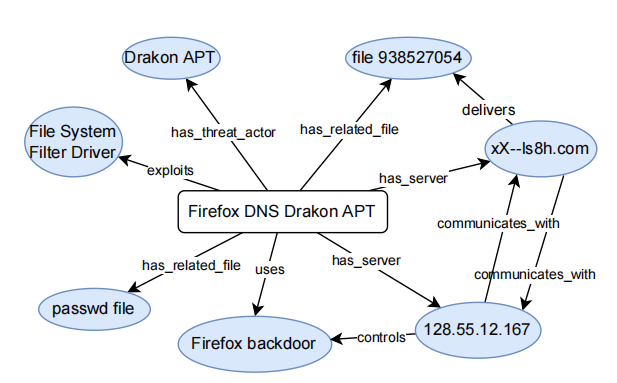
(A′)使用正则表达式（RegExp）提取IoCs，并将结果与ChatGPT的响应合并。

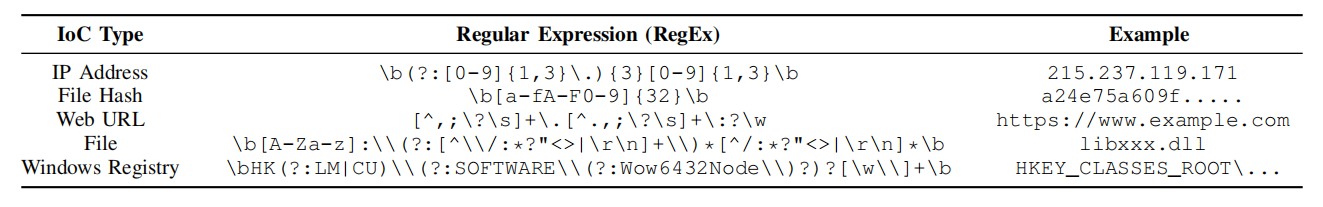
1. 基于元数据（注释了类型的实体）和本体生成潜在的SPO三元组。
2. (C)再次指示ChatGPT从候选中选择正确的SPO三元组，基于CTI。



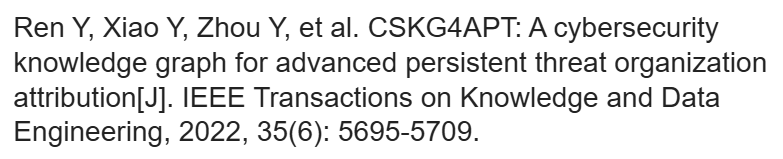








2



a会

CSKG4APT 模型构建

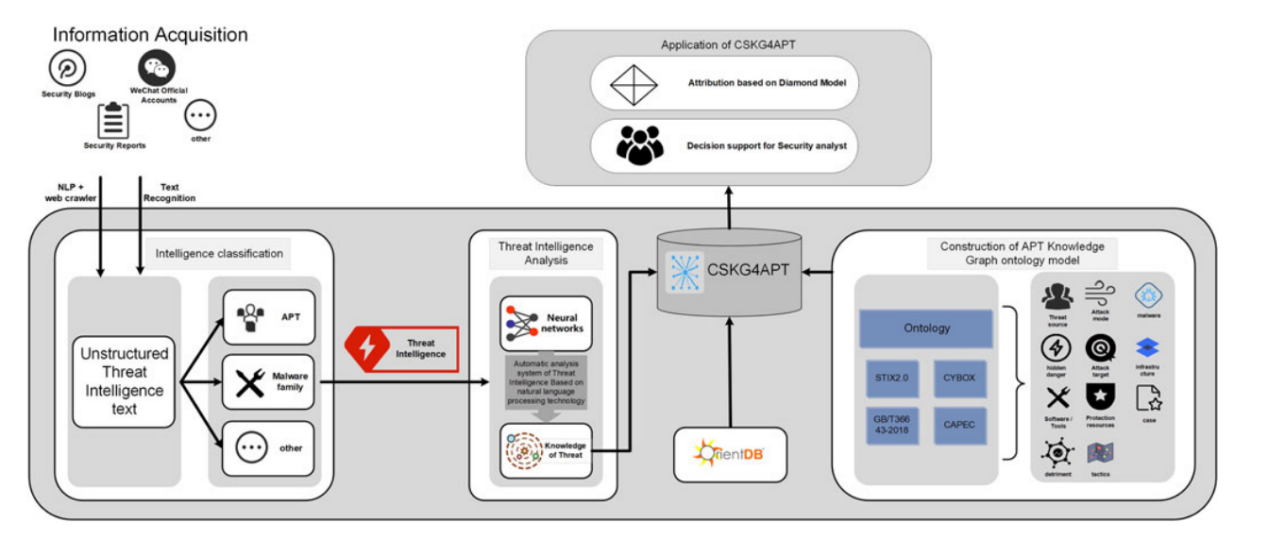
CSKG4APT 是一个面向 APT（高级持续性威胁）组织归属分析的网络安全知识图谱模型，基于本体论理论构建，整合了 OSCTI、STIX、CYBOX 等威胁情报标准的数据。它组织了战术集（攻击行为）、武器集（攻击工具与漏洞）和资产集（关键目标），用于全面表示APT相关知识。

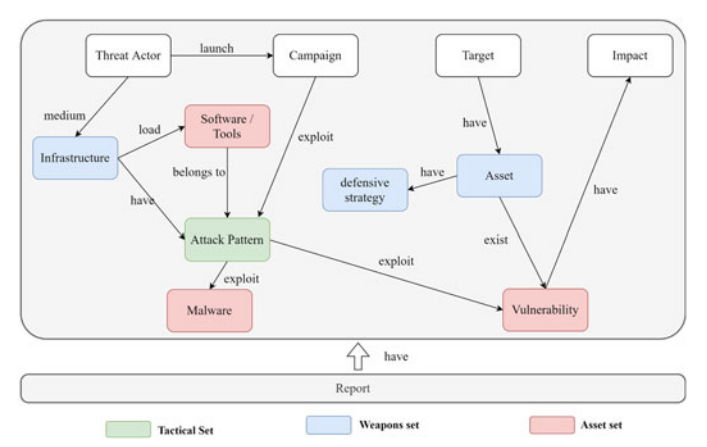
威胁知识的自动提取方法

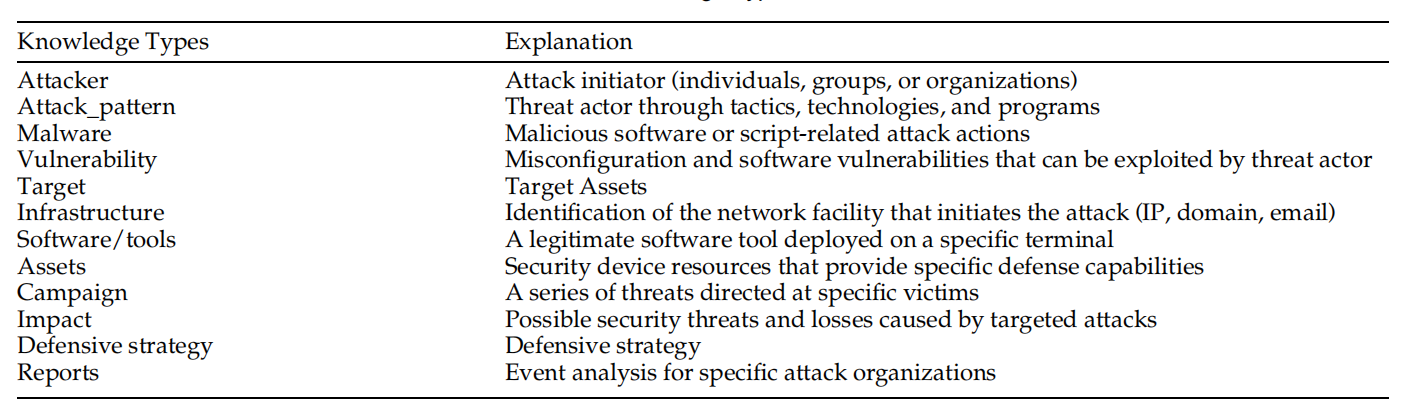
为实现图谱的动态更新，提出了一种基于 BERT 模型的双向编码器算法，能够从中英文威胁情报文本中自动识别出相关实体。该方法在宏观与微观平均指标上均优于现有算法，体现了出色的实体识别效果。

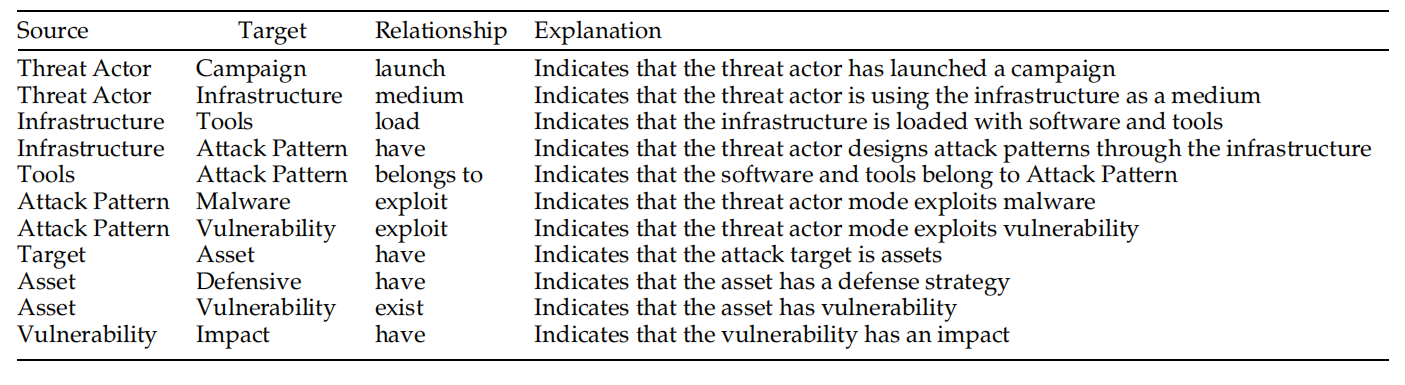
实用应用与测试结果

结合 CSKG4APT 和大规模网络安全数据，设计了一个威胁狩猎分析引擎，利用“钻石模型”对APT组织特征进行建模与追踪。测试表明，该方法有助于提升网络防御能力，具有实战应用价值。

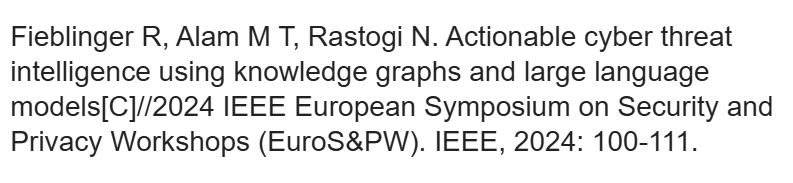


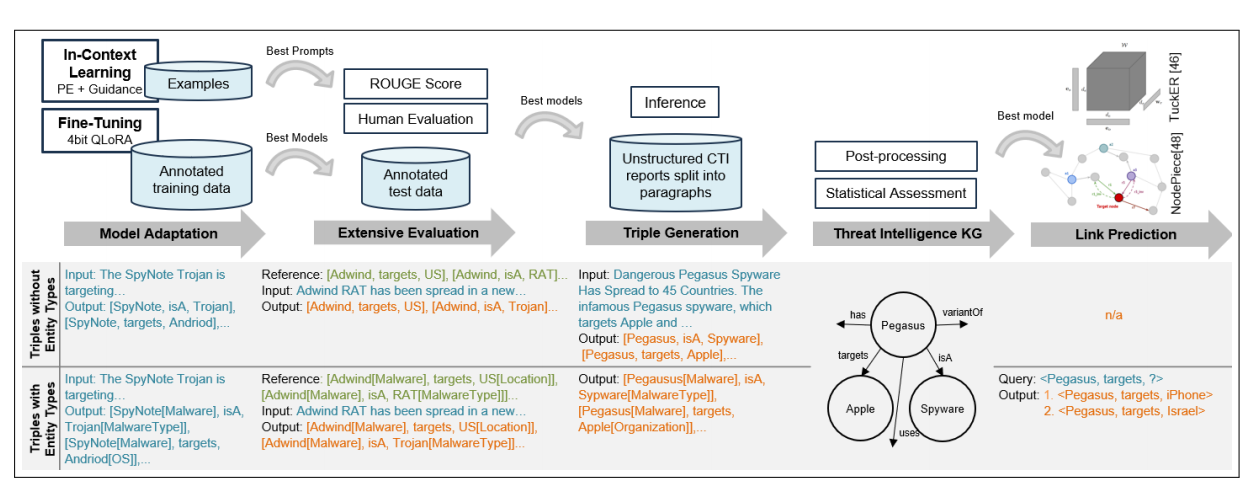




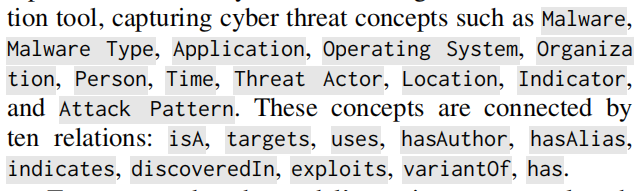


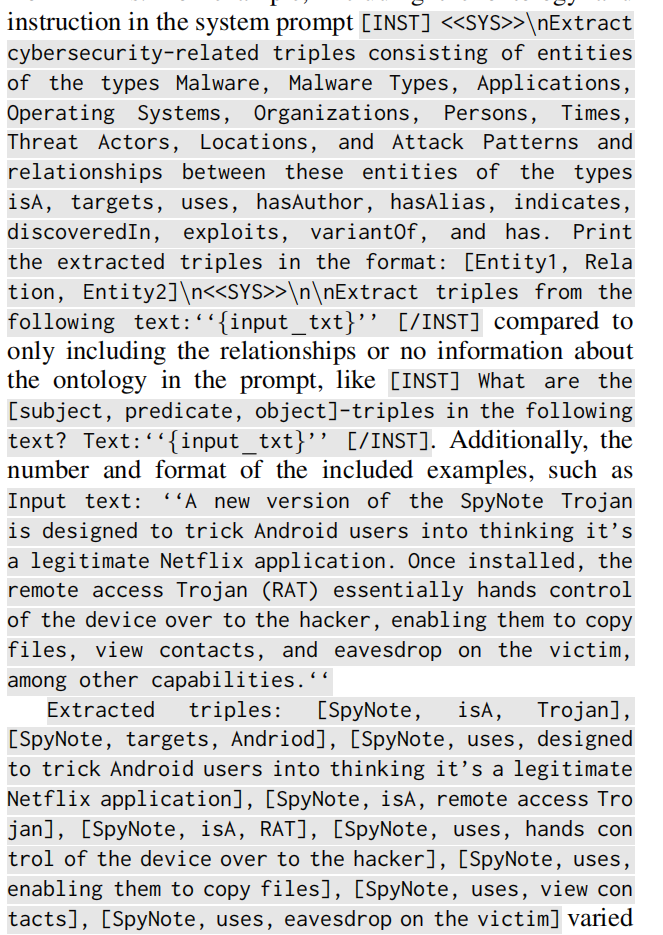
3



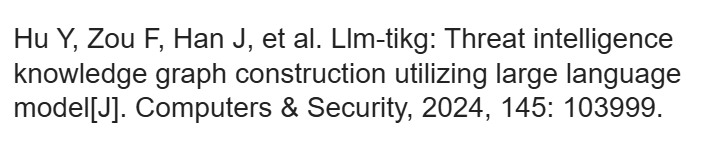


基于语言模型的CTI提取和知识图谱（KG）开发框架。初期阶段，通过少量提示和微调来调整模型，以生成三元组输出。随后，通过广泛的评估确定最佳的模型和提示组合。接着，使用这些顶级模型在KG中生成三元组，从而在最优的KG上进行链接预测，展示从原始数据到可操作情报的转变过程。

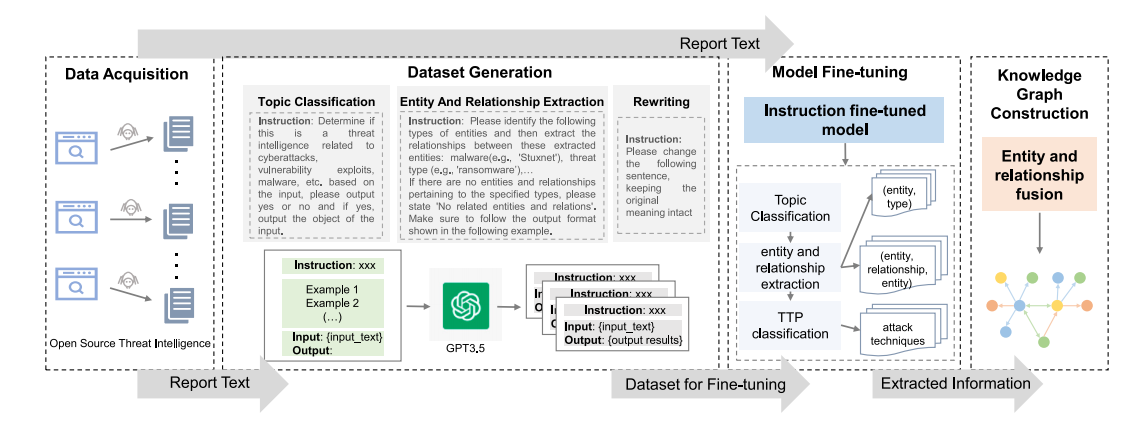




4

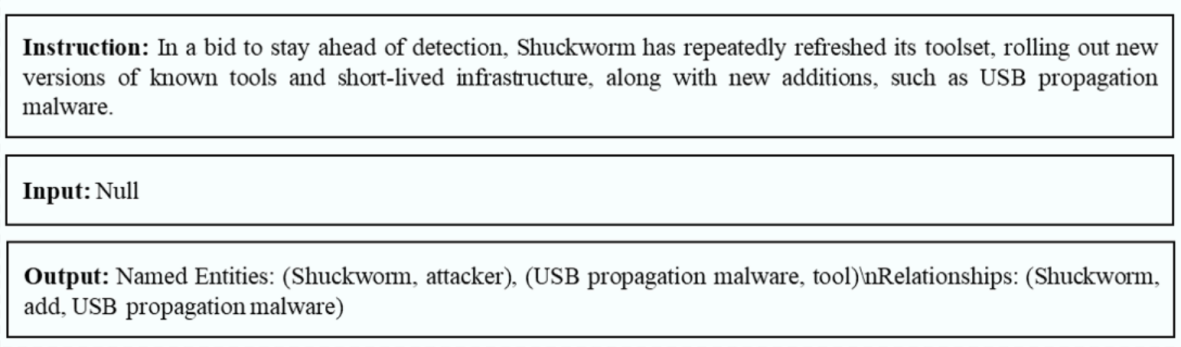


B会

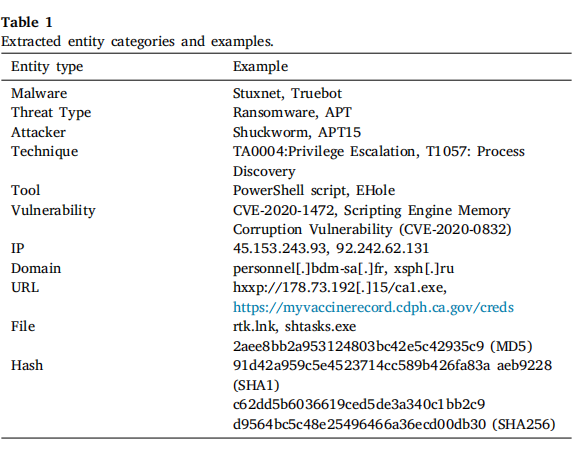


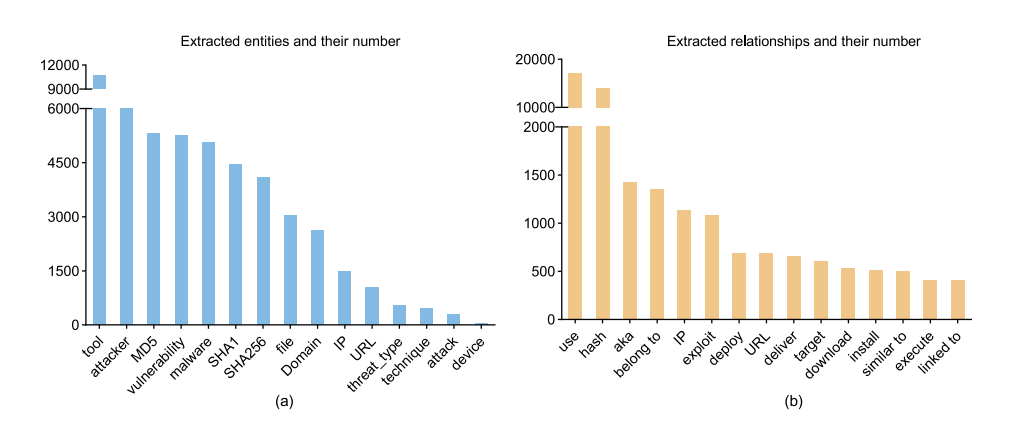
依靠微调的大模型提取实体和关系

用于微调的数据集公开可查

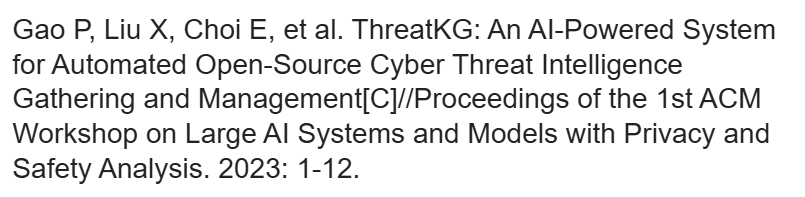


设计的实体和关系如下





5



**ACM CCS（国际计算机安全大会）** 的官方附属研讨会

**THREATKG** 是一个包含约 26K 行代码的自动化系统，用于从多源收集和管理 OSCTI（开放源网络威胁情报）报告，通过过滤、AI 提取等步骤构建并持续更新威胁知识图谱。

系统采用**分层威胁知识本体**建模，涵盖丰富的实体与关联关系，揭示威胁行为与其背景信息。通过**源相关的解析**与**源无关的知识抽取**结合，适应多样化数据格式，并利用**深度学习与数据编程**方法提升提取精度与泛化能力。

THREATKG 拥有**模块化、可扩展的架构**，支持组件的灵活关闭、更新与添加，实现高效并行处理，并能够快速整合新的情报来源或知识类型。

开发了两个典型应用场景：  
(1) **图形用户界面**：用于图谱的可视化、搜索与交互；  
(2) **问答系统**：通过自然语言查询实现威胁知识的智能获取。

