# 面向威胁情报的大语言模型技术应用综述

崔孟娇[1,2], 姜政伟[1,2], 陈奕任[1,2], 江 钧[1], 张 开[1], 凌志婷[1],
封化民[2,3], 杨沛安[1]

[1] 中国科学院信息工程研究所 北京 中国 100093
[2] 中国科学院大学网络空间安全学院 北京 中国 100049
[3] 北京电子科技学院 北京 中国 100070

**摘要** 随着计算机与网络技术的不断发展, 网络空间面临着日益复杂的安全威胁。为了有效防御网络攻击, 网络威胁情报应运而生。然而当前网络威胁如零日漏洞、高级可持续性威胁(Advanced Persistent Threat, APT)等, 具有形式复杂、针对性强、危害性高、隐蔽性强, 时间跨度长等特征, 传统的威胁情报技术难以有效应对。近年来, 大语言模型(Large Language Models, LLM)的兴起不仅降低了攻击成本, 还促进了网络攻击技术的普及化。因此, 本文旨在通过探讨大语言模型在威胁情报领域的技术应用现状, 利用大语言模型的潜能提高对威胁情报聚合, 分析及应用的能力, 从而更为精准地识别、分析和应对网络威胁。本文首先概述了网络威胁情报背景知识, 接着介绍大语言模型的概念、发展历程和研究现状, 以发掘大语言模型在威胁情报领域应用的可能。然后深入分析了威胁情报与大语言模型结合的相关文献, 围绕威胁情报生命周期系统地梳理了大语言模型在增强威胁情报聚合、驱动威胁情报分析以及赋能威胁情报应用方面的成果, 并从技术应用场景和主要方法等角度对其进行分类归纳。此外, 针对这三个方面分别总结了研究现状、技术特点和潜在发展方向。最后本文讨论了大语言模型应用于威胁情报和网络安全领域所面临的挑战, 并给出了未来研究方向, 进一步推动网络威胁情报的发展。

**关键词** 网络威胁情报; 大语言模型; 情报聚合; 情报分析; 情报应用
**中图法分类号** TP391.1 **DOI 号** 10.19363/J.cnki.cn10-1380/tn.2024.09.09

# Applications of Large Language Models Technology for Threat Intelligence: A Survey

CUI Mengjiao[1,2], JIANG Zhengwei[1,2], CHEN Yiren[1,2], JIANG Jun[1], ZHANG Kai[1], LING Zhiting[1],
FENG Huamin[2,3], YANG Peian[1]

[1] Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100093, China
[2] School of Cyber Security, University of Chinese Academy of Sciences, Beijing 100049, China
[3] Beijing Electronic Science & Technology Institute, Beijing 100070, China

**Abstract** With the continuous development of computer and network technology, cyberspace faces increasingly complex security threats. To effectively defend against cyber attacks, cyber threat intelligence has emerged. However, the current network threats such as zero-day vulnerability and Advanced Persistent Threat (APT) are characterized by their complex form, strong targeting, high harm, high covert, and long time span, which are difficult to be effectively dealt with by the traditional threat intelligence technology. In recent years, the rise of Large Language Models (LLM) has not only reduced the costs of attacks but also facilitated the widespread adoption of cyber attack techniques. Therefore, the goal of this article aims to explore the current state of technology application of LLM in the field of threat intelligence and to utilize the potential of LLM to improve the ability to aggregate, analyze, and apply threat intelligence, so as to identify, analyze, and respond to cyber threats more accurately. This paper first outlines the background knowledge of cyber threat intelligence and then introduces the concept, development history, and research status of large language models to explore the possibility of applying large language models in the field of threat intelligence. Then, we analyze in-depth the relevant literature on the combination of threat intelligence and large language model. Around the threat intelligence life cycle, we systematically combine the results of the large language model in enhancing threat intelligence aggregation, driving threat intelligence analysis, and empowering threat intelligence application, and categorize them from the perspectives of technical application scenarios and main methods. In addition, the research status, technical characteristics and potential development directions are summarized for each of these three aspects. Finally, this paper discusses the challenges faced by the

**通讯作者:** 杨沛安, 博士, 高级工程师, Email: yangpeian@iie.ac.cn。
本课题得到中科院战略先导项目课题(No. XDC02030200), 国家自然科学基金(No. 62202466), 中国科学院青年创新促进会(No. 2020166), 中科院网络测评实验室、北京市网络安全防护技术重点实验室资助。
收稿日期: 2024-01-16; 修改日期: 2024-04-08; 定稿日期: 2024-07-24

application of large language models to threat intelligence and cyber security and gives future research directions to further promote the development of cyber threat intelligence.

**Key words** cyber threat intelligence; intelligence aggregation; intelligence analysis; intelligence application

## 1 引言

在当今这个信息技术迅速发展的时代, 我们正面临着日益严峻的全球网络安全挑战。近年来全球网络安全攻击态势严峻, 攻击数量不断攀升, 攻击手段也日趋复杂和先进, 传统的网络安全防御手段已无法应对这些网络威胁。据 Check Point 统计, 2022 年全球网络安全与 2021 年相比, 网络攻击数量增加了 38%[1]。此外, 地缘政治冲突引发的网络战也逐渐从幕后走向前台, 形势越发紧张。在这样的背景下, 网络安全已成为事关各国国家安全的重要问题, 为了有效应对和减少这些网络安全威胁, 具有“一点发现, 全局共享, 协同联动”特点的威胁情报机制已成为一个不可或缺的关键工具。

网络威胁情报是关于现有或潜在的威胁信息, 经过采集、处理和分析, 包含了攻击场景、机制、技术指标和可采取行动的建议等, 能够辅助组织对网络威胁进行分析决策和响应。威胁情报是对各种网络安全数据采用多技术手段进行深入挖掘, 关联分析后产生的分析结果[2]。威胁情报能有效缩短攻击响应时间、减小攻防不对称性和发现预测新型威胁, 是政府企业一直竭力获取的网络安全战略资源。

随着不同国家、组织之间的网络博弈深入, 网络攻击正朝着形式复杂、针对性强、危害性高、隐蔽性强, 持续时间久等方向发展[3]。而威胁情报不仅包含了基础信息和信誉(如恶意软件签名、IP 地址等), 还能作为高级知识的表达(如攻击者的攻击手法、动机等), 这种多层次的情报极大地助力政府与企业快速识别和应对持续演变的网络威胁活动。云计算、大数据、人工智能等技术的快速发展和应用, 正不断地打破传统网络边界的定义[4], 在这种背景下, 威胁情报能对政府和企业数字化转型期间遗留的安全问题提供有效的解决方案。《“十四五”国家信息化规划》中详细阐述了国家网络安全面临的新形势和发展新需求[5], 提高企业网络威胁应对能力和安全管理水平迫在眉睫, 而网络安全产业智能化升级的过程必然离不开威胁情报这类宝贵的数据资源。

尽管威胁情报在政府和企业网络安全体系的建设管理上存在巨大的应用价值, 但在其技术实现方面仍然存在一些关键性问题, 包括情报要素提取准确性不足、情报的可解释性差、情报时效性低、共享难等。虽然不少研究者将深度学习、自然语言处理、机器学习等技术应用于威胁情报分析处理及应用[6-10], 缓解了部分原先存在的问题, 但日益严峻的安全形势使得安全人员仍然无法从海量的安全数据中摆脱出来。

因此为了进一步降低威胁情报生产的成本, 包括时间和人力资源的投入, 同时提高情报质量和可解释性, 并实现从被动防御转向主动防御, 该领域应聚焦于:

(1)提高情报质量: 提升情报的准确性和针对性, 确保提供的信息对防御措施具有实际价值。

(2)降低生产成本: 探索自动化和半自动化的工具和方法, 以减少人力资源的投入和加快情报分析处理的速度, 生成情报从而缩短攻击响应时间。

(3)增强情报的可解释性: 提供更清晰的情报上下文, 帮助分析人员更容易理解情报内容。

(4)实现主动防御: 识别潜在的威胁, 从传统的反应式防御转向更主动和预先防范的安全策略。

随着 ChatGPT 的火爆, 大语言模型应用受到了安全界的广泛关注[11], 也为解决上述问题提供了可能。相较于传统机器学习模型, 大模型具备生成式人工智能、类人推理、智能交互等涌现能力[12], 适合应用在威胁信息提取、知识图谱生成、情报结构化等威胁情报处理过程, 能够极大节省这部分的资源开销。因此, 本文使用大语言模型、威胁情报关联的检索语句, 对 Web of Science、ACM、IEEE 等数据库在该交叉方向近几年内的重要会议或期刊论文进行了搜集。同时, 由于研究主题属于近两年内新兴的交叉领域, 按照表 1 的标准对参考文献进行了筛选和梳理, 以帮助研究人员更好地了解该领域的研究现状、把握研究方向、开展后续研究工作。

本文围绕大语言模型在网络威胁情报领域中的技术成果与潜能进行梳理总结, 重点在于情报分析和应用, 并给出在该领域的未来研究方向, 便于网络安全学者的更深入研究。

文章结构如下, 第 1 节引言介绍了系统综述写作背景; 第 2 节介绍了威胁情报和大语言模型的关键背景知识主要包括概念、发展历史以及技术研究现状; 第 3 节至第 5 节分别梳理总结了大语言模型在增强威胁情报聚合、驱动威胁情报分析和赋能威胁情报应用三个方面的技术研究; 第 6 节对该领域进行了未来展望和更多的问题讨论; 最后是结束语。

表 1 参考文献筛选标准表

Table 1 Reference selection criteria table

| 选用文献的特征 | 剔除文献的特征 |
|---|---|
| 1. 选取材料来源于官方数据库的正式期刊、会议论文或预印版论文等。 | 1. 文献来源不明, 或所属期刊、会议可信度不够, 已被移出数据库的。 |
| 2. 内容完整, 与研究主题直接关联。 | 2. 仅与单个主题关联, 或研究范围不完整。 |
| 3. 文献来源于重要会议、影响因子较高的期刊, 或在某研究方向的工作具有代表性。 | 3. 文献所属会议或期刊影响力过低, 或属于某项后续研究的前置工作的。 |
| 4. 文献配套开源代码, 或实验描述详实可靠、便于复现的。 | 4. 文献不具备开源解决方案, 同时实验过程较模糊、不利于复现和继续研究的。 |
| 5. 优先考虑近三年以内的研究成果。 | 5. 尽量不选用近五年以外的研究成果。 |
| 6. 使用英语写作。 | 6. 使用其他语言写作。 |

## 2 背景介绍

深度学习和威胁情报的最早关联可以追溯到 20 世纪 90 年代, 当时已有安全人员尝试使用神经网络去进行威胁识别、入侵检测等工作[13]。而大语言模型作为深度学习技术产物, 最早诞生于计算机视觉领域, 近几年在网络威胁等安全领域出现了相关应用。本节将简要介绍威胁情报和大语言模型的知识背景, 便于后续更好地探索二者交叉的研究工作。

### 2.1 威胁情报

威胁情报(Threat Intelligence, TI)的早期实践起源于军事领域, 在这一阶段, 人类决策者和专家指导、收集、处理情报并将其传播给其他相关利益方[14]。2014 年, Gartner[15]对威胁情报进行了正式定义: 威胁情报是一种基于证据的知识, 包括上下文、机制、指标、影响、含义和可执行的建议等, 这些知识与资产与所面临已有的或潜在的威胁或危害有关, 可为实际的决策提供信息支持。随着计算机网络技术的发展, 威胁情报的应用范围扩展至网络空间, 演变成网络威胁情报(Cyber Threat Intelligence, CTI)。CTI 是从安全数据中提炼的与网络空间威胁相关的信息, 包括威胁来源、攻击意图、攻击手法、攻击目标信息, 以及可用于解决威胁或应对危害的知识。CTI 以空间换时间, 知己知彼, 协同联动, 可用于入侵防御、威胁发现、攻击溯源、态势感知及预警、主动防御等业务场景, 显著提升网络空间安全防御能力。以下所提威胁情报均指网络威胁情报。

针对威胁情报的分类, Bianco[16]根据情报价值及其获取的难易程度, 制作了一个名为“痛苦金字塔”的层次模型以描述威胁情报体系, 如图 1 所示。该模型从下到上依次为: 文件哈希、IP 地址、域名信息、网络或主机特征、攻击工具和 TTPs(Tactics, Techniques & Procedures, 战术、技术和过程)。其中哈希值、IP 地址、域名这种低层级威胁指标较易获取, 常被加入特征库用于对比检测恶意样本, 属于传统失陷指标(Indicator of Compromise, IOCs)类型。这类指标攻击者可通过沙箱和数据分析等自动化手段生成, 轻易改变网络特征就能逃避检测, 且失效快, 关联性差, 价值也较低。相比之下, 金字塔上层的网络或主机特征、攻击工具和 TTPs 等高级威胁信息通常需要人工分析研判才能得出, 且攻击者不易改变攻击手法和攻击工具的特征, 具有较强的关联关系和更高的价值。当防御者掌握这些指标, 就会给攻击者带来一定程度的攻击代价或痛苦。MITRE 曾在 ATT&CK 框架中特别指出, 关注战术和技术对于超越“传统的 IOCs”非常重要[17]。

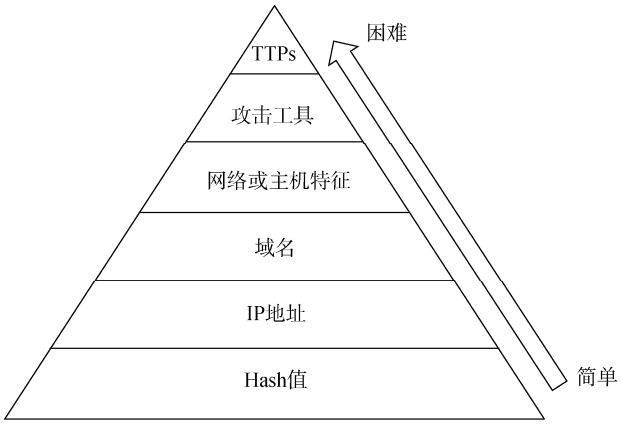


图 1 “痛苦金字塔”层次模型[16]

Figure 1 “The Pyramid of Pain” hierarchical model[16]

除此之外, 可以按照情报服务对象将情报分为战略威胁情报、运营威胁情报和战术威胁情报[18]。战略威胁情报主要帮助组织的管理者了解当前安全态势并做出安全决策, 涵盖了网络活动或攻击趋势等带来的影响以及威胁活动的历史数据或预测。运营威胁情报主要帮助安全分析师或者安全事件响应人员分析及响应, 包括已知或即将发生的攻击信息如利用已知的攻击者技战术手法, 主动的查找攻击相关线索。战术威胁情报主要用于发现威胁事件以及对报警确认或优先级排序, 一般应用于防护系统

或设备。代表性的是失陷检测指标、攻击指征, 如C&C(Command & Control Server, C2)地址、IP 黑名单, 都是可机读的情报, 可自动完成威胁发现, 甚至通过联动实现威胁阻断。

威胁情报生命周期可以大致划分为威胁情报计划定向、威胁数据收集、威胁数据处理、威胁情报分析、威胁情报传播、威胁情报反馈六个阶段[19]。从各类威胁数据到完整的威胁情报通常需要经过复杂的分析生产过程, 安全人员首先会根据需求, 采集特定的原始安全数据, 包括流量、日志、恶意代码、开源信息等。然后对原始数据进行清洗、标准化、结构化, 挖掘分析其中存在的威胁信息, 并进行推理、研判、高级知识归纳等生成完整威胁情报, 发挥其更高的价值。同时, 安全人员将生成的威胁情报进行共享和应用, 以融合丰富情报内容, 并不断改进情报收集、分析和应用的方法技术。

在此过程中, 虽然已有的技术研究很大程度上提高了威胁情报生产、管理和应用的自动化、智能化水平, 但威胁情报技术目前仍面临以下困境:

(1)多源异构情报采集汇聚困难。威胁情报的多源异构特性导致情报质量参差不齐, 特别是源于缺乏标注的数据集或单条的可操作情报, 缺乏上下文信息。此外, 针对多个数据源的情报难以有效融合, 面临信息缺失或断裂、信息矛盾以及由此导致的情报混淆等问题。

(2)情报的分析处理技术不足。情报信息提取过度依赖专家先验知识, 并且机器学习等方法通常针对特定任务训练分类器, 泛化能力差。此外网络安全领域的专有术语和表达方式, 在结合自然语言技术时会带来语义歧义, 难以理解关联上下文内容从而准确识别威胁要素。

(3)威胁情报的评估缺乏统一公认的标准。细粒度的情报质量评估通常难以开展, 主要因为当前对于情报的可用性和可信性的评估, 缺乏公认的指标, 也缺乏支撑评估过程的相关信息, 现有的做法往往只能在数据源层级对批量情报进行统一评估。

(4)威胁情报的时效性较低, 情报共享成本较高。威胁信息更新快, 而威胁情报的生产、整理、传播和应用过程又需要一定时间, 易出现当情报如C2 地址, 被安全人员接收时已经失效的情况。同时, 情报中的保密和隐私信息也会对情报的共享造成阻碍。

(5)威胁情报的可解释性差, 应用具有局限性。情报难以在缺乏相关上下文的情况下单独发挥应用价值。同时, 目前直接应用威胁情报进行 APT 攻击组织溯源归因的工作仍然比较困难。

这些问题让威胁情报难以发挥其真正的价值, 还可能会影响威胁情报技术的发展和相关安全产品商业化的进一步扩大。因此, 采纳新技术成为解决这些难题的关键。在这方面, 大语言模型的独特优势, 为解决部分问题提供了可能。它们的高级自然语言处理能力, 能有效解析多源异构数据, 从而提高情报的集成和上下文分析。此外, 大语言模型的上下文学习和强大的泛化能力, 能够从大量数据中学习到特征, 准确识别关键威胁要素, 减少对专家知识的依赖, 提高分析的自动化程度。最后, 大语言模型在提升情报的可解释性和拓展其应用方面具有重要作用, 不仅提供深度的数据分析, 还具备与安全专家有效的交互, 使得威胁情报的应用更为直观和容易理解, 这种辅助和交互能力显著提升了安全人员的执行效率和准确性, 从而增强网络安全防护的整体性能。

## 2.2 大语言模型

伴随目标计算需求和深度模型规模的扩大, 研究者们惊喜地发现神经网络开始表现出思维链推理、情感分析、信息摘要等一系列传统模型无法表现的高级能力。这种能力在文献[20]中被定义为“智力涌现”(Intellectual Emergent), 即小模型中不存在但大模型中存在的能力。它在模型规模达到一定水平时, 性能显著提升。而这种涌现能力可以作为大模型和传统模型区分的重要标志之一。

由于大模型的“智力涌现”能力, 大语言模型开始在社会和学术界引起广泛的关注。这一趋势在下图中得到了清晰的体现。图 2 展示了自 2018 年以来, 关于大语言模型话题在谷歌趋势上的热度变化[21]。从图中可以明显看出, 随着 2022 年 11 月 ChatGPT 的发布, 该话题的整体热度有了显著增长。而在 2023 年 1 月至 5 月期间, 随着 GPT-4 的推出, 这一热度更是出现了剧烈增长。图 3 则反映了大语言模型在学术界的影响力。与 2021 年相比, 2022 年和 2023 年在 Web of Science 数据库中与 LLM 相关的论文数量急剧增加[22]。这一趋势表明, LLM 已经吸引了众多研究者的关注, 并在学术领域引发了广泛的探索和研究。

大语言模型主要应用于自然语言处理领域, 能够很好地完成文本生成、对话问答、智能检索、机器翻译等任务[23]。大语言模型通常具备数十亿到数百亿个参数, 拥有极其复杂的网络结构和极高的表示能力。目前流行的大语言模型包括各类 GPT 系列模型及开源大模型, 如 OpenAI 的 GPT-4-turbo 模型、Google 团队的 PaLM2 模型和 Meta 的 LLaMA 模型等[24], 它们在金融、医疗、教育等行业均取得了一些应用成果。

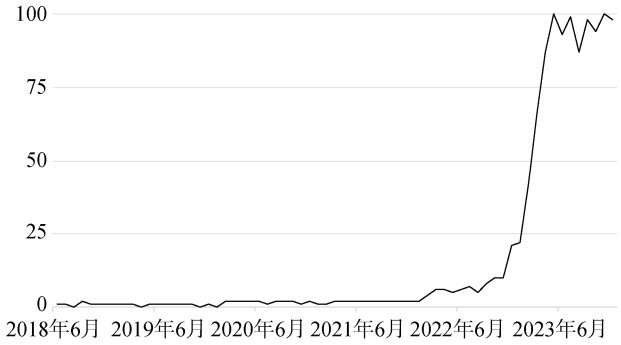


**图 2 “Large Language Model”话题的谷歌趋势[21]**
**Figure 2 Google Trends on the topic of “Large Language Model” [21]**

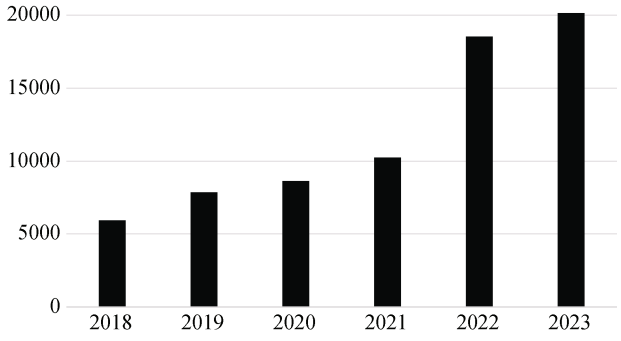


**图 3 “Large Language Model”主题的论文累计数量[22]**
**Figure 3 The cumulative number of papers on the topic of “Large Language Model”[22]**

大语言模型的涌现能力可以体现在以下几个方面[25]:

(1)上下文学习: 175B 参数的 GPT3 在许多复杂的 NLP 任务上如问答、机器翻译、文章生成等表现较好, 而 GPT1 和 GPT2 则表现不佳, 这展现出大语言模型在理解和处理上下文方面的优势。

(2)指令遵循: 通过对自然语言描述格式的多任务数据集进行微调(称为指令微调), LLMs 能够在同样以指令形式描述的新任务上表现出色[26-28]。通过指令微调, LLMs 可以在不使用具体示例的情况下, 根据任务指令完成新任务, 从而提高泛化能力。

(3)逐步推理: 小模型很难解决涉及多个推理步骤的复杂任务。相比之下, 大模型可以通过思维链(Chain-of-Thought, CoT)提示策略[29], 将多步骤的复杂任务分解成一系列中间推理步骤, 为模型行为提供可解释的窗口, 从而显著地提高大语言模型的推理能力, 有效地解决复杂任务。

为此, 本小节将简要介绍大语言模型的发展历史和技术研究现状, 以便更好地发掘大语言模型和威胁情报工作交叉的可能。

### 2.2.1 大语言模型的发展历史

大语言模型的发展可以溯源至早期的自然语言处理技术, 其代表工作是 2013 年 Google 的 Mikolov 等人[30]提出的词向量(Word2Vec)技术。Word2Vec 主要包括两种方法, CBOW 和 Skip-gram, 二者均通过浅层神经网络将单词嵌入到连续向量空间中。该过程将语义相近的单词映射到相邻位置, 使模型能够根据上下文或目标词语预测其他词语, 进而捕捉更丰富的语义信息。Word2Vec 首次提出了通过无监督学习从大规模文本数据中学习语言表示的思想, 弥补了传统词袋模型在语义表示上的不足, 为后续BERT 和 GPT 等大语言模型的出现奠定了基础。

与此同时, 在深度学习领域, Cho 等人[31]对传统的长短时记忆循环网络(Long Short-Term Memory, LSTM)做出了优化, 提出门控循环单元(Gate Recurrent Unit, GRU)。该技术在维持网络原有性能的同时, 显著减少了参数数量, 进而降低计算复杂度。GRU 和 LSTM 的引入有效增强了语言模型对长距离依赖关系的学习能力, 为大语言模型提供了高效的序列建模工具和关键技术支持。

随着可用数据增加、计算资源增强和预训练语言模型兴起, 2018 年诞生了两个具有划时代意义的语言模型, 分别是 Google 团队的 BERT 和 OpenAI 团队的 GPT。由 Devlin 等人[32]提出的 BERT 模型采用了双向 Transformer 架构, 能够同时考虑一个词向量两侧的上下文, 有效提升了语言模型的上下文相关性分析和语境歧义处理的能力。BERT 通过在大规模文本数据上预训练来学习通用语言表述, 然后在特定任务上进行微调, 在专家问答、文本分类等任务中表现出色。同时, Radford 等人[33]发布的 GPT-1 模型则选择了自回归 Transformer 架构, 参数规模达到 1.17 亿左右。GPT-1 采用无监督预训练和监督微调的混合方法, 较好地完成了各类复杂的自然语言处理任务。在大语言模型的发展历史中, BERT 和 GPT 共同扮演着里程碑式的角色, 也代表了大语言模型今后发展的两条主流道路, 在这里主要讨论这两种, 如表 2 所示[34]。

BERT 和 GPT 为各类语言模型树立了新的性能标杆, 也引领了后续更大规模、更复杂的大语言模型的研发和应用。仅隔两年, AlecRadford 等人陆续发布了 GPT-2[35]和 GPT-3[36]。GPT-2 的参数规模最多达到 15 亿左右, 使用大型网页数据集 WebText 进行预训练, 并尝试了无监督语言建模来执行任务。GPT-2 具备了比较成熟的内容生成、零样本学习和模型规模自适应等能力, 支持文章写作、智能翻译、专家问答等新服务。GPT-3 则在 GPT-2 的基础上将模型参数激增至 1750 亿左右, 采用更复杂的网络结构和模型

表 2 大语言模型总结对比[34]

Table 2 Summary of Large Language Models[34]

| 模型架构 | 特征 | 代表模型 |
|---|---|---|
| Encoder-Decoder 或 Encoder-Only (BERT 模式) | 掩码语言模型; 单词生成依赖其两侧上下文; 判别式模型 | ELMo, BERT, RoBERTa, DistilBERT, BioBERT, XLM, Xlnet, ALBERT, ELECTRA, T5, XLM-E, ST-MoE, AlexaTM |
| Decoder-Only (GPT 模式) | 自回归语言模型; 单词生成依赖其左侧上下文; 生成式模型 | GPT-3, OPT, PaLM, BLOOM, GLM, MT-NLG, GLaM, Gopher, chinchilla, LaMDA, GPT-J, LLaMA, GPT-4, BloombergGPT |

架构, 并使用代码数据训练和基于人类反馈的强化学习(Reinforcement Learning from Human Feedback, RLHF)等方法做进一步优化。这使得 GPT 在迁移泛化能力、内容生成质量、多任务学习等性能上获得极大提升, GPT-3 能够支持代码生成、创意写作、深度推理等更复杂的服务。与此同时, BERT 类模型也发展出了 RoBERTa[37]、ALBERT[38]和 DeBERTa[39]等优化模型, 被广泛应用于文本分类、实体识别、智能翻译等任务中。

2022 年底, 基于 GPT-3.5 开发的智能对话机器人一夜走红, 成功的技术实践和商业化结果将大语言模型直接推向了人工智能技术和资本的风口浪尖。国内外以 Open AI、微软、Meta、百度等为代表的企业组织纷纷加大了对大语言模型技术和应用的投资力度, 催生出 GPT-4[40]、Bard[41]、LLaMA[24]、文心一言等新一批大语言模型, 能够为用户提供更高质量的内容生成、多模态信息支持、深度语义分析等服务。目前, 国内外大语言模型市场已呈现出“百模大战”的局面, 不同类型的大语言模型在金融管理、医疗问诊、智能客服、网络安全等领域正发生日益深刻的产业渗透。而在 2023 年 11 月 7 日 GPT-4-turbo 的发布会上, Open AI 指出用户可以通过直接提交数据给 GPTs, 定制出不同任务和应用场景的 GPT, 并能在未来的 GPT Store 中进行商业用途使用。该过程几乎不需要任何编程操作, 极大降低了大语言模型的使用和构建门槛。

### 2.2.2 大语言模型的技术研究现状

从最早 word2Vec 的提出, 到 BERT、GPT 的诞生, 再到 GPT-4、Bard 等大语言模型的流行, 大语言模型技术无不遵循着“量变到质变, 暴力出奇迹”的思想, 并表现出了以下技术特征和发展规律:

1)虽然底层架构和感知能力没有明显优化, 但整个模型的复杂性和输出性能逐年提高, 模型的规模和效能在不断均衡优化。具体而言, 大语言模型可以视为深度学习模型朝向“更宽、更深、更多”方向发展的产物。更宽主要指代神经网络每一层拥有更多的参数; 更深主要指代神经网络拥有更多、更复杂的层次结构; 更多主要指代神经网络预训练数据的体量和质量不断提升。以上要素共同推动神经网络具备越来越强的信息表示能力和智力涌现能力, 完成了深度学习语言模型向大语言模型的蜕变。

2)开源大语言模型和闭源大语言模型协同发展。目前市面知名的开源大语言模型包括 GPT-NeoX-20B[42]、BERT、Falcon[43]等, 它们支持各类科研及商业使用; 闭源大语言模型包括 Claude、GPT4 等, 它们通常以具体产品的形式分布在商业市场。不少闭源大语言模型是开源大语言模型优化后的产物, 但也为开源大语言模型提供了新的优化思路和数据资源, 二者的协同发展受企业和研究者之间具体的合作或竞争关系的影响。

3)编码器和解码器的优化是大语言模型发展的重要环节。可以通过 2018 年的 ELMo[44]、GPT-1 等, 2019 年的 RoBERTa 模型, 2020 年的 DeBERTa、GPT-3 等模型的相关工作中明显发现, 编码器和解码器的组件架构优化极大推动了对应模型性能的提升。

4)大语言模型方法多样化, 产品更侧重应用层面创新。例如 LLaMA、Bard 等模型在自然语言处理领域的多任务集成应用场景中发挥了不错的表现, 而通过“预训练-微调”范式定制的各个垂直领域的大语言模型也在各行各业中发挥出越来越明显的优势。

随着技术进步和应用扩展, 对大语言模型的信息内容, 道德伦理、个人信息保护等方面的风险管理也日益受到重视。尽管当前尚未形成全球统一的系统的法规政策与监管体制, 但已有一些重要政策和法规出台[45]。下面从国外和国内介绍大语言模型治理现状。

**国外** 欧盟和美国等地区在大语言模型的研发和应用上采取了一系列监管措施。2023 年 3 月意大利数据保护机构一度对 ChatGPT 发布禁令, 并调查其涉嫌违反欧洲隐私法规的行为。部分欧盟国家表示将遵循欧盟通用数据保护条例(General Data Protection Regulation, GDPR)监控 ChatGPT 个人数据泄露的风险。欧洲数据保护委员会近期宣布成立专门工作组, 旨在促进各国协同调查, 并就可能采取的执法措施进行交流。2023 年 12 月 8 日, 欧盟就《人工智能法案》达成协议, 这将成为全球首部人工智能领域的全面监管法规。此外, 在同年 4 月, 美国国家电信和信息管理局发布《人工智能问责政策征求意见》, 就是否以及如何对生成式人工智能等工具进行监管和问责征求相关利益主体的意见和建议[46]。2023 年 5 月,

拜登政府宣布将对现有生成式人工智能系统进行公开评估, 确保其符合人工智能负责任使用的原则。

**国内** 中国已构成包括《刑法》、《民法典》、《国家安全法》、《数据安全法》、《治安管理处罚法》以及《网络安全法》、《个人信息保护法》、《互联网信息服务管理办法》等法律法规的信息内容安全监管框架, 明确禁止危害国家安全、社会稳定和虚假信息等有害信息, 为生成式大语言模型的内容规制提供了基础。此外, 今年年初推出的《互联网信息服务算法推荐管理规定》和《互联网信息服务深度合成管理规定》进一步为生成式大语言模型的应用提供了基础性规则。2023 年 4 月, 中华人民共和国国家互联网信息办公室发布《生成式人工智能服务管理办法(征求意见稿)》(以下简称《办法》), 这是我国首个针对生成式人工智能产业的规范性政策。《办法》从数据使用、个人信息收集、内容生成、内容提示标注等全流程对生成式人工智能服务提出了一系列监管设想, 并于 8 月 15 日正式施行。目前已超 20 个大模型通过《办法》获得备案。

大语言模型的发展不仅是技术上的进步, 也是伴随着工业实践、学术研究和政策规制相互作用的结果。随着大语言模型在各领域应用的深入, 全球范围内对其进行有效管理和规制的需求日益增强。学术界、工业界和政府机构正共同努力, 以确保大语言模型的发展既能推动技术创新, 又能保障社会和个人的安全与权益, 致力于为用户构建更加安全高效的大语言模型。

### 2.3 小结

通过以上梳理, 可以发现网络威胁情报目前面临着人力和时间成本较高的问题, 而此时通过传统模型(如隐马尔可夫(Hidden Markov Model, HMM)、条件随机场(Conditional random field, CRF)、早期的卷积神经网络(Convolutional Neural Network, CNN))已很难做出大的优化。大语言模型具备类人推理、深层语义理解、语境分析及多模态任务支持等能力, 能够与威胁情报产业, 尤其是威胁情报生命周期后阶段任务发挥更好的交融。同时, 目前既有的开源威胁情报及其他威胁情报获取方式, 能够为大语言模型微调提供较为充足的数据资源。这为通过迁移学习、领域自适应、知识蒸馏等方法定制威胁情报领域或者整个安全垂直领域的大语言模型提供了极大支持。

为了更好地应用大语言模型于威胁情报领域, 学术界和工业界均展开了一系列的研究和实践活动。在这一过程中, 大语言模型在威胁情报的聚合、分析和应用等诸多过程中展现出了巨大的潜力。研究者们通常会按照以下研究线索探究大语言模型的应用方案, 展开技术问题的具体讨论:

(1)提高威胁情报采集的效率与质量:

处理多源异构情报和情报的采集汇聚是一个挑战。传统方法如定制爬虫、数据清洗等一系列流程, 面临着效率低和无法充分处理复杂数据的问题。大语言模型能否更有效地收集、筛选相关数据并减少噪音信息, 提高数据的质量和可用性。这需要将工具与自动化定制的采集策略相结合, 实现从数据源到分析准备的自动化流程, 为后续的分析处理提供更准确的基础数据。

(2)深化威胁情报分析处理的能力:

传统模型在实体识别、检测分类等任务中存在泛化能力较差, 误报率高等问题。大语言模型能否更充分学习上下文, 提供更准确的识别和分类。此外, 探索大语言模型的深度分析, 对于构建和完善威胁情报知识图谱至关重要, 这有望为威胁情报提供更全面和更深入的背景信息。

(3)拓展威胁情报应用的可能性:

传统方法在对新型威胁的预测、检测和溯源方面, 存在诸多挑战。大语言模型能否利用攻击行为分析等技术, 提供新的视角以准确识别潜在威胁。同时, 在威胁溯源阶段, 其推理能力和关联分析能力可以用于攻击者的溯源。将大语言模型集成到现有的威胁检测系统, 有望提高整体系统的自动化和智能化。

结合目前的三条研究线索和威胁情报生命周期的六个阶段, 将面向威胁情报的大语言模型技术应用, 按照威胁情报处理的过程自下向上归纳为 LLM 增强威胁情报聚合、LLM 驱动威胁情报分析和 LLM 赋能威胁情报应用三大阶段, 具体如图 4 所示。

其中, 威胁情报聚合属于威胁情报产出的前置阶段, 通过收集各类情报数据, 并对数据做预处理和评估融合, 为威胁情报内容分析提供基础。威胁情报分析主要包括威胁情报的信息提取、情报生成、知识图谱构建等流程。威胁情报的应用则属于威胁情报产出的后续工作, 也是威胁情报价值实现的重要途径, 这些阶段都在不同程度上存在着大语言模型技术应用的相关工作或研究空间, 具体内容将在后续三个章节中详细探讨。

## 3 LLM 增强威胁情报聚合

威胁情报聚合是指将各类网络安全数据作为输入, 输出可以进行下一步分析处理的威胁情报基础信息。威胁情报的聚合包括情报数据的自动采集、情报预处理、情报评估融合等步骤, 大语言模型的引入对这一流程产生了一定影响, 能够有效增强情报的收集和整合。威胁情报根据来源不同, 一般分为内部情报和外部情报[47], 如表 3 所示。其中内部情报主

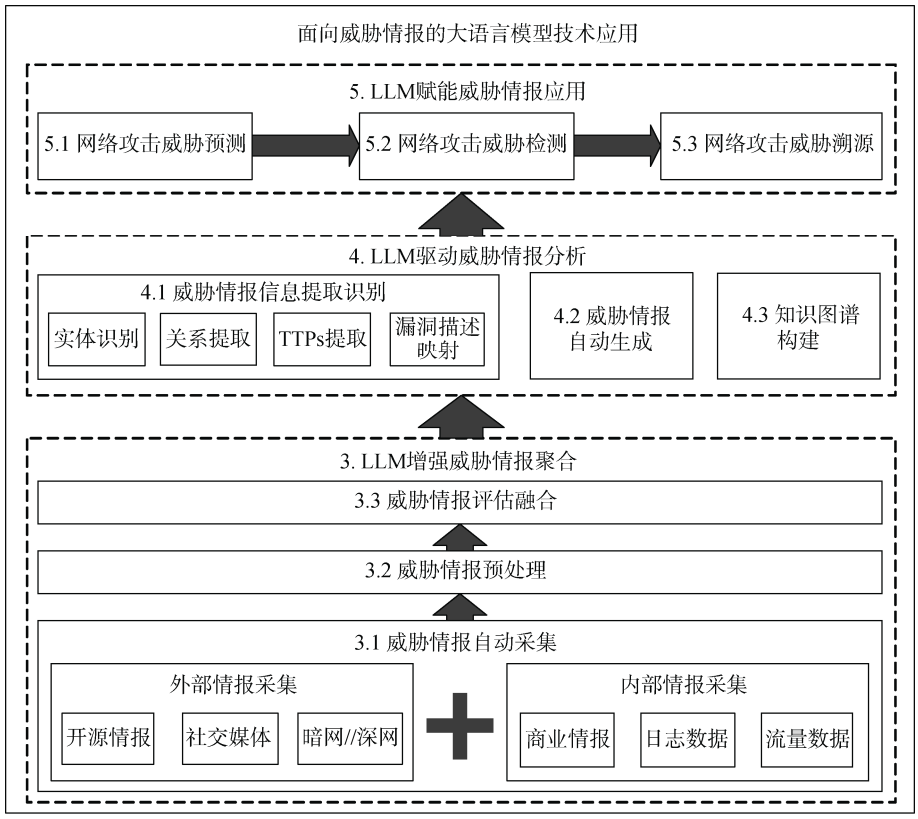


**图 4　面向威胁情报的大语言模型技术应用框架图**

**Figure 4　Application framework of large language model technology for threat intelligence**

**表 3　威胁情报来源分类**

**Table 3　Classification of threat intelligence sources**

| 类型 | 威胁数据实例 | 来源 |
|---|---|---|
| 内部情报 | 安全事件 | 安全专家、SIEM、SOC |
| | 安全日志数据 | 防火墙、IDS、IPS、沙箱、高交互蜜罐 |
| | 流量数据 | 交换机、路由器 |
| 外部情报 | 开源威胁情报 | 安全博客、社交媒体 Twitter、暗网等 |
| | 商业威胁情报 | 安全厂商 |

要源于内部组织掌握的数据, 包括安全设备防火墙、入侵检测系统(Intrusion-detection system, IDS)、入侵防护系统(Intrusion-prevention system, IPS)、高交互蜜罐、沙箱等产生的安全日志数据, 安全专家、SIEM (Security information and event management, 安全信息和事件管理)等分析得出的安全事件信息和交换机路由器等流量监测数据。外部情报主要是源于安全博客、CVE(Common Vulnerabilities & Exposures, 通用漏洞和风险)漏洞库、社交媒体 Twitter、暗网等开源威胁情报, 和各大安全厂商付费的商业情报数据。整合关联内部情报与网站、推特、暗网、论坛等外部情报, 不仅为情报提供了丰富的上下文, 还为后续威胁情报的分析处理奠定了坚实的基础。

## 3.1　威胁情报自动采集

威胁情报自动采集可以从外部情报和内部情报进行探讨。

(1)外部情报

开源威胁情报(Open Source Intelligence, OSINT)是外部情报的主要获取方式, 也是研究者主要的研究对象, 它能够帮助分析野外发现的恶意软件以及情报界获得的恶意软件, 补充 IBM、VirtusTotal 和 Mandiant 等公司收集的资源。下面主要介绍长文本、短文本和暗网等开源威胁情报的自动采集技术。

长文本主要是篇章级的非结构化文本如安全新闻、博客、APT 攻击报告等, 包含了丰富的上下文信息。传统采集方法主要是研究自动化爬虫及解析技术[48], 并利用机器学习[49]、深度学习 CNN[50]等技术收集并筛选出所需要内容的数据。但因不同厂商网页结构不一致, 需要定制爬虫来获取特定厂商页面的报告文本, 而大语言模型可以让采集过程更加自动化。2023 年 11 月 OpenAI 发布了支持 AI 智能助理定制及链接分享的 GPTs 工具。不到一周, 社区公开发布的 GPTs 数量已过万。其中有些 GPTs 的应用场

景是威胁情报聚合检索, 在集成 GPT-4 的语言理解能力上, 还扩展了WEB检索能力。如TheDFIRReport Assistant[51]支持用户从 TheDFIRReport 网站获取并讨论最新报告, 支持对关键词进行搜索, 并抽取搜索结果中的关键词, 将不同攻击路径中的关键词聚合, 还可以定期并自动化追踪相关恶意活动的趋势。Threat Intel Brief[52]可以提供每日特定部门的网络安全威胁情报简报以及来源引文。CyberGPT[53]可以生成针对特定网站(比如安全厂商卡巴斯基、赛门铁克和火眼等网站)的检索爬虫, 并使用 BeautifulSoup 开源库对这些网站进行爬取和内容解析, 从而获取情报。这三个 GPTs 都可以自动化采集公开数据, 帮助研究人员节省时间和资源。文献[54]提出一个基于 GPT3.5 的自动知识情境化系统 LOCALINTE, 该系统根据提示从全球威胁存储库中自动检索威胁报告, 并根据本地知识数据库生成为组织量身定制的威胁情报。对系统的定量评估采用 RAGAS 指标, 它是专为评估检索增强生成(Retrieval-Augmented Generation, RAG)管道设计的综合指标, 结果发现平均 RAGAS 得分为 95%, 证明了该方法能够提供准确且相关的威胁情报, 帮助安全分析师减少手动工作量。

短文本形式的社交媒体内容如实时发布的推文, 常常包含最新且丰富的威胁信息。社交媒体平台特别是推特也成为传播前沿和关键网络威胁情报信息的重要渠道。Bose 等人[55]提出一种使用图网络结构跟踪 Twitter 数据流中与网络威胁相关的用户账户的方法。利用推文上下文内容以及从 Twitter 社区中提取的信息对用户进行评分和排名。这种排名机制是为了确定哪些用户提供的信息与网络威胁最相关, 从而能够按照优先级来获取最相关的威胁信息。Wang 等人[56]提出了一个机器学习框架, 利用主题建模来发现和分析 Twitter 上的网络威胁。暗网的匿名技术实现了互联网的匿名交流和沟通, 也成为了恶意活动的首选平台。很多网络攻击工具和各类情报数据会在此非法售卖。暗网数据情报多为实时信息, 内容多样, 有助于发现新兴威胁。通常暗网情报会使用特殊手段进行一定遮蔽或者加密, 难以识别。为了对抗暗网的极端词汇和结构多样性, Jin 等人[57]提出 DarkBERT, 一个在暗网数据上预训练的语言模型, 以适应暗网领域特定的语言分析任务。作者收集公共存储库的种子地址, 再根据种子地址在暗网中抓取页面, 并以此扩展域列表, 进而爬取更多的数据, 构建了由暗网页面组成的海量文本语料库。

(2)内部情报

内部网络安全情报的监测采集主要源于入侵检测系统、网络流量检测、蜜罐等技术手段。入侵检测被动防御采集的异常数据和蜜罐主动式防御获取的数据, 虽然本身并不直接解释攻击者的完整路径, 但通过分析挖掘, 可以获取攻击者的行为模式、攻击技巧和使用的恶意软件特征等有价值的情报信息。下面从流量和日志两个数据对象介绍内部情报的采集监测技术。

提高入侵检测系统性能有助于从流量中获取高质量情报信息。文献[58]提出 HuntGPT, 基于 GPT3.5 的对话代理与可解释人工智能集成到入侵检测系统。其核心是将检测到的威胁以易于解释的格式提供给用户, 便于人工对异常数据包的检测, 提高了可解释性和可操作性。在评估部分, 本文使用网络安全技术知识来检查生成答案的准确度和精确度, 并评估 GPT3.5 提供响应的质量和适当性。为了减少真实数据获取的成本, 文献[59]提出一种端到端的 PCA-GPT 工具, 利用 GPT-3 生成网络流量, 使用损失率、准确率和成功率等指标来评估性能。这些数据可以替代真实数据, 用于各种网络安全任务如威胁建模等。

安全日志的采集是了解攻击行为和诊断漏洞的关键, 为了获取细粒度情报信息, 需要更好的分析攻击日志。都灵理工学院团队[60]利用语言模型自动分析和总结文本式 Unix shell 攻击日志。性能指标采用二分类准确度和ROUGE评分, 二者分数越高说明性能越好。在识别和理解攻击者的策略上, GPT-3 Davinci 模型以花费更高的推理时间为代价, 获得比 CodeBERT 略差一些的性能。中科院信工所团队[61]提出了 LAAEB, 一种综合利用日志中丰富的文本内容信息的通用无监督内部威胁检测。使用基于 LLM 即 GPT3.5-turbo 的方法消除假警报, 只需给出每个用户的场景描述和告警描述, 以询问 LLM 的用户是否是良性的, 实验表明 AUC 值可以达到 94%。高交互蜜罐可以采集到更多情报数据, 布拉格捷克技术大学团队[62]提出一种基于 GPT3.5 的动态软件蜜罐生成方法。使用 LLM 来模拟 Linux 终端, 输出对应命令的结果, 迷惑攻击者, 实验表明, 命令输出的准确率高达 92%, 逼近真实环境。因此可以通过该方法生成动态蜜罐, 从而提高获取攻击者的情报信息效率。

近几年的研究通常使用爬虫、机器学习和自然语言处理相关技术来从各大平台上采集情报数据, 但是面对数据的复杂性、分散性和庞大规模, 往往需要人工为各种数据源定制不同的采集方案, 严重依赖手工任务, 属于劳动密集型。大语言模型的出现可以提高情报数据的采集聚合效率和准确性, 并且更好的分析, 从而提炼出较高质量的情报数据。

### 3.2 威胁情报预处理

威胁情报预处理通常包括数据清洗、数据标准化, 数据标注等操作。数据清洗通常包括正则表达式过滤、去除 HTML 标签、删除特殊符号, 然后使用 Spacy、NLTK 库对文本内容分段分句分词, 具体内容在这里不展开叙述。以下主要介绍 LLM 用于数据标准化和数据标注的技术。

数据标准化在预处理阶段非常重要, 它可以将数据统一格式, 便于管理、分析和存储。Siracusano 等人[63]提出了一个新的大型开放基准数据集和 aCTIon 一种结构化 CTI 信息提取工具。该数据集包括 204 个现实世界的公开报告及其相应的 STIX (Structured Threat Information eXpression, 结构化威胁信息表达)格式的结构化 CTI 信息。利用 GPT3.5 设计提示, 让LLM完成两个自定义的信息提取(实体关系提取和攻击模式提取), 并根据输出结果交互执行自检, CTI 分析师负责模型结果验证, 无误则输出 STIX 标准化格式报告。

目前开源的威胁情报数据集较少, 研究人员大多根据自身需求, 对数据打上特定的标签, 缺少泛化强的标签数据。而使用神经网络等模型训练, 需要大量的标签数据, 单靠人工去标注是一个大工程, 普遍采用专业人员或者众包方式完成标注工作, 耗时短则几周, 长则几月且需要一定财力的支出, 质量也无法保障。Markus Bayer 等人[64]结合了三种不同的低数据机制技术, 即迁移学习、数据增强和少样本学习, 从极少的标签实例中训练出高质量的分类器。在数据增强部分, 利用 GPT-3 模型根据现有的少数标记实例生成新实例, 大大减少数据标注工作量。分类效果方面, 标注少量数据实例(使用 32 个实例进行测试), 效果与使用 1800 个实例训练的分类器相当。

大语言模型的引入正在改变传统的情报预处理方式, 极大得提高了威胁情报预处理的质量和效率。

### 3.3 威胁情报评估融合

威胁情报评估融合包括基于情报源的评估、基于情报内容的评估和知识融合。

基于情报源的评估, 文献[65]定义了一组威胁情报数据源的度量标准, 即数量、差异贡献、独家贡献、延迟性、覆盖范围和准确性。并使用这些度量标准系统地评估了开源和商业的情报源, 其中重点分析提供 IP 地址和文件 hash 的情报源。结果表明, 使用现有威胁情报数据来实现其声称的目标存在重大限制和挑战。

基于情报内容的评估, Shin 等人[66]提出 twiti 系统, 它能够自动从 twitter 中筛选出 IOC 相关的推文, 提取 IOC 并广泛的评估了 IOC 的准确性、延迟性、排他性等性能。Miao 等人[67]针对网络领域, 推出了一个网络知识能力评测公开数据集 NetEval, 基于此数据集对当前各类大语言模型在网络领域的知识能力进行了评估。结果显示, 目前只有 GPT-4 能够在 NetEval 数据集上达到近似人类的知识水平。作者表示 NetEval 数据集在未来可以被广泛应用于分析基础模型在网络领域的优势和不足, 推动大语言模型在网络领域的发展。据此可以推断, 也可以用该数据集评估大语言模型在网络威胁情报内容的分析性能。麻省理工团队[68]评估了 GPT-3.5 在理解和生成与 MITRE ATT&CK 威胁行为相关内容的能力, 并对其生成内容的准确性和质量进行了评估。结果表明, LLM 可以产生关于威胁行为和缓解措施的问题和答案。但是 GPT-3.5 可能很难评估生成问题的质量, 需要通过专家的评价来补充。

不同来源的情报数据中会存在重复和互补的内容, 需要使用知识融合来将多个来源的统一实体或概念进行融合。知识融合需要在两个层面进行, 一个是本体层面即本体匹配, 另一种是实体级别即实体对齐。本体匹配是将威胁情报领域中的一些概念和复杂关系抽象成语义网络。Syed 等人[69]提出了统一网络安全本体(Unified Cybersecurity Ontology, UCO), 提供对网络安全领域的通用理解。UCO 本体能够较好的实现对 CVE、CAPEC、STIX 等本体的映射。实体对齐主要包括两大类, 基于相似性和基于嵌入的方法。Azevedo 等人[70]提出基于相似性度量方法来关联融合来自不同 OSINT 源的 IOC, 以丰富 IOC 的形式。Chen 等人[71]通过改进 TransE 形成了 MtransE, 将每种语言的实体和关系编码在单独的嵌入空间中, 以实现跨语言实体对齐。以往的知识融合研究成果, 为大语言模型提供了一定思路。大语言模型的语言理解能力可以进行跨语言的实体对齐, 例如将英文和中文报告中的同一恶意软件实体进行对齐。

大语言模型在威胁情报评估融合领域目前还未有较多研究成果, 目前的研究大多是评估大语言模型自身的性能。推测可能是与情报处理研究相比, 情报评估融合的研究基础较薄弱, 研究成果较少, 同时缺乏统一的威胁情报评估标准, 再加上大语言模型本身产生的结果具有一定的不可信性, 这些因素共同限制了其在评估融合领域的广泛应用。因此, 目前该领域的成果相对有限。在下一节中, 我们将讨论大语言模型在未来情报评估领域方向的潜在应用及发展方向。

### 3.4 讨论与总结

本节总结论述了 LLM 增强威胁情报自动采集、预处理和评估融合相关研究工作, 其中不少研究将大语言模型应用并产生了较好的效果。基于技术应用场景的不同, 将现有 LLM 增强威胁情报聚合相关技术进行分类, 具体内容如表 4 所示, 讨论总结如下:

**表 4 LLM 增强威胁情报聚合技术汇总**

**Table 4 Related works of LLM-enhanced threat intelligence aggregation**

| 研究方向 | 技术应用场景 | 文献 | 年份 | 主要方法/基于模型 | 数据集 | 数据集开源 | 实现效果 |
|---|---|---|---|---|---|---|---|
| 威胁情报自动采集 | 长文本 | [48] | 2016 | 博客爬虫+话题过滤 | 博客文章 | × | 从博客网站采集包含 IOC 的技术报告类文章 |
| | | [49] | 2018 | SVM | 安全文章 | × | 从开源数据中过滤掉安全不相关信息, 精度达到 85%, 召回率 91% |
| | | [50] | 2022 | TextRCNN | 安全文章 | × | 从安全论坛、博客和其他平台获取网络威胁情报数据 |
| | | [51-53] | 2023 | GPT4 | 未提及 | × | GPTs 工具: 实现威胁情报采集检索聚合 |
| | | [54] | 2024 | GPT3.5 | 安全报告 | × | 答案相关性得分高达 95% |
| | 短文本 | [55] | 2021 | Leiden 社区检测算法 | 推文 | × | 跟踪 Twitter 网络威胁相关用户账户, 获取情报 |
| | | [56] | 2023 | 主题建模 | 推文 | √ | 发现 Twitter 中的威胁信息 |
| | 暗网 | [57] | 2023 | DarkBERT: RoBERTa | DUTA-10K、CoDA | √ | 与其他预训练语言模型相比, DarkBERT 更适合暗网网络安全相关的任务。 |
| | 流量数据 | [58] | 2023 | HuntGPT:GPT3.5 | KDD99 | √ | 领域知识准确率 72%-82.5% |
| | | [59] | 2023 | GPT3 | ToN IoT | √ | 生成网络流量 |
| | 日志 | [60] | 2023 | GPT3 | NLP2Bash HaaS Cyberlab AcmeData | √ | GPT3 分析日志性能上略差于 CodeBERT |
| | | [61] | 2023 | GPT3.5 | CERT、LANL | √ | 假警报识别 AUC 值高达 94% |
| | | [62] | 2023 | GPT3.5 | 未提及 | × | 验证对命令的回复输出准确率高达 92% |
| 威胁情报预处理 | 数据标准化 | [63] | 2023 | GPT3.5 | 204 份报告及其 STIX | √ | 生成 STIX 结构化化报告 |
| | 数据标注 | [64] | 2023 | GPT3 | 2021 Microsoft Exchange Server data breach | √ | 数据增强, 生成标注数据 |
| 威胁情报评估融合 | 情报源评估 | [65] | 2019 | 自定义一组基本威胁情报指标 | OSCTI 付费情报 | × | 定义 IP 和 HASH 情报源的评估标准 |
| | 情报内容评估 | [66] | 2021 | BERT | 推文 | √ | 评估推文中的 IOC 可信可用性 |
| | | [67] | 2023 | GPT4 | NetEval | √ | GPT4 在网络领域的知识能力达到类人效果 |
| | | [68] | 2023 | GPT3.5 | MITRE ATT&CK | × | 评估模型对 MITRE ATT&CK 中威胁行为的理解 |
| | 情报融合 | [69] | 2016 | UCO | NVD | √ | 融合并集成网络安全中的异构数据 |
| | | [70] | 2019 | 相似性度量 | MISP TIP | × | 提高 IOC 的完整性 |
| | | [71] | 2017 | MTransE | WK3l | √ | 实现跨语言的知识对齐 |

(1)威胁情报采集预处理

实现细粒度和潜在个性化的威胁情报数据采集是一大难点, 目前大部分是粗粒度的采集处理 CTI, 通过训练分类器进行信息分类。在采集预处理过程中, 大语言模型的融入可以显著提高自动化水平, 减轻人工负担, 还提升了整体处理的效率和准确性。

LLM 可以自动定制化采集威胁情报数据如 GPTs[51-53], 提高采集效率。威胁情报通常是用不同的自然语言和格式编写, LLM 的多模态优势可以汇聚不同语言的报告内容, 如使用中文、英文、韩文等编写的报告。在处理文本内容时, LLM 在理解复杂上下文和语义关系方面表现突出, 能够筛选提炼高质量

的情报数据, 从而实现细粒度的情报收集; 在数据预处理部分, LLM 可以自动生成数据清洗规则和预处理操作流程的代码, 提高效率。面对威胁情报数据集的不足, 尤其是异常数据和良性数据间的不平衡问题, LLM 能够有效地生成补充数据, 扩展数据集的覆盖范围。受文献[59]启发, 可以微调 LLM, 设定提示词, 生成某一标签的数据。如生成描述钓鱼邮件攻击的句子, 从而补充恶意数据, 减少人工采集和标注的时间; 但同时 LLM 也面临一些问题, 例如被攻击者恶意利用生成虚假数据进行数据中毒攻击。因此其生成的数据真假难辨, 利用 LLM 生成的数据的可信性和可用性还有待进一步验证。在情报采集处理过程中, 传统 ML 方法在性能评估上依赖于准确率、精确率、召回率和 F1 值。而 LLM 的评估除了使用准确率和精确率, 还需要不同的指标, 如使用 NLP 中机器翻译指标 ROUGE 来衡量生成情报的质量。

威胁情报多源异构, 时效性较短、可能随时失效, 因此可以运用 LLM 技术自动化探索未知的情报来源, 集成更多样化的数据源和类型, 从而提高采集效率并缩短情报在外流通时间。此外利用 LLM 的生成能力构建高交互蜜罐等主动式防御手段, 可以获取更全面和主动式的威胁情报。

(2)威胁情报评估融合

在情报评估融合领域, 可以结合 LLM 的理解能力对情报内容进行融合, 并利用其分析理解关联能力, 对情报源和情报内容进行质量和可靠性的评估。可以让 LLM 提供支撑情报有效性验证的证据, 便于人工进行判断。或者采用思维链或思维树[72]的方法, 训练 LLM 以模仿主题专家(Subject Matter Expert, SME)的方式, 逐步推理并自行决定下一步的选择, 以完成对情报内容质量的评估。还可以在 LLM 基础上引入额外的事实来源, 如 BRON 属性图[73-74]这样的额外事实来源, 以提高评估的准确性和深度。此外, 文献[75]已证实交互语言模型具有捕获事实错误的潜力, 因此可以使用两个 LLM 交叉校验, 一个生成, 一个审查, 后者引入问题来发现不一致, 提高对情报评估的准确性。LLM 还可以被用来识别情报信息中的不一致性和冲突问题。例如, 通过比对不同来源的信息, LLM 可以识别并标记出相互矛盾或冲突的情报。针对情报冲突, 能够理解和解释数据之间的差异, 帮助安全人员更好地分析冲突的原因。

## 4 LLM 驱动威胁情报分析

威胁情报分析作为整个生命周期的关键环节, 涵盖了威胁情报信息提取识别, 威胁情报自动生成和知识图谱构建三大方面。大语言模型在这些环节中的应用, 显著提升了威胁情报质量和分析的深度。

### 4.1 威胁情报信息提取识别

威胁情报信息提取识别这一过程涉及到实体识别(Named Entity Recognition, NER)、关系抽取(Relation Extraction, RE)、TTPs 提取、漏洞描述映射等多个方面, 这些步骤共同构成了威胁情报分析的基础, 为后续的情报生成和知识图谱构建打下坚实的基础。

**实体识别** LLM 已被证明可以有效的从 CTI 文本中识别实体, 例如漏洞利用、目标、恶意软件、漏洞、IOC 等。Ranade 等人[76]通过微调 BERT, 可以识别专门的网络安全实体, 可高精度高效的用于特定网络安全下游任务。Aghaei 等人[77]基于 RoBERTa, 通过定制分词器策略和权重调整方法构建了网络安全领域的语言模型 SecureBERT, 实验发现在网络安全相关的 NER 任务上对效果最佳。文献[78]创建了一个用于网络安全命名实体识别的开源 Python 库, 使用 RoBERTa 训练, 用于提取不同 IOC 的启发式方法以及用于通用实体类型的公开 NER 模型 cyNER。Bayer 等人[79]提出一种基于 BERT 的网络安全语言模型, 在 CTI 文本分类、相关性分类和网络安全 NER(识别 CVSS(Common Vulnerability Scoring System, 通用漏洞评分系统)中的恶意软件名称、版本、NVD 描述的攻击复杂性)任务上优于其他模型。

**关系提取** LLM 可以推理语义关联, 完成关系提取任务。文献[80]使用 RoBERTa 模型的学习深层语义知识能力, 推理出攻击技术和防御对策描述文本之间存在的语义关联。实验结果表明该方法能以良好的准确性为安全分析师提供有用的建议, 特别是与无法表现出进行此类关联所需的语义和上下文理解的基线方法(word2vec、TF-IDF)相比。Wadhwa 等人[81]评估 LLM 在标准 RE 任务上的表现, 发现 GPT-3 的少样本提示实现了接近 SOTA 的性能。Flan-T5 在少样本设置中的能力不强, 但通过思想链的解释(由 GPT-3 生成)对其进行监督和微调, 可以产生 SOTA 的结果, 作者发布这个模型作为 RE 任务的新基线。这证明了 LLM 模型可以高效完成 RE 任务, 后续可以尝试使用 LLM 做威胁情报领域的关系提取。

**TTPs 提取** LLM 可以从 CTI 中获取更强的语义关系, 理解上下文内容, 从而提取威胁行为、攻击类型, 提供对 TTPs 描述的解释。这种提取相比单独的 IOC 提取, 为防御者采取有效行动提供了上下文背景。LADDER[82]可以大规模地从 CTI 报告中推理攻

击特征, 并提取出攻击模式。ThreatQA[83]通过问答的方式从 CTI 报告中自动获取有关网络威胁的知识。对于情报分析, 较难的是从自然语言文本中提取出 TTPs 如攻击技术, 因文本描述的模糊性以及人们的安全专业知识不同, 人们对文本中描述的攻击技术会有不同见解。文献[84]研究了 LLMs 解释网络攻击描述并将其映射到 ATT&CK 的战术的效果。作者使用带标签的 ATT&CK 描述对小规模语言模型 SecureBERT 监督微调, 使用 CAPEC 数据集评估模型面对新的或者未见过的 TTPs 描述时, 能否正确理解文本并匹配到对应战术。结果表明 LLMs 存在固有的模糊性, 解释 ATT&CK 描述方面的能力比微调模型 SecureBERT 差。因此建议当标注数据较多的时候采用训练小规模语言模型会更加精确。反之, 在标注数据少的情况下, GPT3.5 能供提供更广泛以及更可解释的结果。文献[85]提出一种从推文中提取攻击技术特征的方法, 将基于语义角色标注(Semantic Role Labeling, SRL)的原型与 OpenAI 的 GPT-3.5-Turbo 模型和 text-embedding-ada-002 模型的性能进行了比较, 证明了可以利用大语言模型来准确处理推文并提取可操作的网络威胁信息。使用大语言模型替换原型中的词嵌入模型, 可以获得更好的结果。

**漏洞描述映射** 随着漏洞数量日益增加, 安全管理需要越来越多的结构化数据。除了对漏洞进行文字描述外, 安全工程师还必须对漏洞进行分类和评估, 并阐明其相关技术。这就需要漏洞描述映射(Vulnerability Description Mapping, VDM), 它是指将漏洞映射到常见弱点枚举(CWE)、常见攻击模式和 ATT&CK 技术。研究者[86-87]发现使用基于 BERT 的模型可以将 CVE 漏洞映射到 ATT&CK 技术。Liu 等人[88]通过评估 ChatGPT 将漏洞映射到 CWE ID 和 ATT&CK 技术 ID 任务上的性能, 探讨了闭源 LLM 在现实安全管理场景中的应用。其中漏洞描述映射到 CWE-ID 效果较好, 而映射到 ATT&CK 技术 ID 效果较差。推测可能因为 CVE 公开的数据集 CVE-CWE 质量高, GPT 能够较好的学习上下文内容, 而 ATT&CK 技术描述具有多样性及其数据量少, 数据质量不高, GPT 不能较好的正确理解相关技术内容, 导致误报率过高。普渡大学团队[89]提出了一个框架 VWC-MAP, 使用 BERT 和 T5 大模型进行 CVE 漏洞到 CWE、CWE 到 CAPEC 攻击技术的映射, 实验显示 CVE 与 CWE 关联起来的准确率达到 87%, 而 CWE 与 CAPEC 关联的准确率达到 80%。

## 4.2 威胁情报自动生成

非结构化 CTI 报告数量持续激增, 急需生成简化报告的自动化工具。结构化威胁信息表达 STIX 可以使组织以尽可能富有表现、灵活、自动化和可读的方式共享网络威胁信息。Perrina 等人[90]提出一种自动生成威胁情报报告的工具 AGIR, 通过 ChatGPT 自动化生成 STIX 标准的威胁情报, 在不引入幻觉的情况下实现了 99%的召回值。将信息抽取系统与 AGIR 相结合, 可以自动处理多个报告, 有助于 CTI 任务的自动化。评估采用定量和定性两种。定量评估使用精确度、召回率和 F1 分数三个指标来评估 AGIR 的准确性。对 AGIR 生成报告的定性评估, 主要关注以自然语言表达信息的流畅性、正确性和实用性, 分为句法评估和语言评估。句法评估采用 SLOR(Syntactic Log-Odds Ratio), 一种无参考评估中评估文本流畅性的事实标准。语言评估则是让专家从流畅性、正确性、实用性三个维度对报告质量进行评级。EclecticIQ 公司[91]演示了使用 ThreatIntelGPT 提取报告信息, 自动生成 STIX 机读情报, 还可以根据 STIX 中的威胁指标, 自动编写 Sigma 检测规则。

在情报的定义中提到了可执行的建议, 在这里可以理解为安全策略。McIntosh 等人[92]使用 GPT-4 模型生成网络安全策略以应对勒索软件攻击和数据泄露的问题。在涉及数据泄露的勒索软件攻击的情况下, 根据既定安全供应商的策略评估 GPT-4 生成的网络安全策略。研究发现, 通过输入提示词, GPT 生成的策略在某些情境下比人类生成的策略更有效、高效和完整, 为了解决 GPT-4 模型的不确定性, 研究采用了多次生成结果并选择最连贯和相关的输出。然而, 由于网络安全威胁的动态性和模型的限制, 需要人工验证以确保合规性。

## 4.3 知识图谱构建

知识图谱构建通常步骤是使用预训练语言模型、NLP 技术等进行实体识别、关系提取、形成实体关系三元组, 从而构建图谱。Li 等人[93]提出的 AttacKG 框架, 是首个在技术层面上从多个非结构化 CTI 报告中聚合攻击知识的工作。从 MITRE ATT&CK 技术示例和 CTI 报告获取数据, 利用 NLP 技术和基于学习的 NER 模型来识别文本中的实体, 自动提取结构化攻击行为图, 识别相关的攻击技术, 并将攻击行为图增强为技术知识图谱。Sarhan 等人[94]提出 Open-CyKG 框架, 提供了第一个基于开放信息抽取三元组构成的网络安全知识图谱。通过开发一个神经网络安全 NER 模型来识别相关实体, 标记模型生成的关系三元组, 能够从非结构化的 APT 报告中高效提取有价值的网络威胁信息, 并将检索到的数据用于构建知识图谱。Gao 等人[95]提出了一种自动收集和管理高质量的威胁

情报系统 SecurityKG。采用 CRF 模型来提取非结构化文本中与安全相关的实体, NLP 技术抽取高质量的威胁行为知识, 构建安全知识图谱。

而大语言模型正在重构 NLP 本身的技术体系。Pan 等人[96]系统地探究了黑盒的 LLM 和结构化的知识图谱(Knowledge Graph, KG)结合, 利用各自优势互补的可能性。提出三种统一 LLMs+KGs 的框架, 包括 KG 增强的 LLM、LLM 增强的 KG 和协作的 LLMs + KGs。LLM 增强的 KG 框架主要是利用 LLMs 的潜力, 来帮助处理 KGs 中的原始语料库并提取关系和实体。Sewak 等人[97]认为使用 OSINT 在很大程度上仍然是一个人类专家精心策划的知识检索任务。因此利用 GPT3.5、GPT4、ChatGPT、Bing Chat 构建威胁智能图谱的经验, 提出了利用 LLM 设计和开发企业知识图谱(Enterprise Knowledge Graphs, EKG)的新方法。评估使用分类性能指标 PRF 值即精确度、召回率和 F1 值。在检测恶意脚本的二分类任务上, LLM-EKG 达到 99%的召回率, 优于仅使用 LLM 的方法。这证明了 LLM 与知识图谱结合使用的有效性。这为构建更全面的威胁情报知识图谱, 提高威胁检测能力提供了可能性。Neo4j 团队[98]从漏洞描述中自动识别具体的应对措施, 并生成全面的分步指南。支持生成式 LLM(OpenAI GPT)+知识图谱(D3fend KG)的混合方法, 使从业人员能够将网络安全建议无缝应用到现实世界的使用案例中。文献[99]提出一个用于构建 CTI 知识图谱的端到端方法, 使用 ChatGPT 自动提取与攻击相关的实体及关系, 构建 CTI 知识图谱。在 13 份报告上评估, 采用 PRF 指标, 实验发现与 AttacKG(NER 任务)和 REBEL 模型(RE 任务)相比, ChatGPT 展示了更好的性能, 同时需要更少的手动干预和计算资源。这证明了 LLM 在资源匮乏的情况下的可行性和适用性, 特别是网络威胁情报领域。

### 4.4 讨论与总结

本节总结论述了 LLM 驱动威胁情报信息提取识别、威胁情报自动生成和知识图谱构建相关研究工作。基于技术应用场景的不同, 将现有 LLM 驱动威胁情报分析的相关技术进行分类。具体内容如表 5 所示, 讨论总结如下:

(1)威胁情报信息提取识别

传统威胁情报信息提取识别方法使用深度神经网络、NLP 技术等来学习上下文和语义关系。但面临以下几个问题: 首先标注数据稀缺且定制化。基本是针对特定任务标注专有数据。准确的数据标注和评估还需要人类专家辅助。其次, 威胁情报领域快速发展的性质需要不断更新基准数据集, 使其更加资源密集和耗时。第三, 威胁情报提取规则模板高度依赖专家知识。面对以上问题, 大语言模型能够很好的解决。它能够在少样本的情况下达到微调预训练模型效果, 近似 SOTA。甚至已被证明在零样本的情况下以高准确率完成 NER 任务, 比最先进方法表现还要好[100], 并且能够学习更深层的语义关联, 理解上下文, 准确提取关键信息, 提高效率的同时也显著提升了准确性。后续可以考虑利用 LLM 来完成在少样本条件下进行更高层面的威胁情报要素提取如安全事件的提取。在信息提取识别部分, 基本可以看作是分类任务, 因此 LLM 的评估指标与 ML 无异, 也是使用准确度等值来衡量性能。

(2)威胁情报自动生成

机读情报和人读情报都是重要的输出产物。可以给定提示模板, 利用 LLM 自动化从文本中提取关键信息, 自动生成 STIX1、STIX2 等结构化报告, 或者利用 LLM 对获取的情报进行总结, 生成自然语言人读报告。因为涉及文本内容的生成, 在对 LLM 方法评估时, 会增加定性评估。一般采用文本生成领域的指标如句子通顺度指标或者以专家填写调查问卷的形式评估生成内容的质量。

(3)知识图谱构建

传统的知识图谱构建过程采用 Pipeline 形式, 实体识别-关系抽取-构建实体关系三元组-构建知识图谱, 这种方式存在一定局限性比如上下游任务错误的传播, 容易忽略任务之间存在的关系, 并产生一些冗余信息。而研究者们发现知识图谱与 LLM 二者互补, LLM 出色的自然语言理解和表达能力, 结合知识图谱结构化特点, 可以实现自动化提取、分析和增强威胁情报数据, 有效分析和解释复杂的威胁情报, 从而构建更全面的威胁情报知识图谱。在这一过程中, 可以有两种方法, 一种利用 LLM 来增强现有的威胁情报知识图谱; 第二种将 LLM 与知识图谱进行交互融合, 以实现更优的结果。这样的结合不仅提高了数据提取和分析的自动化效率, 还有助于构建更为全面和深入的威胁情报知识图谱, 从而更有效地应对安全威胁。此外, 还可以发展 LLM 在提高用户交互性和知识图谱可解释性方面的能力。而知识图谱也可以为 LLM 补充世界知识。为解决 LLM 数据过时和领域特定限制。Feng 等人[101]提出采用知识编辑和检索增强两种策略结合外部信息来增强 LLM。因此可以尝试使用知识编辑和检索增强结合外部威胁情报库信息, 来增强大语言模型对威胁情报领域的理解能力。

**表 5 LLM 驱动威胁情报分析技术汇总**

**Table 5 Related works of LLM-driven threat intelligence analysis**

| 研究方向 | 技术应用场景 | 文献 | 年份 | 主要方法/基础模型 | 数据集 | 数据集开源 | 实现效果 |
|---|---|---|---|---|---|---|---|
| 威胁情报信息提取识别 | 网络安全实体识别 | [76] | 2021 | CyBERT: BERT | 网络安全文本 | × | NER 准确率 80% |
| | | [77] | 2022 | SecureBERT: RoBERTa | MalwareTextDB | √ | NER 准确率 86% |
| | | [78] | 2022 | CyNER: RoBERTa | MITRE ATT&CK 中软件类别下引用的约 60 份威胁情报报告 | √ | NER 准确率 75% |
| | | [79] | 2022 | CySecBERT: BERT | CVSS、CySecAlert、MS Exchange | √ | 软件版本、名称和攻击复杂性的实体识别 F1 值分别为 93%、89%、35% |
| | 攻击防御关系提取 | [80] | 2022 | RoBERTa | ATT&CK 和 D3FEND 的攻击及防御文本描述 | × | 攻击技术关联效果比基线方法(word2vec、TF-IDF)好 |
| | 通用领域关系提取 | [81] | 2023 | GPT3、Flan-T5 large | ADE、CoNLL、NYT、DocRED | √ | 少样本下效果接近 SOTA |
| | 攻击模式提取 | [82] | 2023 | LADDER:BERT,RoBERTa,XLM-RoBERTa | 恶意软件标注数据集 | √ | 提取攻击特征并映射到 ATT&CK |
| | 提取威胁知识 | [83] | 2022 | ThreatQA: RoBERTa | OSCTI | × | 知识库问答来促进网络威胁知识获取 |
| | TTPs 描述映射 | [84] | 2023 | GPT-3.5、Bard 与 BERT、SecureBERT | ATT&CK 和 CAPEC 描述文本 | × | TTP 描述映射 ATT&CK 技术方面 GPT 没有 SecureBERT 精确 |
| | 攻击技术提取 | [85] | 2023 | GPT-3.5-Turbo 和 text-embedding-ada-002 | IoCMiner 和爬取的 21 年推文 | √ | 实现了 88.59%的精确度 |
| | CVE 映射到 ATT&CK | [86] | 2022 | CVE2ATT&CK: BERT、SciBERT、SecBERT | ATT&CK 技术 | √ | F1 分数达到了 47.84% |
| | CVE 映射到 ATT&CK | [87] | 2023 | SMET: ATT&CK BERT, LLM | ATT&CK 技术、CVE 条目 | √ | SMET 映射效果比 LLM 聊天机器人好 |
| | CVE 映射到 CWE-ID、ATT&CK 技术 ID | [88] | 2023 | ChatGPT | CVE-CWE、CVE-ATT&CK、CVE-ATT&CK-builtOnBRON | √ | 漏洞映射到 CWE 效果比映射到 ATT&CK 技术 ID 好 |
| | CVE 映射 CWE、CWE 映射 CAPEC 攻击技术 | [89] | 2022 | VWC-MAP: BERT、T5、RoBERTa,DistilBERT | CVE、CWE、 CAPEC | × | CVE 映射到 CWE 成功率 87%, CWE 映射到 CAPEC 技术成功率 80%。 |
| 威胁情报自动生成 | STIX 报告生成 | [90] | 2023 | ChatGPT | STIX 图的 JSON 表示 | √ | 召回率 99%, 报告编写时间减少了 40% 以上 |
| | | [91] | 2023 | ChatGPT | 未提及 | × | 自动提取报告信息生成 STIX 情报 |
| | 网络安全策略生成 | [92] | 2023 | GPT-4 | 安全建议文章 | × | 某些情况下优于人类生成的策略 |
| 知识图谱构建 | 传统知识图谱构建 | [93] | 2022 | NLP+图神经网络 | ATT&CK 技术示例和 CTI 报告 | √ | 从 CTI 报告中构建技术知识图谱 |
| | | [94] | 2021 | 基于注意力的开放信息提取模型 | MalwareTextDB、Microsoft Security Bulletins、CTI 报告 | √ | Open-CyKG 的组件优于最先进的模型 |
| | | [95] | 2021 | CRF+NLP | OSCTI 报告 | × | 提取高保真的威胁行为知识, 构建安全知识图谱 |
| | | [96] | 2023 | LLMs+ KGs | 未提及 | × | 探索 LLM 与 KG 结合的方法 |
| | 威胁智能图谱构建 | [97] | 2023 | CRUSH:GPT3.5、 GPT4、ChatGPT、Bing Chat | 恶意脚本 URL TTP | × | LLM-EKG 在检测恶意脚本的任务上召回率高达 99% |
| | | [98] | 2023 | OpenAI GPT+知识图谱(D3fend KG) | 未提及 | × | 自动生成网络攻击对策 |
| | | [99] | 2023 | ChatGPT | ATT&CK 报告 | × | NE 和 RE 任务中 F1 分别取得 78%和 56% |

## 5 LLM 赋能威胁情报应用

威胁情报应用按照时序事前事中事后可以分为网络攻击威胁预测、网络攻击威胁检测和网络攻击威胁溯源等关键阶段。大语言模型在这些阶段中扮演着重要的赋能角色。本节梳理已有或潜在的相关工作, 以发掘基于大语言模型和威胁情报技术的应用和技术融合可能, 并通过这种技术融合, 探索新的应用场景, 从而为威胁情报应用提供更强大的赋能支持。

### 5.1 网络攻击威胁预测

网络攻击威胁预测指根据历史威胁情报数据和当前趋势, 预测和警告即将发生的安全威胁。大语言模型可以辅助人员更好的分析情报数据, 解释攻击行为模型和意图, 从而有效预测识别新威胁。

网络攻击威胁预测通常关注攻击模式、安全报告和社交媒体、日志记录和流量等信息。为了对具有共同攻击模式并使用相同攻击工具的复杂网络威胁及时调查, Noor 等人[102]提出一种基于机器学习的网络威胁识别框架。通过提取网络威胁事件报告中的 TTPs 和威胁之间的关联形成语义网络, 最后依据概率关系以 92%的准确度和较低的误报率预测网络威胁。

研究者发现一些漏洞会在官方数据库发布之前就已经在社交媒体上开始了讨论, 这为早期的漏洞预警提供了可能性。已有研究者证明, 使用 Twitter 来预测软件漏洞的可利用性, 以提前产生告警[103-106]。除了对漏洞告警, 还可以预测漏洞的严重性。Zong 等人[106]分析推特用户对漏洞威胁严重性的评论, 基于这些评论预测漏洞严重性, 进而预测真实世界的漏洞。

预测攻击方式可以有效防御网络威胁。Ghafir 等人[107]为了应对 APT 攻击, 提出基于机器学习的系统 MLAPT, 主要包括威胁检测、警报相关性和攻击预测。其中攻击预测是供网络安全团队用于确定早期警报发展为完整 APT 攻击的概率。实验表明该系统可以在早期阶段预测 APT, 预测准确率为 84.8%。Zhao 等人[108]使用属性异构注意力网络和转导学习来预测网络攻击偏好。对攻击者、漏洞、被利用脚本、受损设备、被入侵平台以及描述它们之间相互依赖关系的 20 种元路径进行建模, 构建了一个攻击事件的属性异构信息网络。然后分别提出了基于注意力机制和转导学习的攻击偏好预测模型。最后通过叠加这两个基本的预测模型, 构建了一个网络攻击偏好的自动预测模型, 该模型能够集成元路径和元图中更全面、更复杂的语义信息来描述入侵者的攻击偏好。基于真实数据的实验结果证明, 该方法在预测入侵者的网络攻击偏好方面优于最先进的方法。

由于大语言模型尤其是 GPT 类型自身特性, 更适合于生成而非判别, 并且不能确保其本身生成内容的正确和可靠性, 因此少有直接使用大语言模型进行威胁预测预警的应用。但可以借助大语言模型的语言理解能力, 来分析和概括文本类型的攻击数据, 帮助安全分析人员更高效地理解和处理大量信息。基于这一初步分析, 再采用专门的预测模型来进行威胁预测, 从而帮助安全团队缩短分析时间, 并迅速响应潜在的网络威胁。

### 5.2 网络攻击威胁检测

网络攻击威胁检测可以按照应用场景分为网络入侵检测、恶意 URL 检测、恶意代码检测和钓鱼邮件检测。大语言模型与威胁检测阶段结合可以显著增强对潜在威胁的发觉能力并且提高检测准确性。

各平台恶意软件攻击不断升级, 急需开发自动化方法来检测防御恶意软件。文献[109]基于 GPT-2 构建了恶意软件检测的垂直领域语言模型。在恶意软件检测性能实验上, 发现特定领域语言模型 DSLM-GPT2 优于通用语言模型 GLM-GPT2 取得了 86.0%的 F1 分数。恶意代码分析常需要静态分析工具, 但其具有一定的误报率, 对开发人员的生产力提出了挑战。文献[110]发现 FalconLLM 在识别复杂模式和复杂漏洞方面有巨大潜力, 因此基于 FalconLLM 微调形成 SecureFalcon 模型。该模型经过训练可以区分易受攻击和不易受攻击的 C 代码示例, 检测软件漏洞准确率高达 94%。Sun 等人[111]基于 GPT-3.5-Turbo 设计了 GPTScan 框架, 使用自然语言进行漏洞描述, 通过静态分析、情报匹配等将最终代币合约的漏洞检测精度提升到 90%以上, 召回率超过 70%, 同时系统检测到了人类未发掘到的真实逻辑漏洞。此外, GPTScan 在分析不易受攻击的合约数据集时, 误报率极低, 仅 4.39%。评估该方法在时间和性能上的表现时, 发现可以达到扫描每千行 Solidity 代码平均只需 14.39 s 和 0.01 美元。在恶意代码分析时重要的是对代码程序的理解。模糊测试是一种常见的软件测试和动态分析技术。Zhao 等人[112]使用 CodeBERT 和 UniXcoder 代码大模型, 提出一种模糊调优, 在给定预训练的 LLM 的情况下提高程序理解和代码表示学习的性能。实验证明在包括代码克隆检测和代码分类在内的两个程序理解任务上大大优于当前的最先进技术, 并取得了新的 SOTA。文献[113]通过情报知识蒸馏、监督微调等方法构建了基于 BERT 的 URLTran 分类器。相较于传统 BERT 方法, 该模型对

URL 分类的准确率提高了 9%, 参数减少了 175 倍。通过情报关联分析和协同检测, 大语言模型能更高效地分析实时数据流中的文本信息, 完成潜在威胁的快速检测。

Heiding 等人[114]探讨了使用 LLM 在生成和检测钓鱼邮件的应用。在检测网络钓鱼意图上, 使用四种流行 LLM(GPT、Claude、PaLM、LLaMA)和人工检测做对比, 结果显示, 大语言模型表现出强大的检测恶意意图的能力, 即使面对不明显的钓鱼邮件, 其检测性能有时会超过人类水平, 但准确度往往会略低于人工审核。

大语言模型可以推理各类威胁行为的上下文信息, 针对性地使用威胁情报进行攻击模式识别, 以及时发掘攻击行为。文献[115]基于 SecurityBERT 和 FalconLLM 构建了网络威胁检测领域的预训练语言模型 SecurityLLM, 该模型在网络威胁检测任务中准确度方面超过了大部分传统机器学习和常用深度学习方法。其在收集的安全数据集上进行实验, 能识别 14 种不同类型的攻击, 总体准确率为 98%, 但存在数据集需要频繁更新的局限性。Tran 等人[116]将 GPT-2、BERT 和 8 个传统机器学习模型应用于入侵检测过程, 发现两个语言模型的精确度明显优于传统机器学习模型, 同时通过优化数据集降低了数据分布差异、重复性数据等对模型性能的损失。通过情报关联分析和行为特征映射, 大语言模型能更准确地检测出各类威胁行为, 完成对安全事件的实时判定。

### 5.3 网络攻击威胁溯源

网络攻击威胁溯源主要是确定和追踪网络攻击的来源及其背后的行为者。通过关联分析、推理等步骤完成对攻击主机、控制主机、攻击者和攻击组织的溯源归因。例如通过分析恶意样本时间戳、钓鱼邮件使用的语言, 可以推断出攻击者时区和国家。一般网络攻击溯源研究包括基于蜜罐、流量和威胁情报。下面主要介绍基于威胁情报的攻击溯源技术。

对攻击事件的挖掘能够实现对攻击者的溯源。Huang 等人[117]提出一种基于图模型的网络攻击溯源, 通过建立网络攻击事件溯源本体模型, 融合网络攻击事件中提取的线索数据和威胁情报数据, 形成溯源关系图, 然后使用图算法学习攻击事件特征, 基于历史攻击事件特征向量训练 SVM 分类器。结果显示使用 SVM 能够准确的对 APT 攻击事件溯源, 判定网络攻击事件与攻击者的归属关系, 从而达到攻击者的溯源, 其准确率高达 95%。黑客组织的识别为黑客的追踪溯源、攻击组织画像等工作奠定了基础。Xu 等人[118]提出一种基于异构图注意力网络的黑客组织识别方法 HGHAN。通过从图网络中提取黑客组织特征, 对黑客群体的 WEB 攻击链建模, 发现该算法相比其他异构图节点嵌入算法具有更好的识别效果。网络攻击的溯源分析需要结合威胁情报和既有检测结果, 完成各类推理工作。攻击路径追踪溯源是网络攻击溯源的重要部分, Zang 等人[119]设计了一种异构威胁情报融合方法来实时重建攻击场景, 以发现关键攻击路径。通过分析每个异构情报的因果关系并关联, 将多阶段攻击场景重建为社区发现问题, 利用语义相关权重和社区检测算法挖掘攻击场景, 实验表明该方法可以准确的发现隐蔽的 C2 路径。文献[120]提出了一种基于攻击溯源知识库和图卷积网络的 DDoS 攻击路径追踪系统(AT-GCN), 能够利用知识库中的图结构重现攻击路径。此外, 还能够根据管理员对溯源性能的要求, 动态推荐最佳溯源算法。

Scanlon 等人[121]尝试将 GPT-4 应用在电子取证的各个环节, 发现其强大的思维链生成能力能够辅助溯源工作者借助各类威胁情报完成场景还原等分析工作。

以往研究是采用图算法结构学习特征, 使用机器学习进行分类, 完成溯源。后续将大语言模型的推理关联能力和威胁情报技术的有机结合, 能极大提升全面溯源取证分析的自动化水平, 进而缩短对攻击响应的时间。

### 5.4 讨论与总结

本节总结论述了 LLM 赋能网络攻击威胁预测、网络攻击威胁检测和网络攻击威胁溯源的相关研究工作。基于技术应用场景的不同, 将现有 LLM 赋能威胁情报应用相关技术进行分类。具体内容如表 6 所示, 讨论总结如下:

(1)网络攻击威胁预测

网络攻击威胁预测通常使用回归模型、机器学习等方法, 少有语言模型, 其本身不适合做状态预测。因此没有发现大语言模型直接应用于威胁预测预警的研究。但是受回归模型在威胁预测推文方面工作[106]的启发, 大语言模型可以通过分析漏洞报告等情报层级的文本数据, 结合系统当前态势来预测哪些漏洞可能被攻击者利用, 从而帮助安全团队提前采取措施修复潜在的安全风险。另一方面, 也可以将大语言模型融入威胁预测预警的前置处理阶段, 借助其对数据的深度分析和理解能力, 汇总和提炼关键的威胁特征, 这些特征随后可以作为预测模型的输入数据或者中间步骤的关键信息, 从而增强整个威胁预测系统的性能和准确性。

**表 6　LLM 赋能威胁情报应用技术汇总**

**Table 6　Related works of LLM-empowered threat intelligence application**

| 研究方向 | 技术应用场景 | 文献 | 年份 | 主要方法/基础模型 | 数据集 | 数据集开源 | 主要效果 |
|---|---|---|---|---|---|---|---|
| 网络攻击威胁预测 | 网络威胁预测 | [102] | 2019 | 基于概率的机器学习 | CTIR 和 ATT&CK 文章 | × | 预测网络威胁, 准确度 92% |
| | 漏洞预警 | [103] | 2017 | 词典匹配 | 推文 | × | 基于 Twitter 数据源, 预测真实世界漏洞 |
| | | [104] | 2016 | 语义 WEB RDF | | × | |
| | | [105] | 2015 | 机器学习 | | × | |
| | | [106] | 2019 | | | √ | |
| | 攻击方式预测 | [107] | 2018 | 机器学习相关性 | 告警 | × | 能够早期阶段预测 APT, 准确度为 84.8% |
| | 攻击偏好预测 | [108] | 2021 | 属性异构注意力网络、转换学习 | 攻击事件描述 | × | 预测入侵者的网络攻击偏好方面优于最先进的方法 |
| | 恶意软件检测 | [109] | 2022 | BiLSTM GPT-2 | 可执行文件 | × | 领域特定语言模型(DSLM-GPT2)和通用语言模型(GLM-GPT2)的 F1 分数分别达到 86.0%和 76.2% |
| | 软件漏洞检测 | [110] | 2023 | FalconLLM | FormAI | √ | 检测软件漏洞准确率高达 94% |
| | 智能合约漏洞检测 | [111] | 2023 | GPT-3.5-Turbo | Top200 Web3Bugs DefiHacks | √ | 精确率 90%以上, 召回率 70%以上, 能发掘新漏洞 |
| 网络攻击威胁检测 | 模糊测试 | [112] | 2023 | CodeBERT 和 UniXcoder | POJ104、CodeNet | √ | 代码克隆和代码分类任务上取得 SOTA |
| | URL 内容检测 | [113] | 2023 | URLTran | 客户遥测数据 | × | 相比 BERT 准确率提升 9%, 模型效能极大优化 |
| | 钓鱼邮件检测 | [114] | 2023 | GPT、Claude、PaLM、LLaMA | 钓鱼邮件 | × | 能够检测出不明显的恶意钓鱼邮件 |
| | 威胁模式识别 | [115] | 2023 | SecurityLLM | EdgeIIoTset | √ | 能识别 14 种不同类型的攻击, 总体准确率为 98% |
| | 入侵检测系统 | [116] | 2022 | GPT-2、BERT | 入侵检测数据 | √ | 语言模型相对传统机器模型准确率、鲁棒性更高 |
| | 攻击者溯源 | [117] | 2021 | 图算法+SVM | OSCTI | × | 将攻击事件判定到攻击者, 准确率 95% |
| | 黑客群体溯源 | [118] | 2022 | 异构图注意力网络 | HCRL | √ | 较好的识别黑客群体 |
| 网络攻击威胁溯源 | 攻击路径溯源 | [119] | 2023 | 语义关联+社区算法 | DARPA 2000、CICIDS 2017、CERNET | √ | 重建攻击场景, 发现攻击路径 |
| | | [120] | 2023 | 知识库+GCN | DDoS 数据 | √ | 重现 DDoS 的攻击路径 |
| | 溯源取证辅助 | [121] | 2023 | GPT-4 | 未提及 | × | 辅助攻击场景还原 |

(2)网络攻击威胁检测

大语言模型能够促进威胁情报更好地适应威胁检测任务, 完成网络入侵检测、恶意邮件检测等威胁检测应用, 并提高检测效率。大语言模型可以通过分析大型数据集、识别可疑指标和关联事件来帮助主动寻找威胁。它们可以帮助检测高级持续威胁 APT 组织并发现容易被忽视的秘密攻击活动。除此之外, 大语言模型可以基于预设模板和已知的威胁情报信息(如 IOC), 自动生成用于威胁检测的规则, 如 YARA, Sigma, Snort 规则等, 减少人工编写规则的时间。大语言模型可以通过用户交互或系统状态采集, 自适应修改威胁检测策略, 促使系统更高效地调取各类威胁情报, 采取更加合适的威胁检测方案。例如, 当资产价值较高时, 偏向于选择查全率较高的威胁检测模式, 以减少威胁漏报; 当资产价值较低时, 可以选择查准率较高的威胁检测模式, 以降低威胁误报。大语言模型具备的多类模型输出能力使得其能根据实际资产价值、系统算力资源、数据规模质量、情报获取难易等因素自适应调节系统检测性能, 在一定程度上能缓解误报率与漏报率、查准率与查全率间的固有矛盾。在使用 LLM 进行威胁检测时, 评估该方法的性能指标除了 PRF, 还会将误报率、假阴性率、运行时间、花费成本等作为参考。

(3)网络攻击威胁溯源

威胁溯源的挑战在于构建完整的攻击链, 将分散的攻击事件有效地关联起来。此外, 基于历史情报

数据进行分析和逻辑推理, 以实现对攻击源的精确追踪也是一大难题。而大语言模型具备的多模态内容支持、深层逻辑推理、关联分析等能力在数字取证、攻击路径还原、攻击场景重构、攻击组织画像构建、溯源图生成、溯源报告和关联情报生成等中均存在较大的应用潜力, 后续可以考虑通过大语言模型提升溯源取证整体的工作效率。例如可以将大语言模型融入威胁响应溯源系统, 与组织内部数据库(如 Elasticsearch 数据库)结合。根据用户提供的某一个可疑行为或 IOC, 自动生成并执行相应的数据库查询语句, 快速检索并提供相关联的数据。随后, 利用大语言模型的推理分析能力, 系统能够进行初步的溯源分析, 辅助安全专家迅速追踪历史信息并对潜在的恶意活动做出有效响应。

## 6 挑战与未来

根据大语言模型在威胁情报领域的既有相关工作和技术研究现状, 可以提出四个未来可能热门的研究方向。

1)通过大语言模型降低威胁情报生产成本, 提高威胁情报输出质量。大语言模型能够有效提升威胁情报全生命周期的自动化水平, 显著节省威胁情报的分析、研判、结构化生成、共享、反馈等环节的时间人力成本。由于大语言模型具备良好的类人表达方式和人类偏好习惯信息, 可以考虑使用大语言模型对情报生命周期后半段的工作进行着重优化。例如通过社交舆论、漏洞公告、安全新闻等信息的挖掘, 推测新生成的威胁情报短期内的安全价值, 优化情报推送、传播、解释机制, 降低情报共享成本。

2)通过大语言模型推动威胁情报更好地完成威胁检测、威胁溯源等应用层级工作。可以使用大语言模型自动识别和提取各类安全报告、威胁情报、端点日志等数据中的威胁信息, 并进行规则筛选和信息匹配, 以促进“数据和情报结合”的更全面的威胁检测方式, 提升系统对 APT 攻击的检测分析能力。可以使用大语言模型辅助完成安全问题的推理分析, 在威胁情报的基础上更好地完成实体威胁查询、线索关联拓展、威胁预测与预警、威胁画像和主动防御等工作。可以使用大语言模型帮助用户完成交互式安全任务, 通过编排安全流程、自动化安全报告等提升用户或系统的安全问题处置能力, 也可以将自然语言信息转化为机器可读的查询语句、执行代码等来提升查询效率, 实现安全人机交互。

3)将大语言模型融入威胁情报平台、网络威胁分析平台等安全产品的生态体系中。通过研究传统模型与大语言模型在威胁情报子任务或其他安全任务中的具体表现, 探索新的安全产业技术路线和安全解决方案。例如, 在资源受限的条件下进行安全产品开发, 可以考虑在入侵检测环节沿用基于规则的方式, 威胁分析环节仍使用深度学习模型, 而在威胁情报生成、利用和安全人机交互等环节考虑使用大语言模型技术优化输出, 以确保整个网络安全系统具备更好的安全效果和更低的安全成本。

4)通过解决模型自身问题, 设计新应用场景来促进大语言模型更好地应用于威胁情报工作。需要更规范高效的安全评测、安全过滤和安全框架等, 确保用户使用大语言模型过程的安全合规, 满足保密性、完整性、可用性、不可否认性等信息安全基本要素。同时, 也需要构建更大规模优质的安全数据集来提升模型性能上限, 或探寻新的安全需求和安全场景, 设计大语言模型解决威胁情报或者其它安全问题的具体方案, 实现该交叉领域应用层级上的更多创新。

同时, 在大语言模型和威胁情报交叉的后续研究过程中, 我们也可以思考以下四组问题。

1)大语言模型技术的安全效益问题。虽然大语言模型在数据处理能力、迁移泛化能力、多模态学习支持、内容生成等方面展现出了区分于传统模型的明显优势, 但过高的计算资源开销、环境依赖严重、数据偏见等问题对大语言模型安全应用产生了一定程度的限制。在实际应用时必须考虑当前任务场景大语言模型是否具备不可替代性, 以及整个安全解决方案的投入产出比是否符合用户实际条件。

2)新的攻防辩证关系。大语言模型的出现并不能让安全工作一劳永逸, 反而推动了自动化黑客攻击的浪潮。在未受限的情况下, 强大的内容生成能力降低了黑客攻击门槛, 缩短了攻击过程时间。但大语言模型自身也为应对这些新问题和更好地解决旧问题提供了支持, 各类新的安全服务能有效缩短攻击响应时间。“大语言模型对于网络安全产业是机遇还是威胁? 攻防不对称关系是愈演愈烈还是逐步减小? 今后的安全定义是小范围内的动态平衡, 还是随着攻防博弈加速呈现出周期性的此消彼长? ”这些都是大语言模型到来后较为热门的安全话题。

3)大语言模型和网络安全的关系。大语言模型的本质是目前较先进的人工智能技术工具, 对于网络安全来说是一把双刃剑, 既能生成虚假的威胁情报干扰安全系统, 也能输出高质量的威胁情报解决安全问题。同时, 大语言模型和网络安全存在相辅相成

的发展关系。安全的人工智能是实现人工智能安全应用的重要前提, 而人工智能的安全应用也能反向促进人工智能自身安全问题的解决。

4)技术应用过程存在的一些挑战。在实际应用过程, 用户和厂商可能出现不信任、利益冲突等问题, 例如“QA 对齐问题”指由于一些输出控制机制的存在, 用户有时无法获得自己想要的答案。而部分国家地区则是担心第三方大模型会威胁自身的信息安全, 严令禁止了 GPT 等大语言模型相关的产品。技术的发展必须要遵循法律法规, 符合资本市场规律才能走得长远。同时, 新的安全技术往往需要新的安全管理模式, 才能发挥最大效益。

## 7 总结

本文首先介绍了网络威胁情报的定义、分类和技术研究现状, 接着介绍了大语言模型的定义、发展历史、研究现状等信息, 以发掘大语言模型在威胁情报领域应用的可能。然后, 围绕威胁情报聚合、威胁情报分析和威胁情报应用三个大方向对大语言模型驱动威胁情报的相关技术研究进行了全面细致的分类和归纳, 总结了不同环节任务的研究现状、技术特征和潜在方向等信息。最后, 分析了大语言模型与威胁情报交叉, 以及在整个网络安全领域应用面临的问题, 并给出未来研究方向, 进一步推动该领域的繁荣发展。

## 参考文献


[1] Check Point. Check Point Research Reports a 38% Increase in 2022 Global Cyberattacks. http://blog.checkpoint.com/2023/01/05/38-increase-in-2022-global-cyberattacks/. Jan. 2023.

[2] TianJi Partners. Cyber Threat Intelligence Technical Guide[M]. SHANDONG: SHANDONG UNIVERSITY PRESS, 2021:200.
(天际友盟 网络威胁情报技术指南[M]. 山东: 山东大学出版社, 2021:200.)

[3] Xiang G, Shi C, Zhang Y S. An APT Event Extraction Method Based on BERT-BiGRU-CRF for APT Attack Detection[J]. *Electronics*, 2023, 12(15): 3349.

[4] Hou Y L, Xie X C, Hong Z K, et al. A Brief Discussion on the Current Situation and Development Trend of Network Information Security in the Era of Digital Economy[J]. *Journal of Shandong Industrial Technology*, 2023(3): 60-64.
(侯艳玲, 谢兴昶, 洪之坤, 等. 数字经济时代网络信息安全基本现状与发展趋势[J]. *山东工业技术*, 2023(3): 60-64.)

[5] 中国政府网. 《“十四五”国家信息化规划》. http://www.gov.cn/xinwen/2021-12/28/5664873/files/1760823a103e4d75ac681564fe481af4.pdf. 2021 年 12 月.

[6] Farukee M B, Zaman Shabit M S, Haque M R, et al. DDoS Attack Detection in IoT Networks Using Deep Learning Models Combined with Random Forest as Feature Selector[M]. Communications in Computer and Information Science. Singapore: Springer Singapore, 2021: 118-134.

[7] Mansoor A, Anbar M, Bahashwan A, et al. Deep Learning-Based Approach for Detecting DDoS Attack on Software-Defined Networking Controller[J]. *Systems*, 2023, 11(6): 296.

[8] Rahman M R, Hezaveh R M, Williams L. What Are the Attackers Doing Now? Automating Cyberthreat Intelligence Extraction from Text on Pace with the Changing Threat Landscape: A Survey[J]. *ACM Computing Surveys*, 2023, 55(12): 1-36.

[9] Lu Y L, Huang X H, Dai Y Y, et al. Blockchain and Federated Learning for Privacy-Preserved Data Sharing in Industrial IoT[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(6): 4177-4186.

[10] Vevera A V, Cîrnu C E, Radulescu C Z. A Multi-Attribute Approach for Cyber Threat Intelligence Product and Services Selection[J]. *Studies in Informatics and Control*, 2022, 31(1): 13-23.

[11] Gupta M, Akiri C, Aryal K, et al. From ChatGPT to ThreatGPT: Impact of Generative AI in Cybersecurity and Privacy[J]. *IEEE Access*, 2023, 11: 80218-80245.

[12] Schaeffer R, Miranda B, Koyejo S. Are emergent abilities of Large Language Models a mirage?[J]. *Advances in Neural Information Processing Systems*, 2024, 36.

[13] Cannady J, Harrell J. A comparative analysis of current intrusion detection technologies[C]. *Proceedings of the Fourth Technology for Information Security Conference*. 1996, 96.

[14] Ainslie S, Thompson D, Maynard S, et al. Cyber-Threat Intelligence for Security Decision-Making: A Review and Research Agenda for Practice[J]. *Computers & Security*, 2023, 132: 103352.

[15] McMillan R, Pratap K. Market guide for security threat intelligence services. http://www.gartner.com/en/documents/2874317.2014.

[16] FireEye. Response to Increase Your Adversary’s Cost of Operations. http://rvasec.com/slides/2014/Bianco_Pyramid%20of%20Pain.pdf. 2014.

[17] Strom B E, Applebaum A, Miller D P, et al. Mitre att&ck: Design and philosophy[M]. Technical report. The MITRE Corporation, 2018.

[18] Tounsi W, Rais H. A Survey on Technical Threat Intelligence in the Age of Sophisticated Cyber Attacks[J]. *Computers & Security*, 2018, 72: 212-233.

[19] RecordedFuture. The Intelligence Handbook, Fourth Edition. http://go.recordedfuture.com/the-intelligence-handbook-fourth-edition. 2022.

[20] Wei J, Tay Y, Bommasani R, et al. Emergent Abilities of Large Language Models[EB/OL]. 2022: 2206.07682.https://arxiv.org/

abs/2206.07682v2.

[21] GoogleTrends. Large Language Model. http://trends.google.com/trends/explore?date=2018-06-01%202023-12-31&q=Large%20Language%20Model&hl=zh-CN. Jan. 2024.

[22] Web of Science. Large Language Model. http://webofscience.clarivate.cn/wos/alldb/analyze-results/a18df363-8f55-44a7-aac5-a414b00d62b2-c47b6887. Jan. 2024.

[23] Shaaban, Omar. THE ROLE AND IMPACT OF LARGE LANGUAGE MODELS IN CYBERSECURITY: A CASE STUDY ON THE RUSSIAN-UKRAINIAN CONFLICT. http://www.researchgate.net/publication/374807061_THE_ROLE_AND_MPACT_OF_LARGE_LANGUAGE_MODELS_IN_CYBERSECURITY_A_CASE_STUDY_ON_THE_RUSSIAN-UKRAINIAN_CONFLICT. May. 2023.

[24] Touvron H, Lavril T, Izacard G, et al. LLaMA: Open and Efficient Foundation Language Models[EB/OL]. 2023: 2302.13971. https://arxiv.org/abs/2302.13971v1.

[25] Zhao W X, Zhou K, Li J Y, et al. A Survey of Large Language Models[EB/OL]. 2023: 2303.18223. https://arxiv.org/abs/2303.18223v13.

[26] Victor S, Albert W, Colin R, et al. Multitask prompted training enables zero-shot task generalization[C]. *International Conference on Learning Representations*. 2022.

[27] Ouyang L, Wu J, Xu J, et al. Training Language Models to Follow Instructions with Human Feedback[EB/OL]. 2022: 2203.02155. https://arxiv.org/abs/2203.02155v1.

[28] Wei J, Bosma M, Zhao V Y, et al. Finetuned Language Models Are Zero-Shot Learners[EB/OL]. 2021: 2109.01652. https://arxiv.org/abs/2109.01652v5.

[29] Wei J, Wang X Z, Schuurmans D, et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models[EB/OL]. 2022: 2201.11903. https://arxiv.org/abs/2201.11903v6.

[30] Mikolov T, Chen K, Corrado G, et al. Efficient Estimation of Word Representations in Vector Space[EB/OL]. 2013: 1301.3781. https://arxiv.org/abs/1301.3781v3.

[31] Cho K, van Merrienboer B, Gulcehre C, et al. Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation[EB/OL]. 2014: 1406.1078. https://arxiv.org/abs/1406.1078v3.

[32] Devlin J, Chang M W, Lee K, et al. Bert: Pre-training of deep bidirectional transformers for language understanding[C]. *Proceedings of naacL-HLT*. 2019, 1: 2.

[33] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by generative pre-training. http://www.mikecaptain.com/resources/pdf/GPT-1.pdf. 2018.

[34] Yang J F, Jin H Y, Tang R X, et al. Harnessing the Power of LLMS in Practice: A Survey on ChatGPT and beyond[J]. *ACM Transactions on Knowledge Discovery from Data*, 2024, 18(6): 1-32.

[35] Radford A, Wu J, Child R, et al. Language models are unsupervised multitask learners. http://insightcivic.s3.us-east-1.amazonaws.com/language-models.pdf. 2019.

[36] Brown T, Mann B, Ryder N, et al. Language models are few-shot learners[J]. *Advances in neural information processing systems*, 2020, 33: 1877-1901.

[37] Liu Y H, Ott M, Goyal N, et al. RoBERTa: A Robustly Optimized BERT Pretraining Approach[EB/OL]. 2019: 1907.11692. https://arxiv.org/abs/1907.11692v1.

[38] Lan Z Z, Chen M D, Goodman S, et al. ALBERT: A Lite BERT for Self-Supervised Learning of Language Representations[EB/OL]. 2019: 1909.11942. https://arxiv.org/abs/1909.11942v6.

[39] He P C, Liu X D, Gao J F, et al. DeBERTa: Decoding-Enhanced BERT with Disentangled Attention[EB/OL]. 2020: 2006.03654. https://arxiv.org/abs/2006.03654v6.

[40] OpenAI. “GPT-4 Technical Report”. http://cdn.openai.com/papers/gpt-4.pdf. 2023.

[41] Google. Bard. http://bard.google.com/. 2023.

[42] Black S, Biderman S, Hallahan E, et al. GPT-NeoX-20B: An Open-Source Autoregressive Language Model[EB/OL]. 2022: 2204.06745. https://arxiv.org/abs/2204.06745v1.

[43] Almazrouei E, Alobeidli H, Alshamsi A, et al. Falcon-40B: an open large language model with state-of-the-art performance[J]. *Findings of the Association for Computational Linguistics: ACL*, 2023, 2023: 10755-10773.

[44] Peters M, Neumann M, Iyyer M, et al. Deep Contextualized Word Representations[C]. *The 2018 Conference of the North American Chapter ofthe Association for Computational Linguistics*: *Human LanguageTechnologies*, *Volume 1* (*Long Papers*), 2018: 2227-2237.

[45] Song K, Qu L L, Yang M K. Research on Governance Strategy of Generative Artificial Intelligence[J]. *Information and Communications Technology and Policy*, 2023(7): 83-88.
(宋恺, 屈蕾蕾, 杨萌科. 生成式人工智能的治理策略研究[J]. *信息通信技术与政策*, 2023(7): 83-88.)

[46] National Telecommunications and Information Administration. AI Accountability Policy Request for Comment. http://www.ntia.gov/issues/artificial-intelligence/request-for-comments. 2023.

[47] Cui L, Yang L B, He Q L, et al. Survey of Cyber Threat Intelligence Mining Based on Open Source Information Platform[J]. *Journal of Cyber Security*, 2022, 7(1): 1-26.
(崔琳, 杨黎斌, 何清林, 等. 基于开源信息平台的威胁情报挖掘综述[J]. *信息安全学报*, 2022, 7(1): 1-26.)

[48] Liao X J, Yuan K, Wang X F, et al. Acing the IOC Game: Toward Automatic Discovery and Analysis of Open-Source Cyber Threat Intelligence[C]. *The 2016 ACM SIGSAC Conference on Computer and Communications Security*, 2016: 755-766.

[49] Li K, Wen H, Li H, et al. Security OSIF: Toward Automatic Discovery and Analysis of Event Based Cyber Threat Intelligence[C]. *2018 IEEE SmartWorld*, *Ubiquitous Intelligence & Computing*, *Advanced & Trusted Computing*, *Scalable Computing & Communications*, *Cloud & Big Data Computing*, *Internet of People and Smart City Innovation*, 2018: 741-747.

[50] Tang B H, Qiu H R. Indicators of Compromise Automatic Identification Model Based on Cyberthreat Intelligence and Deep Learning[C]. *2022 5th International Conference on Pattern Recognition and Artificial Intelligence*, 2022: 282-287.

[51] OpenAI. TheDFIRReport Assistant. http://chat.openai.com/g/g-lFYMXc3sn.2023.

[52] OpenAI. Threat Intel Briefs. http://chat.openai.com/g/g- 8K32VQvgD-threat-intel-briefs. 2023.

[53] OpenAI. CyberGPT. http://chat.openai.com/g/g-GGqU669bx-cybergpt. 2023.

[54] Mitra S, Neupane S, Chakraborty T, et al. LOCALINTEL: Generating Organizational Threat Intelligence from Global and Local Cyber Knowledge[EB/OL]. 2024: 2401.10036. https://arxiv.org/abs/2401.10036v1.

[55] Bose A, Sundari S G, Behzadan V, et al. Tracing Relevant Twitter Accounts Active in Cyber Threat Intelligence Domain by Exploiting Content and Structure of Twitter Network[C]. *2021 IEEE International Conference on Intelligence and Security Informatics*, 2021: 1-6.

[56] Wang Y, Bashar M A, Chandramohan M, et al. Exploring Topic Models to Discern Cyber Threats on Twitter: A Case Study on Log4Shell[J]. *Intelligent Systems with Applications*, 2023, 20: 200280.

[57] Jin Y, Jang E, Cui J, et al. DarkBERT: A Language Model for the Dark Side of the Internet[C]. *The 61st Annual Meeting of the Association for Computational Linguistics* (*Volume 1*: *Long Papers*), 2023: 7515-7533.

[58] Ali T, Kostakos P. HuntGPT: Integrating Machine Learning-Based Anomaly Detection and Explainable AI with Large Language Models (LLMS)[EB/OL]. 2023: 2309.16021. https://arxiv.org/abs/2309.16021v1.

[59] Kholgh D K, Kostakos P. PAC-GPT: A Novel Approach to Generating Synthetic Network Traffic with GPT-3[J]. *IEEE Access*, 2023, 11: 114936-114951.

[60] Boffa M, Valentim R V, Vassio L, et al. LogPrécis: Unleashing Language Models for Automated Malicious Log Analysis[EB/OL]. 2023: 2307.08309. https://arxiv.org/abs/2307.08309v3.

[61] Fei K, Zhou J, Zhou Y, et al. Laaeb a Comprehensive Log-Text Analysis Based Approach for Insider Threat Detection. http:// papers.ssrn.com/sol3/papers.cfm?abstract_id=4582921. 2023.

[62] Sladić M, Valeros V, Catania C, et al. LLM in the Shell: Generative Honeypots[EB/OL]. 2023: 2309.00155. https://arxiv.org/abs/2309.00155v2.

[63] Siracusano G, Sanvito D, Gonzalez R, et al. Time for aCTIon: Automated Analysis of Cyber Threat Intelligence in the Wild[EB/OL]. 2023: 2307.10214. https://arxiv.org/abs/2307.10214v1.

[64] Bayer M, Frey T, Reuter C. Multi-Level Fine-Tuning, Data Augmentation, and Few-Shot Learning for Specialized Cyber Threat Intelligence[J]. *Computers & Security*, 2023, 134: 103430.

[65] Li V G, Dunn M, Pearce P, et al. Reading the tea leaves: A comparative analysis of threat intelligence[C]. *28th USENIX security symposium,* 2019: 851-867.

[66] Shin H, Shim W, Kim S, et al. #Twiti: Social Listening for Threat Intelligence[C]. *The Web Conference 2021*, 2021: 92-104.

[67] Miao Y K, Bai Y, Chen L, et al. An Empirical Study of NetOps Capability of Pre-Trained Large Language Models[EB/OL]. 2023: 2309.05557. https://arxiv.org/abs/2309.05557v3.

[68] Garza E, Hemberg E, Moskal S, et al. Assessing Large Language Model's knowledge of threat behavior in MITRE ATT&CK. http://ai4cyber-kdd.com/KDD-AISec_files/assessing_llm_qa_2023_kdd_wkshp.pdf. 2023.

[69] Syed Z, Padia A, Finin T, et al. UCO: A unified cybersecurity ontology[J]. *UMBC Student Collection*, 2016: 1-8.

[70] Azevedo R, Medeiros I, Bessani A. PURE: Generating Quality Threat Intelligence by Clustering and Correlating OSINT[C]. *2019 18th IEEE International Conference on Trust*, *Security and Privacy In Computing and Communications/13th IEEE International Conference on Big Data Science and Engineering*, 2019: 483-490.

[71] Chen M H, Tian Y T, Yang M H, et al. Multilingual Knowledge Graph Embeddings for Cross-Lingual Knowledge Alignment[C]. *The Twenty-Sixth International Joint Conference on Artificial Intelligence*, 2017: 1511-1517.

[72] Yao S Y, Yu D, Zhao J, et al. Tree of Thoughts: Deliberate Problem Solving with Large Language Models[EB/OL]. 2023: 2305.10601. https://arxiv.org/abs/2305.10601v2.

[73] Hemberg E, O'Reilly U M. Using a Collated Cybersecurity Dataset for Machine Learning and Artificial Intelligence[EB/OL]. 2021: 2108.02618. https://arxiv.org/abs/2108.02618v1.

[74] Hemberg E, Turner M J, Rutar N, et al. Enhancements to Threat, Vulnerability, and Mitigation Knowledge for Cyber Analytics, Hunting, and Simulations[J]. *Digital Threats*: *Research and Practice*, 2024, 5(1): 1-33.

[75] Cohen R, Hamri M, Geva M, et al. LM vs LM: Detecting Factual Errors *via* Cross Examination[C]. *The 2023 Conference on Empirical Methods in Natural Language Processing*, 2023: 12621-12640.

[76] Ranade P, Piplai A, Joshi A, et al. CyBERT: Contextualized Embeddings for the Cybersecurity Domain[C]. *2021 IEEE International Conference on Big Data*, 2021: 3334-3342.

[77] Aghaei E, Niu X, Shadid W, et al. SecureBERT: A Domain-Specific Language Model for Cybersecurity[M]. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering. Cham: Springer Nature Switzerland, 2023: 39-56.

[78] Alam M T, Bhusal D, Park Y, et al. CyNER: A Python Library for Cybersecurity Named Entity Recognition[EB/OL]. 2022: 2204.05754. https://arxiv.org/abs/2204.05754v1.

[79] Bayer M, Kuehn P, Shanehsaz R, et al. CySecBERT: A Domain-Adapted Language Model for the Cybersecurity Domain[EB/OL]. 2022: 2212.02974. https://arxiv.org/abs/2212.02974v1.

[80] Akbar K A, Halim S M, Hu Y B, et al. Knowledge Mining in Cybersecurity: From Attack to Defense[M]. Lecture Notes in Computer Science. Cham: Springer International Publishing, 2022: 110-122.

[81] Wadhwa S, Amir S, Wallace B C. Revisiting Relation Extraction in the Era of Large Language Models[EB/OL]. 2023: 2305.05003. https://arxiv.org/abs/2305.05003v2.

[82] Alam M T, Bhusal D, Park Y, et al. Looking beyond IoCs: Automatically Extracting Attack Patterns from External CTI[C]. *The 26th International Symposium on Research in Attacks*, *Intrusions and Defenses*, 2023: 92-108.

[83] Ji Z J, Choi E, Gao P. A Knowledge Base Question Answering System for Cyber Threat Knowledge Acquisition[C]. *2022 IEEE 38th International Conference on Data Engineering*, 2022: 3158-3161.

[84] Fayyazi R, Yang S J. On the Uses of Large Language Models to Interpret Ambiguous Cyberattack Descriptions[EB/OL]. 2023: 2306.14062. https://arxiv.org/abs/2306.14062v2.

[85] Das Purba M, Chu B. Extracting Actionable Cyber Threat Intelligence from Twitter Stream[C]. *2023 IEEE International Conference on Intelligence and Security Informatics*, 2023: 1-6.

[86] Grigorescu O, Nica A, Dascalu M, et al. CVE2ATT&CK: BERT-Based Mapping of CVEs to MITRE ATT&CK Techniques[J]. *Algorithms*, 2022, 15(9): 314.

[87] Abdeen B, Al-Shaer E, Singhal A, et al. SMET: Semantic Mapping of CVE to ATT&CK and Its Application to Cybersecurity[M]. Data and Applications Security and Privacy XXXVII. Cham: Springer Nature Switzerland, *20*23: 243-260.

[88] Liu X, Tan Y, Xiao Z H, et al. Not the End of Story: An Evaluation of ChatGPT-Driven Vulnerability Description Mappings[C]. *Findings of the Association for Computational Linguistics*: *ACL 2023*, 2023: 3724-3731.

[89] Das S S, Dutta A, Purohit S, et al. Towards Automatic Mapping of Vulnerabilities to Attack Patterns Using Large Language Models[C]. *2022 IEEE International Symposium on Technologies for Homeland Security*, 2022: 1-7.

[90] Perrina F, Marchiori F, Conti M, et al. AGIR: Automating Cyber Threat Intelligence Reporting with Natural Language Generation[EB/OL]. 2023: 2310.02655. https://arxiv.org/abs/2310.02655v1.

[91] EclecticIQ. ThreatIntelGPT STIX from Chaos. http://www.first.org/resources/papers/amsterdam23/ThreatIntelGPT-Structure-from-Chaos.pdf. Apr. 2023.

[92] McIntosh T, Liu T, Susnjak T, et al. Harnessing GPT-4 for Generation of Cybersecurity GRC Policies: A Focus on Ransomware Attack Mitigation[J]. *Computers & Security*, 2023, 134: 103424.

[93] Li Z Y, Zeng J, Chen Y, et al. AttacKG: Constructing Technique Knowledge Graph from Cyber Threat Intelligence Reports[M]. *Computer Security - ESORICS 2022*. Cham: Springer International Publishing, 2022: 589-609.

[94] Sarhan I, Spruit M. Open-CyKG: An Open Cyber Threat Intelligence Knowledge Graph[J]. *Knowledge-Based Systems*, 2021, 233: 107524.

[95] Gao P, Liu X Y, Choi E, et al. A System for Automated Open-Source Threat Intelligence Gathering and Management[C]. *The 2021 International Conference on Management of Data*, 2021: 2716-2720.

[96] Pan S, Luo L, Wang Y, et al. Unifying Large Language Models and Knowledge Graphs: A Roadmap[EB/OL]. 2023: arXiv preprint arXiv:2306.08302.

[97] Sewak M, Emani V, Naresh A. CRUSH: Cybersecurity Research using Universal LLMs and Semantic Hypernetworks. http://ceur-ws.org/Vol-3532/paper5.pdf. 2023.

[98] Neo4j. Cyberattack Countermeasures Generation With LLMs & Knowledge Graphs. http://neo4j.com/blog/unifying-llm-knowledge-graph/. 2023

[99] Liu J H, Zhan J Y. Constructing Knowledge Graph from Cyber Threat Intelligence Using Large Language Model[C]. *2023 IEEE International Conference on Big Data*, 2023: 516-521.

[100] Das S, Deb N, Cortesi A, et al. Zero-Shot Learning for Named Entity Recognition in Software Specification Documents[C]. *2023 IEEE 31st International Requirements Engineering Conference*, 2023: 100-110.

[101] Feng Z Y, Ma W T, Yu W J, et al. Trends in Integration of Knowledge and Large Language Models: A Survey and Taxonomy of Methods, Benchmarks, and Applications[EB/OL]. 2023: 2311.05876. https://arxiv.org/abs/2311.05876v2.

[102] Noor U, Anwar Z, Malik A W, et al. A Machine Learning Framework for Investigating Data Breaches Based on Semantic Analysis of Adversary's Attack Patterns in Threat Intelligence Repositories[J]. *Future Generation Computer Systems*, 2019, 95: 467-487.

[103] Sapienza A, Bessi A, Damodaran S, et al. Early Warnings of Cyber Threats in Online Discussions[C]. *2017 IEEE International Conference on Data Mining Workshops*, 2017: 667-674.

[104] Mittal S, Das P K, Mulwad V, et al. CyberTwitter: Using Twitter to Generate Alerts for Cybersecurity Threats and Vulnerabilities[C]. *2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2016: 860-867.

[105] Sabottke C, Suciu O, Dumitras T. Vulnerability Disclosure in the Age of Social Media: Exploiting Twitter for Predicting Real-World Exploits[J]. *Proceedings of the 24th USENIX Security Symposium*, 2015: 1041-1056.

[106] Zong S, Ritter A, Mueller G, et al. Analyzing the Perceived Severity of Cybersecurity Threats Reported on Social Media[C]. *The 2019 Conference of the North*, 2019: 1380-1390.

[107] Ghafir I, Hammoudeh M, Prenosil V, et al. Detection of Advanced Persistent Threat Using Machine-Learning Correlation Analysis[J]. *Future Generation Computer Systems*, 2018, 89: 349-359.

[108] Zhao J, Liu X D, Yan Q B, et al. Automatically Predicting Cyber Attack Preference with Attributed Heterogeneous Attention Networks and Transductive Learning[J]. *Computers & Security*, 2021, 102: 102152.

[109] Demırcı D, Şahın N, Şirlancis M, et al. Static Malware Detection Using Stacked BiLSTM and GPT-2[J]. *IEEE Access*, 2022, 10: 58488-58502.

[110] Ferrag M A, Battah A, Tihanyi N, et al. SecureFalcon: Are we there yet in Automated Software Vulnerability Detection with LLMS? [EB/OL]. 2023: 2307.06616. https://arxiv.org/abs/2307.06616v2.

[111] Sun Y, Wu D, Xue Y, et al. When gpt meets program analysis: Towards intelligent detection of smart contract logic vulnerabilities in gptscan[EB/OL]. 2023: arXiv preprint arXiv:2308.03314.

[112] Zhao J Y, Rong Y Y, Guo Y W, et al. Understanding Programs by Exploiting (Fuzzing) Test Cases[EB/OL]. 2023: 2305.13592. https://arxiv.org/abs/2305.13592v2.

[113] Vörös T, Bergeron S P, Berlin K. Web Content Filtering through Knowledge Distillation of Large Language Models[EB/OL]. 2023: 2305.05027. https://arxiv.org/abs/2305.05027v2.

[114] Heiding F, Schneier B, Vishwanath A, et al. Devising and Detecting Phishing: Large Language Models Vs. Smaller Human Models[EB/OL]. 2023: 2308.12287. https://arxiv.org/abs/2308.12287v2.

[115] Ferrag M A, Ndhlovu M, Tihanyi N, et al. Revolutionizing Cyber Threat Detection with Large Language Models: A Privacy- Preserving BERT-Based Lightweight Model for IoT/IIoT Devices [EB/OL]. 2023: 2306.14263. https://arxiv.org/abs/2306.14263v2.

[116] Tran N, Chen H H, Bhuyan J, et al. Data Curation and Quality Evaluation for Machine Learning-Based Cyber Intrusion Detection[J]. *IEEE Access*, 2022, 10: 121900-121923.

[117] Huang K Z, Lian Y F, Feng D G, et al. Method of Cyber Attack Attribution Based on Graph Model[J]. *Journal of Software*, 2022, 33(2): 683-698.
(黄克振, 连一峰, 冯登国, 等. 一种基于图模型的网络攻击溯源方法[J]. *软件学报*, 2022, 33(2): 683-698.)

[118] Xu Y J, Fang Y, Huang C, et al. HGHAN: Hacker Group Identification Based on Heterogeneous Graph Attention Network[J]. *Information Sciences*, 2022, 612: 848-863.

[119] Zang X D, Gong J, Zhang X C, et al. Attack Scenario Reconstruction *via* Fusing Heterogeneous Threat Intelligence[J]. *Computers & Security*, 2023, 133: 103420.

[120] Li K, Zhou H C, Tu Z, et al. AT-GCN: A DDoS Attack Path Tracing System Based on Attack Traceability Knowledge Base and GCN[J]. *Computer Networks*, 2023, 236: 110036.

[121] Scanlon M, Breitinger F, Hargreaves C, et al. ChatGPT for Digital Forensic Investigation: The Good, the Bad, and the Unknown[J]. *Forensic Science International*: *Digital Investigation*, 2023, 46: 301609.

**崔孟娇** 于 2021 年在北京信息科技大学信息安全专业获得学士学位。现在中国科学院信息工程研究所网络空间安全专业攻读博士学位。CCF 会员。研究领域为网络安全、威胁情报。Email: cuimengjiao@iie.an.cn

**姜政伟** 于 2014 年在中国科学院大学获得博士学位。现在任中国科学院信息工程研究所正研级高级工程师, 中国科学院大学网络空间安全学院岗位教授。CCF 会员。研究领域为威胁情报、网络威胁发现与溯源。Email: jiangzhengwei@iie.ac.cn

**陈奕任** 于 2022 年在中南大学计算机学院通信工程系获得学士学位。目前于中国科学院信息工程研究所攻读博士学位。研究领域为网络安全态势感知、网络威胁检测与发现。Email: chenyiren@iie.ac.cn

**江钧** 于 2016 年在北京交通大学电子科学与技术专业获得硕士学位。现任中国科学院信息工程研究所高级工程师。研究领域为网络安全、威胁情报。Email: jiangjun860@iie.ac.cn

**张开** 于 2020 年在北京邮电大学计算机科学与技术专业获得博士学位, 现任中国科学院信息工程研究所工程师, 研究领域为威胁情报、域名安全、网络威胁发现与溯源。Email: zhangkai0216@iie.ac.cn

**凌志婷** 于 2019 年在北京电子科技学院获得硕士学位。现任中国科学院信息工程研究所中级工程师。研究领域为网络安全、威胁情报分析。Email: lingzhiting@iie.ac.cn

**封化民** 于 1984 年获得博士学位。现在任北京电子科技学院教授, 中国科学院大学网络空间学院客座教授。研究领域为网络安全态势感感知、网络空间资产测绘等。Email: Oliver_feng@yeah.net

**杨沛安** 于 2018 年在中国科学院大学获得博士学位。现任中国科学院信息工程研究所高级工程师。研究领域为威胁情报、威胁发现与威胁溯源。Email: yangpeian@iie.ac.cn