# 第一部分

# 准备工作

本书的第一部分包括人工智能驱动的网络安全和威胁情报的介绍，重点介绍了人工智能变体及其潜力（第 1 章）；基本的网络安全知识，包括该领域使用的常用术语、攻击框架和安全生命周期（第 2 章），为本书提供所需的背景知识和主题。

# 第 1 章

# 人工智能驱动的网络安全和威胁情报概述

**摘要** 随着人工智能与网络安全的融合，不断演进的数字威胁防御方式出现了一种新范式。本书探讨了人工智能驱动网络安全和威胁情报的新动态，强调了人工智能计算、分析以及决策能力如何彻底改变网络安全检测、预防和响应。人工智能和机器学习算法可以快速分析大量数据集、识别模式并预测潜在威胁，促使政企组织机构主动加强其数字基础设施。在本书中，对人工智能驱动的网络安全这一主题进行了全面的研究，不仅探讨了人工智能用于网络威胁情报的潜力，还探讨了如何使用不同的人工智能方法（如机器学习建模、深度学习建模、数据科学过程、生成式人工智能建模、具有大型语言建模的自然语言处理等）来提供智能网络安全服务。讨论了多种重要的现实应用领域，例如物联网和智慧城市、工业控制系统和运营技术环境、关键基础设施、网络物理系统、数字孪生和其他相关领域，其中人工智能驱动的网络安全和威胁情报可用于有效和自动化的解决方案。在本书中，还强调了相关的研究问题和挑战，以及它们在基于人工智能的网络安全和威胁情报背景下的潜在解决方案方向。

**关键词** 网络安全；威胁情报；人工智能；可解释人工智能；机器学习；数据科学；智能决策；下一代网络应用

## 1.1 引言

当前网络互联互通和数字化技术发展既创造了惊人的机遇，也带来了网络安全挑战。随着组织、政府和个人比以往任何时候都更加依赖网络与数字技术，威胁形势变得越来越复杂和精密。传统的网络安全技术已不足以应对这场高风险的“猫捉老鼠游戏”，网络攻击者不断使用创造性的方法来突破防御。因此，人工智能驱动的网络安全和威胁情报——一种利用不同人工智能技术的计算和分析能力的前沿解决方案，已成为一股革命性力量，改变了传统网络安全和威胁情报的处理

方式。

人工智能驱动的网络安全的基础在于它能够从历史数据中学习，即机器学习[1]，并不断完善对网络、系统和应用程序中正常和恶意行为的理解。人工智能已经展示了其在网络安全领域的潜力，因为它能够处理大量数据、识别趋势并调整响应。传统的安全方法虽然在某些情况下仍然有效，但往往落后于网络犯罪分子不断发展的策略。人工智能驱动的网络安全通过提供适应性强的主动防御方法填补了这一空白。此外，由人工智能驱动的威胁情报将网络安全的能力扩展到预防措施之外。人工智能系统可以通过分析来自各种来源的数据（包括暗网论坛、社交媒体和其他在线平台）来发现新的威胁和漏洞，甚至预测未来的攻击媒介。这种预测能力使政企组织能够主动加强防御，在漏洞成为问题之前修复漏洞，并采取强有力的策略来减轻潜在风险。

人工智能驱动的网络安全有望彻底改变我们保护数字资产和信息的方式。它结合了复杂的机器学习算法、深度学习、高级数据分析、自然语言处理和自动化，构建了一个动态自适应的防御系统[2]。人工智能驱动的系统可以实时学习和适应，领先网络威胁一步，这与主要基于预定规则和签名的传统网络安全技术不同。人工智能系统可以使用机器学习算法和深度神经网络实时检测异常和潜在威胁，使安全团队能够快速有效地做出反应。人工智能为安全专业人员提供了领先网络对手一步所需的能力，从检测复杂的恶意软件到识别未经授权的访问尝试。机器学习、深度学习、自然语言处理和其他人工智能方法[2]的强大功能被用于人工智能驱动的网络安全和威胁情报，不仅可以检测和缓解攻击，还可以在攻击造成损害之前预测和预防攻击。

在当前对人工智能驱动的网络安全和威胁情报的探索中，我们将深入研究正在重塑我们保护数字环境方式的前沿应用和技术。将研究人工智能如何增强威胁检测、自动化事件响应，并提供对新型威胁的分析研判，帮助政企组织获得战略优势。在面对不断变化的网络安全格局时，了解人工智能在防范网络威胁方面的作用并利用其潜力提高我们的数字系统安全弹性至关重要。我们将进一步研究人工智能驱动的安全解决方案所带来的问题和挑战，并为实现创新与负责任使用之间的平衡而不断做出努力。本书引导读者进入人工智能驱动的网络安全和威胁情报世界的旅程，旨在阐明人工智能的革命性前景及其对如何更好地保护数字领域的深远影响。

总体而言，人工智能驱动的网络安全解决方案提供了一条充满希望的前进道路，使我们能够防御各种不断发展和更先进的网络威胁。本书旨在介绍人工智能驱动的网络安全和威胁情报的多种方法，包括机器学习和数据科学建模以及现实世界的应用。因此，本书让读者一睹这一新兴研究领域具有革命性的更多可能性。随着我们深入研究这一领域，还将探索人工智能在现实世界中的广泛应用，它所带来的新的困难与挑战，以及随着研究更加深入而不得不面对的伦理问题。

## 1.2　网络安全与威胁情报

本节旨在从多个维度剖析网络安全与威胁情报的定义，并探讨二者之间的关系。

### 1.2.1　什么是网络安全

在过去的半个世纪里，我们的现代数字文明与信息和通信技术（ICT）的联系更加紧密。随着智能计算机和系统通过互联网实现全球互联，数据泄露和网络攻击的威胁也日趋严重。因此，信息通信技术安全——检测与防御 ICT 系统免受各类先进网络攻击或威胁的能力，近年来已成为安全专业人士和决策者们的首要关切[2-3]。

企业为确保其数据和系统的机密性、完整性和可用性，纷纷实施了一系列保障措施、政策和流程，这些均基于信息通信技术安全的原则，通常被概括为“CIA 三要素”（confidentiality，integrity，availability）。简而言之，网络安全即是保护那些因利用信息与通信技术而容易遭受攻击的资源的过程。

网络安全是一个具有多重含义的术语，在当今社会被广泛提及。诸如“信息安全”“数据安全”“网络安全”“互联网或物联网安全”[4]等关键术语常常相互交织，给该领域的读者和专家带来了一定的困惑。然而，在这些术语中，“网络安全”一词在全球范围内受到的关注度最高，且其重要性仍在持续增长[5]。

对于网络安全的定义，不同研究者给出了多种表述。例如，网络安全被定义为一系列旨在保护信息与通信技术系统免受威胁或攻击的活动或政策[6]。Craigen 等[7]则将网络安全视为一套工具、实践和指南的集合，这些工具、实践和指南共同作用于保护计算机网络、软件程序和数据，使之免受攻击、损坏及未授权访问。

根据 Aftergood 等[8]的说法，“网络安全是一套旨在保护计算机、网络、程序和数据免受攻击和未经授权的访问、更改或破坏的技术和流程”。因此，网络安全不仅涵盖了对各类网络攻击或威胁的精准识别，还涉及针对这些威胁的防御策略的制定与实施，最终确保系统的稳固安全。这一过程与数据的机密性、完整性和可用性息息相关，它们是网络安全不可或缺的三个支柱。

正如之前所述，机密性、完整性和可用性的 CIA 三要素构成了推动企业信息安全政策的核心框架。任何对这些原则或其组合的破坏，都将被视为对信息安全的严重威胁。如图 1.1 所示，这些网络威胁在现实中常常表现为“数据窃取”“数据篡改”“拒绝访问数据”等情形。

基于上述安全政策的 CIA 原则，可以进一步阐释：机密性确保了数据、对象和资源免受未经授权的访问和滥用；完整性则保证了数据在未经授权的情况下不被更改；而可用性让授权用户或合适的实体能够随时访问所需的系统和资源。

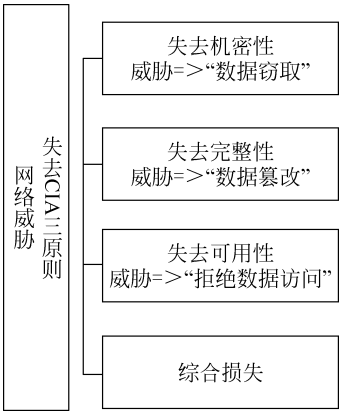


**图 1.1 Sarker 等采用的关于失去 CIA(机密性、完整性和可用性)三原则的网络威胁的说明,用于推动企业内的信息安全政策**

综上所述,网络安全可以被定义为一个全面性的防护体系,旨在保护计算机系统、网络和数字信息不受任何形式的未经授权访问、攻击、损坏或盗窃。它依赖于技术、流程和实践的有机结合,以有效防范网络威胁,并持续维护数据的机密性、完整性和可用性。

## 1.2.2 什么是威胁情报

网络威胁情报(CTI)是网络安全的重要组成部分,它赋予组织主动防御的能力,有效保护其系统和数据免受潜在入侵的威胁。它涉及收集、分析和传播有关数字领域网络安全威胁和漏洞的数据。如图 1.2 所示,威胁情报的生命周期模型凸显了质量和运营管理的重要性。

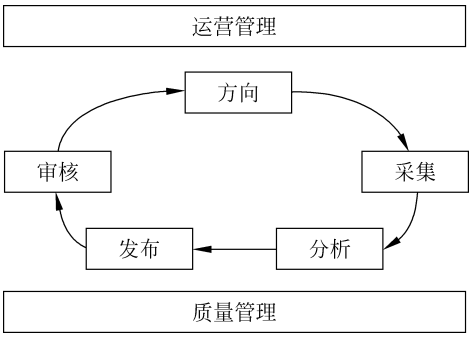


**图 1.2 威胁情报生命周期模型示意图**

CTI 的主要目标是向组织提供有关网络威胁的全方位背景信息和实用的操作建议,从而助力组织做出明智的决策,更有效地抵御网络攻击。这一目标的实现需要依赖于资深的分析师团队、尖端的技术工具和持续追踪不断变化的威胁态势的

坚定决心。

CTI 可以定义为："基于知识、技能和经验的信息，涉及网络和物理威胁以及威胁行为者的发生和评估，旨在帮助减轻网络空间中可能发生的潜在攻击和有害事件的影响"[9-10]。其信息来源丰富多样，包括但不限于开源情报、社交媒体情报、人力情报、技术情报、设备日志文件、从互联网交易中获取的合规法律数据或情报，以及来自暗网(Dark Web)的敏感数据。

近年来，随着网络威胁的复杂性和频繁性不断增加，威胁情报在企业网络安全规划中的地位愈发重要。它使得企业能够主动识别那些对其业务运营构成最大风险的攻击，从而采取更为积极的防御措施。尽管传统的风险管理方法在处理已知风险时表现不俗，但网络威胁情报在各类风险场景中都显示出了其独特的价值，尤其是在传统管理方法显得力不从心的情境下，如图 1.3 所示。这些情境的定义如下。

- 我们知道的：这通常表明我们熟悉的特定威胁行为者正在攻击我们的公司。
- 我们知道我们不知道的：我们似乎很容易受到特定形式的威胁，但我们需要使用威胁情报来进一步发现。
- 我们不知道我们不知道的：在某种情况下，我们面临风险，但除非它发生，或者除非威胁情报提醒我们，否则我们不会知道。

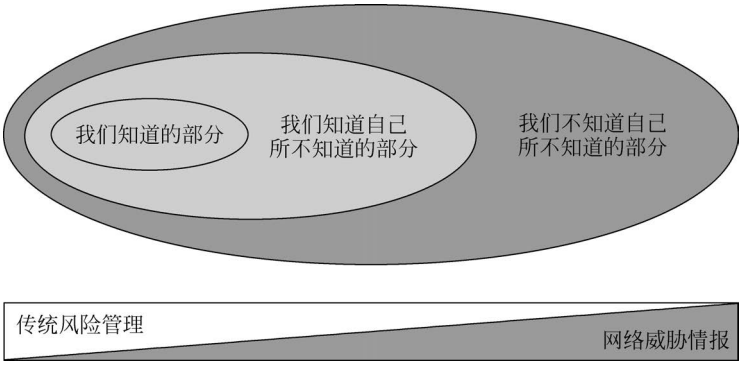


**图 1.3　三层知识分类示意图**[9]

总体而言，威胁情报在揭示威胁行为者的目标、动机和攻击手段上起到了至关重要的作用，它是现代网络安全实践中不可或缺的一环。如图 1.2 所示，威胁情报的生命周期模型展示了一种方法，该方法能够将原始数据转化为有价值的情报，进而支持我们做出明智的决策和迅速的行动。因此，威胁情报不仅让我们能够转变对威胁的应对方式，从被动防御转为主动出击，更使我们能够依据实时数据快速、准确地做出安全决策。一个有效的 CTI 计划能够显著提升组织应对网络威胁的能力，减轻潜在事件的影响，并为网络安全投资提供科学的指导。

基于上述讨论可以清晰地看到，网络安全是一门博大精深的学科，它涵盖了保护数字资产免受网络威胁的多种战略、技术和实践。而威胁情报作为网络安全中的一个重要分支，专注于收集和分析有关潜在威胁的详尽信息，以此提升组织的自我防护能力。因此，网络威胁情报为组织提供了宝贵的背景信息和前瞻意识，使组织能够在新兴威胁面前保持领先地位。为了在不断变化的网络环境中调整防御策略并有效应对各种威胁，一个组织的网络安全体系必须依赖于及时、准确的威胁情报。

## 1.3 了解网络安全中的人工智能

随着每天新发布的恶意软件数量激增，威胁传播的速度呈现指数级增长。单纯依赖人工分析来全面、有效地应对当今海量的网络威胁，实际上已经变得不再可行。在网络安全行业中，人工智能正发挥着引领革新的重要作用，它使得识别、预防和应对在线威胁变得更加便捷和智能化。通过 AI 技术，能够更有效地应对网络安全挑战，保障网络环境的稳定与安全。这正是本书所关注的焦点。接下来，将在网络安全的背景下深入探讨 AI 的关键潜力及其多样化的应用范畴。

### 1.3.1 人工智能的潜力

人工智能的知识正在全球范围内引发网络安全行业的颠覆性变革。凭借其在特定问题领域中展现出的智能决策和自动化计算能力，AI 已成为推动智能服务发展的关键力量[5]。然而，传统的安全解决方案，如防病毒软件、防火墙、用户身份验证和加密技术等，可能已无法完全满足当今复杂多变的网络安全需求[2]。传统系统的主要局限性在于，它们通常依赖于少数具备深厚专业知识的安全人员来维护，且数据处理往往基于临时性方案进行。这种模式极大地限制了系统智能、自动地适应和响应当今复杂多变的安全威胁的能力[2,11]。因此，AI 驱动的网络安全的能力与特性主要由其三大核心方面——“自动化”“智能化”“健壮性”来定义。具体如下：

- **网络安全的自动化**：AI 驱动的网络安全系统凭借其无须人工干预即可自动执行任务和流程的能力，展现了自动化的核心价值。通过运用先进的算法、脚本和机器学习模型，这些系统极大地简化了诸如威胁检测、事件响应和常规安全操作等重复且耗时的任务。自动化的应用使得系统能够迅速响应安全事件，并执行预定义的操作，从而显著提升运营效率。它优化了安全操作的流程，加快了响应速度，并减轻了网络安全团队的工作负担。因此，这种自动化能力有助于组织机构更加高效地处理庞大的数据量和安全警报，进而缩短对潜在危险威胁的响应时间。
- **网络安全的智能化**：智能系统具备卓越的数据分析能力、上下文理解能力，

并能基于复杂的模式和深刻见解作出精准决策。在网络安全领域，智能解决方案凭借其敏锐的洞察力，能够迅速识别并应对那些未被明确编程的威胁模式、异常行为，从而有效防范新颖且复杂的网络攻击。智能网络安全系统的构建离不开高级分析技术的支持、机器学习算法的助力以及威胁情报的精准指引。这些先进的技术和信息不仅使系统能够灵活适应新数据，还能从中不断学习，持续优化自身的性能。这种强大的适应性和提供可执行见解的能力，对于在日益复杂多变的网络威胁环境中保持领先地位具有无可替代的价值。

- **网络安全的健壮性**：一个健壮的、AI 驱动的网络安全系统在各种复杂条件下都能保持其卓越的有效性，并有效抵御任何形式的降级或故障，包括对手企图进行的操纵和利用。其核心在于精心设计的 AI 模型和算法，这些算法能够灵活适应各种场景，并在长期运行中持续展现其高效性。因此，当面临敌对攻击、数据噪声或污染以及不断演变的威胁时，这样的系统能够从容应对，展现出强大的处理能力。更重要的是，它具备出色的弹性，即使在遭遇意外情况或有人故意尝试欺骗时，也能提供可靠且准确的安全结果，确保网络环境的稳固与安全。

简而言之，自动化着重于无须人工干预即可自动执行任务，智能化则体现在对复杂数据模式的深入理解和灵活适应，而健壮性则确保系统在各种挑战面前保持坚韧和高效。人工智能驱动的网络安全解决方案巧妙地将这些元素融为一体，构建了一个强大的防护体系，用以抵御层出不穷的网络威胁。

### 1.3.2　人工智能的分类

AI 模型可以大致划分为三大主要类型：生成式、判别式和混合式 AI 模型。在网络安全领域中，每种模型都发挥着独特的作用，并具备不同的应用场景，具体描述如下：

- **生成式 AI**：生成式模型极具灵活性，其核心功能在于生成全新的、此前未曾出现的数据样本，这些数据通常以图像、文本或其他媒体形式呈现。通过向数据集注入合成数据并提供更为多样化的训练数据，生成式模型能够显著提升机器学习模型的效果。它们专注于模拟数据的潜在概率分布，因此在生成合成训练和测试数据集方面表现出色。例如，生成式 AI 能够生成移动恶意软件样本，助力防御模型的构建。安全研究人员还能借助其强大的能力，运用各种恶意软件类型来评估安全措施的有效性。
- **判别式 AI**：判别式 AI 模型专注于在数据中识别并区分不同的类别或分类。由于其在分类性能上的卓越表现以及区分各种数据类别的能力，判别式模型在网络安全领域得到了广泛的应用。它们能够识别与恶意代码或行为相关的特定模式和特征，从而有效识别恶意软件。这些模型经常用于

入侵检测系统，特别是用于将网络数据或用户行为分类为恶意或良性。基于输入的数据，它们能够做出二元或多类别的精准决策。

- **混合式 AI**：为了最大限度地发挥每种策略的优势，混合式 AI 巧妙地将上述生成式和判别式元素融为一体。其目标在于提供一个既强大又灵活的解决方案。在识别对抗性攻击时，混合式 AI 模型能够借助生成式组件生成人工攻击样本，这些样本随后被用于训练判别式模型。通过这种方式，混合式 AI 模型在网络安全领域同时兼顾了生成式和判别式的需求，从而提供了一个更为全面和均衡的解决方案。它们具备出色的适应性，能够灵活应对多种不同的安全挑战。

总之，特定的用例和需求是决定是否在网络安全中采用生成式、判别式或混合式人工智能的关键因素。面对异常检测、对抗性防御等复杂挑战时，混合式人工智能结合了生成式和判别式人工智能的优势。特别是生成式人工智能能够根据需求生成合成数据和恶意软件变种，这些技术使组织机构能够构建出既高效又适应多变威胁环境的网络安全解决方案。此外，基于计算和数据的本质，人工智能还有其他多种分类方式[2]。接下来，将探讨这些分类及其潜在的应用场景。

- **分析型 AI**：此类人工智能专注于从数据中提炼出观点和模式，以提供决策建议，从而助力数据驱动的决策过程。分析型 AI 能够分析海量的数据，包括网络流量、日志文件、用户活动以及历史攻击趋势，旨在发现异常和潜在的威胁。它能够与入侵检测系统（IDS）、安全信息和事件管理（SIEM）工具以及预测分析技术无缝结合，实现对网络威胁的实时识别与应对。在网络安全领域，分析型 AI 的潜在应用场景广泛，包括威胁预测、异常检测、零日漏洞的发现以及通过数据驱动的洞察来优化事件响应策略。
- **功能型 AI**：它与分析型 AI 类似，但不同之处在于它基于从数据中获取的见解和知识来执行具体操作，而非仅仅提出建议。功能型 AI 专注于执行预定的任务或工作流程，无须人工干预。它能够在多种自动化任务中承担专门的角色，如导航、对象识别或工业自动化。在安全领域，功能型 AI 能够自主执行特定的安全任务，如自动化的漏洞评估、补丁管理以及隔离受感染计算机的网段。为了加快人为干预和响应时间，它还可以自动化执行常规的安全任务，如漏洞评估、合规性检查和安全策略执行。
- **交互式 AI**：交互式 AI 系统的设计旨在以人类认为自然的方式与人们互动，无论是通过语音、手势还是文本。它们极大地促进了双方间的交流与互动。在网络安全领域，交互式 AI 发挥着至关重要的作用，推动了人机之间的无缝协作。例如，虚拟助手能够辅助安全分析师处理事务，解答安全问题，并在应急响应过程中提供宝贵指导。在网络安全实践中，交互式 AI 的应用广泛，包括用于事件报告的聊天机器人、协助安全策略查询的虚拟助手，以及提供实时安全监控的交互式仪表板等。