

【实践研究】

# 人工智能技术赋能情报工作的 历程与当前思考

李广建 潘佳立

**【摘要】**本文系统探讨了人工智能技术的演进及其在情报工作中的应用,揭示了人工智能赋能情报工作在思想、特征、技术方法和场景应用上的新变化。从基于规则的范式到基于基础模型的阶段,人工智能技术为情报工作的智能化提供了新机遇,但同时也引发了伦理、法律、隐私等新挑战。本文指出了情报领域在新时期应该关注的问题和挑战,并提出了针对这些问题和挑战的对策,以为情报领域的稳定发展提供有益的建议。

**【关键词】**人工智能技术;基础模型;数智赋能;情报工作;应用挑战

**【作者简介】**李广建,博士,北京大学信息管理系教授,研究方向为信息管理技术、大数据情报分析;潘佳立(通讯作者),博士生,北京大学信息管理系,研究方向为信息系统、数据挖掘,E-mail: panjiali1998@stu.pku.edu.cn(北京 100871)。

**【原文出处】**《信息资源管理学报》(武汉),2024.2.4~20

**【基金项目】**本文系国家自然科学基金重大项目“数智转型背景下智能情报关键技术应用研究”(23&ZD228)的研究成果之一。

## 1 引言

人工智能技术的飞速发展,引发了包括情报领域在内的各个领域对人工智能技术应用的热情和深刻思考。人工智能新时代的情报工作在思想、方法、技术和场景应用上都迎来了新变化。特别是近年来,诸如GPT等基础大模型技术的崛起,使人工智能技术在全球范围内得到广泛关注和快速应用,为各行业注入了新的活力。在情报领域,通过高效整合大数据、深度学习、知识计算、自然语言处理等技术,可以更为迅速、准确地获取和解析庞大的数据和信息流,为决策者提供了更为精准、深入的情报支持。国内外情报界的理论和实践都表明,人工智能技术的应用,正在重塑情报工作的面貌与格局,为当前和未来的情报工作带来了重大发展机遇。

然而,伴随人工智能技术的广泛应用,情报工作也面临着新的、巨大的挑战。人工智能技术的强大计算和分析能力固然使得情报处理工作更加高效,但也引发了一系列问题。例如,不透明的算法和潜在的被滥用风险是情报工作“准、快、精、全”的大敌。同样,如何提升模型输出结果的可信性也

是新时期情报工作的棘手问题。此外,新旧情报工作范式的切换与融合方式目前仍处于初步的探索阶段<sup>[1]</sup>。最重要的是,在人工智能赋能情报工作的过程中,情报工作者需要不断适应高度技术化和动态化的环境,这既需要深厚的专业知识,也需要具备跨学科的思维与与新技术协同工作的能力,这对情报工作的人才提出了新的要求。

为深刻理解人工智能赋能情报工作的机遇与挑战,本文旨在系统梳理人工智能的发展及其在情报工作中的应用,以期发现情报领域在新时期应该关注的问题和挑战,并提出针对这些问题和挑战的解决方案,为情报领域的稳健发展提供有价值的建议。

## 2 人工智能技术发展及其对情报工作的赋能

人工智能技术从诞生到发展繁荣,其性能不断优化,依托的底层技术原理也在不断演化,人工智能的定位也从最开始的代替人类完成低端重复性工作逐渐转为代替人类进行高端思考,从代替人类解决特定问题的人工智能逐渐转化为具有自主适应新环境、拥有高效的学习和泛化能力、能够完成多种任务或者所有任务的人工智能。目前根据研

究的目的,可以对人工智能从不同的角度进行分类。例如,从能力上分,可以将人工智能分成弱人工智能、强人工智能和超级人工智能三种类型<sup>[2]</sup>,其中,弱人工智能是解决人类特定任务的人工智能,强人工智能是指能够解决多领域问题的人工智能,超级人工智能则是强人工智能的一种更高级别的形式,是一种超越人类智慧的人工智能。从功能上分,可以将人工智能分成只能使用当前数据的反应型人工智能、可以利用最近收集的数据做出决定的有限记忆型人工智能、能够理解人类心理状态的心智型人工智能以及具备自我意识且能够感知和预测人类情感的自我意识型人工智能。从具体应用领域和主要技术方案分,人工智能可以分为机器学习、深度学习、计算机视觉、自然语言处理、机器人技术、专家系统等。本文参考已有的人工智能分类,并结合其在情报工作中发挥的作用,将人工智能技术的发展历程划分为基于规则的人工智能、基于机器学习的人工智能、基于深度学习的人工智能以及基于基础模型的人工智能四个阶段,如图1所示。同时,以这四个阶段为依据,对人工智能赋能情报工作的范例进行分析,以揭示不同时期人工智能对情报工作赋能的特征、方式与影响。

值得注意的是,这四个阶段的划分并非是完全物理隔离的,所做的划分仅仅是为了揭示在特定历史时期(阶段)中主流人工智能技术的特点,并不意味着其他阶段的人工智能技术的销声匿迹。同样,在情报领域应用人工智能技术时,也存在着同时应用不同阶段的人工智能技术的情况。

## 2.1 基于规则的人工智能

### (1) 主要技术

1950年,艾伦·麦席森·图灵提出了“图灵测试”,引发了人们对于机器智能的思考和探索。随后,在1956年的达特茅斯会议上,人工智能(Artificial Intelligence)这一术语被正式采用,这标志着人工智

能学科的诞生。这次会议作为人类历史上第一次召开的人工智能研讨会,奠定了人工智能学科的基础,同时也开启了基于规则的人工智能时代。本文将20世纪50年代至80年代划分为基于规则的人工智能阶段,在这一阶段,关于人工智能的研究主要集中在符号逻辑、专家系统和知识表示方面。受限于计算机性能和可用数据,研究者主要通过基于逻辑和规则的知识表示来实现智能推理。具体而言,这个阶段人工智能技术的主要特征在于对知识的形式化表示,通过人工编写特定函数规则和代码,将人类知识以形式化的规则和概念的方式进行编码,以便实现智能推理,使计算机能够基于先前定义的规则来进行决策和解决问题。

这一阶段的人工智能技术最显著的优势之一是其易于理解和可解释。由于这一阶段的人工智能依赖于手动编写规则和知识,因此具备出色的可解释性,使人能够直观地理解系统的决策过程。这一特性使得基于规则的人工智能技术有了广泛的应用,尤其是在诸如医疗卫生、法律法规等对透明性和解释性要求较高的领域,受到了人们的欢迎,典型的成果就是专家系统。

然而,基于规则的人工智能技术也存在一些明显的劣势。首先,手工编写规则和构造知识库的代价相当高昂,这使得系统的建设烦琐且耗时。其次,编写出的规则通常不易扩展,要想适应新的情境或任务,就需要进行大规模修改和重构已有规则,人工智能系统的灵活性和适应性明显不足。最为重要的是,基于规则的人工智能难以有效处理模糊、不确定和复杂的问题,因而在应对现实世界决策的复杂性方面受到限制。

总体而言,基于规则的技术奠定了人工智能的基础,其强调知识的形式化表示,为后续人工智能的发展提供了宝贵经验,但随着数据量的增加和任务复杂性的提升,基于规则的人工智能的局限性愈

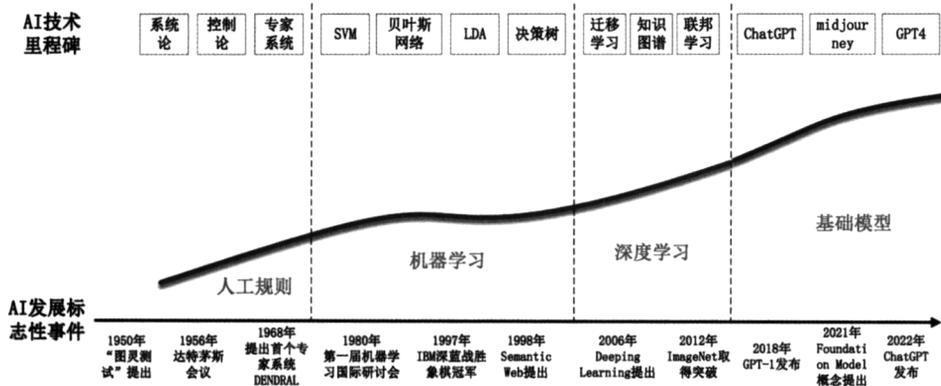


图1 人工智能发展历程与阶段划分

发明显。为了解决这些问题,研究人员开始探索基于统计和概率的方法,如机器学习等。人工智能领域的技术路线逐渐演变,突破了基于规则时代的限制,朝着更加灵活、智能的方向发展。

## (2) 赋能情报工作范例

基于规则阶段人工智能技术主要通过专家系统对情报工作进行赋能。1968年爱德华·费根鲍姆及同事提出并开发第一个专家系统 DENDRAL,将“知识库”和“推理机”结合起来解决问题,这种模式对包括情报工作在内的众多领域产生了巨大影响。DENDRAL 专家系统通过知识库存储有机化学家的专业知识,包括有机分子结构、化学反应规则以及质谱数据的解释,这些知识的表示形式以规则和关系的方式呈现,为系统提供了有机化学领域专业知识的计算机可处理版本。进而,通过推理引擎实现对知识库的应用,模拟有机化学家的推理过程,根据输入的质谱数据,逐步应用知识库中的规则进行推理,从而推断出分子结构。

此后,1972年,斯坦福大学研制了 MYCIN 医疗专家系统,该系统在对血液感染性疾病的诊断和治疗方案选择中取得了显著成就。首先,MYCIN 系统的成功核心在于其丰富的知识库,其知识库整合了大量的领域专业知识,收录了 200 余条规则。其次,MYCIN 采用似然推理技术使系统能够通过对不同症状和疾病之间的关系进行概率推断,这种推理方式在处理不确定性信息上具有优势,此外,该系统的可理解性和可修改性,使医生能够理解系统的推理过程,并可以根据新的医疗知识对系统进行修改和扩充。

同时,在这一阶段,传统的科技情报领域也已经意识到了将人工智能技术应用于情报检索工作具有巨大的前景。尤其是美国的 DIALOG、ORBIT、BRS 以及欧洲的 ESA 等大型联机检索系统的普及,为基于专家系统的智能信息检索带来了良好的发展契机和需求牵引。传统的联机情报检索系统要求检索人员要具备数据库、检索语言、检索系统功能与指令等专业的检索技术知识,这使得许多缺乏经验的非专业用户很难通过自行与这些联机检索系统交互来满足其情报需求,在这一背景下,信息检索领域的专家系统初露头角,并试图解决信息检索门槛高这一问题。关于在信息检索领域开发并利用专家系统的机理,Shoval<sup>[3]</sup>首先从理论层次指出该类系统主要由语义网络形式的知识库和运行于知识库的程序规则组成,系统与用户的交互过程也可以被抽象为“用户检索”与“系统反馈”两个阶

段。随后,陈光祚<sup>[4]</sup>从七个方面进行了更详细的总结与分析,包括情报需求、数据库、检索词、检索语言知识、检索策略、通信系统协议以及系统容错与反馈能力。其中,针对检索策略的优化,提出了基于检索经验丰富的用户的检索记录建立知识库,例如,通过捕获“或”逻辑连接的检索词建立控制词表,供新手用户调整检索式时进行参考。与此同时,相关理论也逐渐向应用落地,黄祥喜等<sup>[5]</sup>提出了一种经典的智能情报检索系统的结构,包括智能接口、知识库和文本处理三个部分,并认为相比于传统的情报检索系统,融入人工技术的智能情报检索系统建立在知识库的基础上,且允许用户利用自然语言进行检索,具备更强的学习和自适应能力。随后,基于前述理念、机制与结构,汪红秋<sup>[6]</sup>提出了一种名为 ESFFIR 的反馈情报检索专家系统,该系统具备知识表达和推理结构,可以使缺乏检索经验的用户在专家系统的协助下进行成熟的检索操作。

DENDRAL 系统在当时体现了将领域专家的知识转化为计算机可处理形式的创新思维。其知识库的建构、推理引擎的设计以及特征提取的实现,展示了专家系统在解决复杂专门领域问题上的实用性。MYCIN 系统在实际医疗中的成功应用,源于其对专家系统特征和方式的深刻理解,并验证了在专家系统中引入似然推理和自然语言交互等技术是可行的。信息检索作为情报工作的核心业务,从理论到应用层面,不断理解并融入知识库、规则等具有人工智能色彩的技术要素与组件,并通过构建专家系统提升用户的学术情报、科技情报检索能力,深刻体现了该阶段人工智能技术对情报工作赋能的学术价值与科技价值。

## 2.2 基于机器学习的人工智能

### (1) 主要技术

20 世纪 80 年代,人工智能进入了应用发展的新高潮。在这个阶段,机器学习,特别是神经网络等技术,日益被重视起来并得以大量应用,逐渐成为人工智能的主要研究方向。具体地说,机器学习时代人工智能技术的主要特征在于其强调通过数据自动学习知识,从而避免了手工编写规则的烦琐过程。这一阶段涌现出了一系列代表性技术,包括决策树、支持向量机、随机森林和集成学习等。这些技术使得计算机能够从大量数据中学习并自动调整模型,以解决相对复杂的问题。机器学习的繁荣推动了人工智能技术的不断进步和应用领域的扩展。

机器学习的优势在于其能够自动从数据中学

习知识,极大地降低了手工编写规则的难度。更重要的是,机器学习算法具备较强的泛化能力,这意味着它们能够将从一部分数据中学习到的知识应用于新的、未见过的数据,从而在不同的任务和情境中表现出良好的适应性。这种泛化能力使得机器学习算法在面对具有复杂性和不确定性的问题时具有更高的鲁棒性,为其在众多领域中的广泛应用提供了可能。

虽然机器学习技术取得了显著的进展,但它也有自身的一些局限性。在处理大规模数据和复杂任务时,传统的机器学习算法可能会遇到效率瓶颈,导致训练和优化过程变得耗时且需要大量的计算资源。此外,机器学习算法在处理结构化数据方面表现出色,但对于如图像和文本等非结构化数据,其处理能力相对有限,往往需要借助专家参与,利用特征工程等方法来人工提取和构建有效的数据特征,以提升模型的性能。这种对特征工程的依赖增加了模型开发的复杂性和主观性,同时也限制了机器学习算法在多样化和复杂场景下的应用。因此,如何有效地处理大规模和非结构化数据,以及减少对特征工程的依赖,仍是机器学习领域亟待解决的问题。

总体而言,机器学习的发展为人工智能领域带来了革命性的变革,克服了基于规则时代专家系统依赖于烦琐规则的弊端,为实现智能系统更广泛的应用奠定了基础。同时,该阶段的研究人员也对机器学习算法的不足之处进行了深入思考,并进一步研究了改进的策略。

## (2) 赋能情报工作范例

机器学习阶段人工智能技术赋能情报工作的方式以指标数据预测为主流,赋能路径为不断开发、适用和优化各种分类、聚类机器学习算法。

以分类算法为例,1995年Cortes等<sup>[7]</sup>开发了支持向量机(SVM)算法,其在小样本、非线性和高维模式识别方面的优势为情报工作提供了独特的赋能,因此被广泛应用于各类情报任务。一方面,在信息检索任务中,SVM通过处理高维稀疏的文本特征,显著提高了自动分类和检索的效率。例如,Joachims<sup>[8]</sup>最早基于SVM在文本数据上进行了实验,发现其具有非常强的稳健性,且无须特征选择;同一时期,伴随互联网的迅猛发展,网页资源作为一种情报源受到关注,Dumais等<sup>[9]</sup>最早考虑了SVM在网页资源这种大型异构数据集上的适应性问题,紧接着李晓黎<sup>[10]</sup>结合网页数据在不同处理阶段的特点对SVM进行针对性优化;后来,Tian等<sup>[11]</sup>将注

意力聚焦于文档检索,先通过特征工程提取文档集合中的语义结构,进而再利用SVM判断给定查询与文档之间的相关性。另一方面,在情报工作更为核心的科技情报分析任务中,SVM在专利分类和分析方面展现了其在情报工作中的应用价值。通过结合专利的文本描述、关键词等信息进行自动分类,SVM使科技情报分析人员能够更快速地理解和归类专利信息。例如,Fall等<sup>[12]</sup>最早通过构建一个专利数据集,测试了包括SVM在内的多种模型对专利文本的分类效果,随后,李程雄等<sup>[13]</sup>也将SVM应用于专利文本分析场景,解决了专利申请量增加背景下专利分类效率低下的问题,同时其也尝试将KNN与SVM进行结合,以优化单个模型在该问题上的性能;SVM在该类情报任务中有着持久的影响,近年来依旧有学者将其应用于专利分类与质量分析,以减少专利审查的时间成本<sup>[14]</sup>。

除了经典的分类算法外,基于机器学习的聚类算法在情报工作中也得到了广泛应用。2003年,Blei等<sup>[15]</sup>提出了LDA模型,LDA模型是一种无监督的机器学习方法,属于文本挖掘领域中的主题模型,目标是通过对文档集的统计分析,发现文档中隐藏的主题,并推测文档主题词的分布情况。LDA模型被众多情报工作人员用于研究文献主题演化、文献个性化推荐等场景或任务中。在文献主题演化分析任务中,通过主题聚类可以分析某领域研究发展脉络和热点的变化,也可以为图书采购和馆藏发展提供参考。例如,AlSumait等<sup>[16]</sup>最早对该模型进行在线化设计,用于动态跟踪文本流的主题随时间的演化并实时检测新主题;同一时期,Kongthon等<sup>[17]</sup>则利用该模型识别文献数据中的实体关系,扩展了传统文献综述维度;后来,Bolelli等<sup>[18]</sup>将LDA用于大规模学术论文集,通过挖掘文档集合中的主题,并结合文档时间顺序来探测数字图书馆主题的主题演化趋势;在个性化推荐场景中,领域相关人员通过对用户查询历史行为和文献元数据进行分析,识别出潜在的文献主题,从而为用户提供更个性化的信息推荐。例如,Jin等<sup>[19]</sup>提出了一种新颖的网络推荐系统,利用LDA识别web文本语义信息,并基于最大熵原则和协同过滤等原理,可用于网络资源的个性化推荐;Harvey等<sup>[20]</sup>也通过整合用户数据到LDA模型中,利用估计的概率分布为用户提供个性化的标签建议,实现了社交标记系统中的个性化标签推荐;在文献资源推荐上,LDA至今仍旧受到较多学者的青睐,被广泛应用于对用户进行图书、学术论文等资源的个性化推荐研究<sup>[21-22]</sup>。

随着机器学习算法种类的增加和性能的提升,情报界在工作中应用机器学习算法的经验也日益丰富,但人们发现没有一种单一的算法能够在各种情报任务下总能表现出最好的效果。因此,对不同的算法进行比较与筛选是情报人员的重要工作,根据情报任务的特点对算法进行筛选和适配成了这一阶段情报学的研究重点之一。例如,为了提升基于协同过滤进行个性化推荐的效果,不同研究针对各种应用场景中存在的微小差异,提出了多元化的适配算法,包括基于用户的协同过滤<sup>[23]</sup>、基于内容的协同过滤<sup>[24]</sup>、基于模型的协同过滤<sup>[25]</sup>等。

### 2.3 基于深度学习的人工智能

#### (1) 主要技术

自21世纪初以来,随着GPU等硬件技术的迅猛发展,计算机的算力显著提升,深度学习成为人工智能领域的关键焦点。深度学习的概念由Hinton等<sup>[26]</sup>于2006年提出,该方法借鉴了人脑由很多神经元组成的特性,利用多个隐藏层的神经网络结构,通过大量的向量计算,学习数据内在信息的高阶表示。通过多层处理,逐渐将初始的“低层”特征表示转化为“高层”特征表示,实现了高效的特征学习和知识表达,从而解决高度非线性的问题。经典的深度学习算法有卷积神经网络(CNN)、循环神经网络(RNN)、长短时记忆网络(LSTM)等,相比前一阶段,深度学习在图像识别、语音识别、自然语言处理等领域研究与应用上取得显著进展。

深度学习技术具备强大的表示能力,在非结构化数据上展现出了卓越的性能。其自动学习特征和层次化知识表示的特点减轻了对人工特征工程的依赖,并具有出色的拟合能力,适用于处理多样化且复杂的任务。深度学习被认为是目前最接近人脑智能学习的方法,一经提出便引起了学术界和工业界的高度重视,为人工智能领域带来了巨大的变革,对产品和服务的发展产生深远影响。

当然,深度学习阶段的人工智能也具有自己的不足。一方面,深度学习模型通常需要大量的训练数据和计算资源,这在很大程度上限制了其在资源有限或数据稀缺的场景下的应用。另一方面,深度学习的可解释性相对较差,这使得模型的决策过程对于人类用户来说难以理解,这种缺乏可解释性的特点在医疗、法律等关键领域中可能会引发信任问题,因为这些领域往往要求模型能够提供明确且可靠的解释。此外,深度学习模型还存在过拟合的风险,这意味着在训练过程中,模型可能会过于关注训练数据中的特定特征,从而无法泛化到新的、未

见过的数据。为了缓解过拟合问题,需要采用正则化技术、数据增强等策略来提高模型的泛化能力。

尽管存在着种种问题,但深度学习被公认为人工智能技术的一个新台阶,实现了对非结构化数据的高效处理,推动了图像识别、语音识别、自然语言处理等领域的进步。

#### (2) 赋能情报工作范例

深度学习阶段的人工智能技术通过其强大的特征学习、非线性建模、大规模数据处理、多模态融合以及迁移学习等特性,基于多层神经网络的向量计算为情报工作提供了更先进和高效的分析工具,同时也关注数据安全问题。

以图像情报分析为例,在深度学习技术出现之前,目标检测与跟踪、地理特征识别、OCR等作为情报工作经典任务的发展非常受限于软硬件工具。而到了深度学习阶段,研究人员通过对CNN、AlexNet、ResNet等经典深度学习算法进行二次开发与优化,将图像情报分析的效率推上了新高度。例如,Szarvas等<sup>[27]</sup>最早利用城市环境照片训练了CNN分类器,实现了行人检测,并发现针对同一任务,该算法在计算需求上远低于前一阶段的SVM;Zhang等<sup>[28]</sup>则将AlexNet应用于遥感影像处理,实现了对大型复杂机场区域的快速探测;近年来,数字人文研究方兴未艾,用于图像处理的深度学习模型在数量和性能上也达到巅峰,研究者借此契机,通过对LeNet、AlexNet、ResNet、VGG等多种深度学习模型进行调参训练与性能比较,提出了甲骨文自动识别的最优方法,为文化遗产的保护和整理提供了新的技术路径<sup>[29]</sup>。值得一提的是,这种在同一情报任务中,不断尝试新算法,并对不同算法模型进行训练、调优与比较,也正是深度学习阶段人工智能技术赋能情报工作的重要特点。

同样,时间序列预测作为情报工作中一类经典的任务,深度学习阶段的人工智能技术对该方面也有深入的聚焦。在早期,RNN一类的循环神经网络被广泛应用于各种场景的时间序列预测,例如,金融和经济周期指数预测<sup>[30]</sup>、电力需求预测<sup>[31]</sup>、工业故障预测<sup>[32]</sup>等。但随着任务和需求复杂度的提升,传统的RNN在处理长时间序列时存在巨大的缺陷,往往会遗漏重要信息,对此,情报领域人员尝试引入LSTM、BiLSTM等网络模型弥补前述不足。后来随着硬件的算力和技术进一步发展,为了更好地解决以往模型在长时间序列上表征能力的缺陷,Google在2017年提出了Transformer架构,并优化了并行化处理模块和可扩展性<sup>[33]</sup>,随后基于Trans-

former 架构的众多时间序列模型被应用于处理各类时间序列预测的情报任务。例如, Zamora-Resendiz 等<sup>[34]</sup>最早提出了一种基于 Transformer 结构的图卷积神经网络, 用于在蛋白质序列分类和回归任务中挖掘可解释的数据特征; 后来, Jumper 等<sup>[35]</sup>通过对 AlphaFold 进行再设计, 实现了以原子精度定期预测蛋白质结构, 在领域内产生了极大的影响。此外, 股票市场指数预测、融合多源数据的个股趋势预测等基于金融数据的时间序列分析与预测也迎来新高度<sup>[36-37]</sup>。总之, 伴随算力与技术的持续发展, 从传统的 RNN 到当前的 Transformer, 在情报工作中发挥各自特色时, 也无一不证实人工智能技术可以为情报工作提供广泛而强大的支持。

此外, 深度学习范式对数据量的高需求, 引起了社会各界对数据安全的担忧。基于此背景, Google 于 2016 年提出了联邦学习范式, 该方法在多个分布式设备或服务端上进行算法训练, 无须共享本地数据<sup>[38]</sup>。通过差分隐私、同态加密和隐私保护集合交集等关键技术, 使得联邦学习成为实现跨部门多源异构数据融合的理论和技术支持。在情报分析中, 这种方法能够解决数据隐私、数据安全和数据访问权限等关键问题, 为多源异构数据的融合分析提供了可能性。例如, 在医疗情报分析场景中, 相关数据在位置(医疗记录、社交媒体)、形式(结构化、半结构化、非结构化)、管理人员(个人、医疗服务机构、政府部门)等方面都非常分散, 这对个人健康和社会福祉的提升造成了较大的障碍。Sun 等<sup>[39]</sup>最早基于联邦学习的理念, 建立了一种稳健、可扩展的数据治理框架, 该框架调和了数据隐私保护与数据融合利用的矛盾; 同时, 基于该框架并利用荷兰统计局等多方数据, 揭示了糖尿病与社会经济因素之间的关系。

人工智能深度学习算法在情报工作中的赋能应用, 显著特点包括强大的特征学习能力、高效处理大规模和序列数据、多模态融合以及迁移学习的灵活性。这些技术不仅提高了情报分析的准确性和效率, 而且为解决传统方法难以应对的挑战提供了新思路。此外, 深度学习还注重数据安全问题, 通过联邦学习范式实现了多源异构数据的隐私保护与融合分析, 为跨部门合作提供了强大的技术支持。总体来说, 人工智能深度学习在情报工作中展现出了深刻的重塑能力, 为相关领域的发展提供了新的技术路径。值得注意的是, 该阶段提出的 Transformer 成为后来众多“基础模型”(如 BERT、GPT)的核心结构, 开启了大模型的浪潮。

## 2.4 基于基础模型的人工智能

### (1) 主要技术

随着深度学习的迅猛发展与技术积累, 包括 Transformer、强化学习等在内的众多新颖网络架构和模式理念应运而生。同时, GPU、量子计算等硬件的飞速发展大大提升了算力, 再加上互联网上积累的海量文本、图像等数据资源, 为参数量更大的神经网络模型的崛起奠定了坚实的基础。在这种背景下, 2018 年 6 月 OpenAI 发布了 GPT-1, 标志着基础模型时代的开端。随后, 微软等科技公司相继推出各类大模型, 使得大模型领域呈现百花齐放、百家争鸣的态势。针对人工智能领域这一异军突起, 美国斯坦福大学的 100 多位学者联名发表了“基础模型: 机遇和挑战”的文章, 将大模型统一命名为“基础模型”(foundation models)<sup>[40]</sup>, 本文对这一阶段的划分, 就沿用了基础模型这一名称。基础模型阶段人工智能技术的显著特征是应用领域通用化以及模型规模巨大化。具体来说, 基础模型的训练数据来源广泛, 且模型目标并未受到局限, 通常定位于跨领域的多任务应用。此外, 这些模型通常拥有非常大的参数容量, 使得它们能够处理和整合来自不同模态的数据, 包括文本、图像、音频等。同时, 基础模型具有高级表示能力和自适应学习能力等特征。

基础模型阶段的主流模型都具备多模态数据处理能力, 多模态基础模型的优势在于其能够将不同模态数据进行融合, 以获得更全面和准确的信息。这种融合方法考虑到了不同模态数据的特点和关系, 使得模型能够更好地理解和处理复杂的任务和问题。同时, 这些模型具有强大的表示能力, 能够学习到更复杂、更细致的信息, 从而更好地适应多样化的任务和场景。此外, 基础模型参数量和训练数据量的提升带来了涌现效应, 也是基础模型的潜在优势。

基础模型技术的不足表现在以下几个方面: 这些模型的庞大规模对计算资源的需求较高, 这一点延续甚至扩大了深度学习的缺陷。再有, 由于模型的规模庞大, 其可解释性进一步变差, 在高度专业化和高信任度需求领域应用时有非常大的挑战。此外, 基础模型目前仍处于探索式研发阶段, 人类社会对其认知程度仍然不够充分, 例如伦理风险等潜在缺陷正在引发关注与研究。

总的来看, 基础模型技术的崛起推动了自然语言处理、计算机视觉等人工智能技术新范式的发展, 对整个人工智能领域产生了深远影响。基础模

型的技术特点使其不局限于语言模型,还包括图像生成、音视频生成等,呈现出趋于融合的姿态。通过无监督的学习模式,基础模型具备了数据标注成本小、任务通用化等特点,可被广泛应用于下游任务,是当前人工智能领域中最受关注的热点。

## (2) 赋能情报工作范例

美国国家情报总监办公室于《2023-2025 年情报界数据战略》<sup>[41]</sup>中指出,人工智能和自动化工具已成为美国情报界实现数据驱动决策的基石,并特别强调了对生成式人工智能在情报领域的关注。基础模型阶段的人工智能技术以模型参数量庞大为特点,以基于大数据进行无监督学习为模型主要训练方式,旨在增强模型的表示能力和泛化性能,同时,通过小样本数据的微调或 Prompt 工程实现模型在多种任务上的通用性目标。该阶段产生的 BERT、T5 以及 GPT 等模型颠覆性地重塑了情报工作的效能。

2018 年 10 月,Google 提出了一种使用双向表征的预训练语言模型 BERT,克服了传统深度学习模型处理长文本时的缺陷,在多项自然语言处理任务上表现出远超越其他模型的效果<sup>[42]</sup>。随后,大量研究人员在情报工作引入该模型用于优化传统情报 workflow。例如,在术语识别工作中,Xue 等<sup>[43]</sup>最早通过微调 BERT 识别中文医学文本中的实体与关系,准确率远超传统模型,开启了该模型在情报领域中解决类似问题的先河;此后,众多研究人员继续基于对 BERT 的微调,高效准确地解决了生物<sup>[44]</sup>、化学<sup>[45]</sup>、医疗<sup>[46]</sup>等多个领域文本的实体或关系抽取。在文本分类工作中,Rodríguez<sup>[47]</sup>最先将包含 BERT 在内的多种预训练模型应用于政策文本的多标签分类;与此同时,还有一些学者针对 BERT 直接进行微调或者融合 CNN 等深度学习算法,将其应用于专利文本的分类,以相对较小的数据标注成本,出色地完成了经典的科技文本处理工作<sup>[48-50]</sup>。至此也可以初步印证,该阶段人工智能赋能情报学的特点是利用模型参数大、表征能力强的优势,使用户在开展下游任务时,通过较少的数据标注即可达到甚至超越深度学习阶段的效果。

伴随 BERT 的广泛应用,不同研究机构争相研发类似训练范式的不同预训练模型,但缺少机制对这些模型的性能与贡献进行综合比较。于是,Google 于 2019 年 10 月提出了一种统一的模型框架 T5,将各种自然语言处理任务都视为 Text-to-Text 任务<sup>[51]</sup>,该模型对于多任务情报工作的流程化、规范化与集成化意义重大。例如,由于在线平台用户

生成内容的繁荣,人工审核并提出意见非常耗时,Laugier 等<sup>[52]</sup>利用垃圾语料数据集对 T5 模型进行微调,协助工作人员生成流畅的审核意见。近来,Pezik 等<sup>[53]</sup>探讨了文本翻译、基于短文本的术语抽取等多任务情报处理的问题,并基于 T5 模型在开放科学元数据语料库上取得了理想的实验效果。

伴随算力的提升,基础模型的参数也在进一步增加,基于生成式路线的基础模型在任务通用性上取得了令人瞩目的效果。2022 年 11 月,基于 GPT-3.5 的人机对话产品 Chat-GPT 上线,更是引起了包括情报工作者在内的各领域人员的关注,进而,情报领域研究人员基于 GPT 对情报工作的方式与理念再度进行赋能。例如,Hassanzadeh<sup>[54]</sup>提出了将 ChatGPT 体现的智能代理技术应用于知识管理的系列想法。国内也有研究人员基于 GPT 开发了学术写作优化产品,该产品具有多功能交互界面,支持自定义和灵活工作流程,优化了用户论文阅读、润色和写作体验<sup>[55]</sup>。同时,还提供强大的分析工具,支持多语言文献翻译,兼容多种应用,提高了文献情报工作的效率和准确性,展现了基础模型在文献情报工作中的广泛应用。此外,众多研究者也聚焦于将知识图谱与 GPT 进行融合,以提升领域知识工程的效率和知识库的质量<sup>[56-57]</sup>。

基础模型在情报工作中的应用展现出了巨大的潜力,通过少量甚至零标注数据,便能够在相关任务中实现理想的效果,显著提高了情报分析的效率和准确性。然而,基础模型对情报的赋能并不仅限于参数微调和文献组织等方面。相反,基础模型对情报工作提出了更高的要求,需要情报系统具备更强的感知、理解和决策能力。为了满足这些要求,当前备受关注的 AI Agent(人工智能代理)具有进一步赋能情报工作的潜力,它为情报工作周期流程化提供了更高的可能性,使情报工作更加高效、自动化和智能化。

## 2.5 小结

在基于规则的阶段,人工智能技术通过利用人工逻辑和规则的形式化表示,模拟了人类情报工作的思维和工作流程,诸如 DENDRAL、MYCIN 和 ESFFIR 等多类情报服务系统在这一阶段崭露头角并繁荣发展。通过预定义的规则和逻辑,计算机能够快速而准确地筛选、分类和分析大量情报数据,尤其是以专家系统为载体,降低了情报检索的技术门槛,使情报分析人员可以聚焦于关键信息,显著提高了情报工作的速度和准确性。

在机器学习阶段,人工智能技术对情报工作的

赋能主要体现在自动化规则建立与高效数据分析两个方面。一方面,通过学习大量标注过的数据样本,机器学习算法能够自动归纳出复杂的工作模型,避免了传统方法中需要手动编写规则的烦琐过程,从而显著提高了工作模型的建立效率,尽管此时针对复杂任务的处理仍需要特征工程进行人为辅助,但机器学习在归纳模型方面的能力为情报工作人员带来了显著的工作效益。另一方面,一旦模型建立,便能够迅速且准确地处理大量情报数据,该阶段基于 SVM、LDA、决策树等机器学习模型的情报工作方式成为主流。这些模型能够相对智能地识别模式、关系和趋势,从海量信息中提取关键数据,使情报分析人员更快速、准确地对情势有所洞察。

在深度学习阶段,人工智能技术为情报工作的范式带来了重大突破。首先,深度学习继承并提升了机器学习阶段的效能,可以通过自动地学习数据特征为情报工作建立模型,显著降低了情报工作建模的难度,但同时模型的可解释性有所降低。其次,深度学习技术直接参与知识组织和加工,如构建知识图谱等知识库,这为情报工作提供了更广泛且自动化的支持,使知识的获取、整理和利用更为便捷,情报工作人员可以更有效地从大量信息中提取有价值的知识,为决策提供更全面的支持。此外,深度学习技术还在大规模情报分析与预测方面发挥关键作用,其强大的模式识别能力使情报分析人员更深入地理解和预测复杂的情报现象,为未来决策提供更为准确的信息基础。

在基础模型阶段,人工智能模型变得更加复杂、对数据需求更大、对算力的要求更高。单一模型应用也由狭隘的场景变得更加通用。然而在当前的情报工作中,尽管基础模型呈现更大的潜力,但由于应用与任务流适配研究尚未深入,学者们仍处于理论性的探讨和场景推测阶段,对于不同大模型的“涌现能力”也尚不明确,因此,仍需以发展的眼光对待基础模型赋能情报工作的前景。

### 3 新时期情报工作应关注的问题

从前述人工智能技术发展及其对情报工作的赋能历程来看,人工智能技术对情报工作发展与变革发挥着巨大的作用。从技术特质与应用场景来看,基础模型之前的人工智能技术基本上属于弱人工智能,它每次只能完成特定或单一类型的任务,情报工作者在应用这类人工智能技术时,关注点通常聚焦于为特定情报任务与应用场景匹配合适的模型并优化模型参数,人工智能技术的角色定位主

要是提高情报工作效率的辅助工具。进入基础模型时代,大模型具备了可观的推理能力,在理论上也具有任务通用性,不需要再为每个情报任务开发或适配不同的模型,特别是以 ChatGPT 为代表的生成式人工智能技术(产品),在人机对话、文本生成、代码生成、图像生成、视频生成、音频生成、音乐生成等方面具备优良的表现。这些技术成果在情报工作的主流业务如资源采集、知识组织、情报分析、科学评价与评估、参考咨询等方面有着广泛的应用前景,甚至会对传统情报工作中的某些环节产生“颠覆性”或“替代性”的影响。因此,笔者认为,为了应对基础模型时代人工智能技术对情报工作带来的巨大机遇与挑战,需要对其涉及的一些问题进行前瞻且深入的思考。

#### 3.1 基础资源建设问题

在基础模型阶段,数据和算法、算力构成了人工智能的三大核心要素。算法可以类比为人的思维方式,算力可以类比为人的反应能力,数据则对应于人所拥有的知识。人区别于其他动物的重要特征是人类能够通过与环境交互学习知识,没有了知识,人的智慧与智能也就无从谈起,人的社会化过程实质上就是学习知识并利用知识提升智慧与智能的过程。同样,数据在人工智能的成功中扮演着至关重要的角色。学者们普遍认识到,人工智能得以快速发展的重要推动因素是用来构建各类智能模型的丰富且高质量的数据,近年来,正是由于数据在人工智能中发挥着越来越重要的作用,科学家们才提出了以数据为中心的人工智能(data-centric artificial intelligence)这一新兴概念,并将注意力逐渐从推进模型设计转向提高数据的质量和数量<sup>[58]</sup>。这里说的数据不是指杂乱无序的数据,而是指经过标注的、有确定标签的数据。人工智能训练的过程就是让计算机去识别这些数据特征并与标签建立函数关系,经过反复地调整,使得人工智能学会数据所代表的内容与规律。以 GPT 为例,据 OpenAI 披露,GPT-3.5 的文本语料多达 45TB,数据量相当于整个维基百科英文版的 160 倍或 472 万套中国四大名著。而 GPT-4 在 GPT-3 和 GPT-3.5 训练数据集的基础上又增加了多模态数据,预训练数据词条(token)达到 13 万亿个,其中,有约 10 万亿个词条来自 Common Crawl 和 Refined Web 等高质量数据源。此外,Facebook 母公司 Meta 发布的首个开源可商用的大语言模型 LLaMA,其预训练时使用的词条数也超过 2 万亿个。然而,有研究机构指出,大模型预估将在 2026 年消耗尽高质量文本数

据,将在2030年到2050年消耗尽图像数据和低质量文本数据<sup>[59]</sup>。由此可见,数据对于训练更加智慧的机器学习、深度学习等模型是不可或缺的,高质量数据在人工智能构建中更是属于战略性的稀有资源。

在基础模型时代,人工智能技术对情报工作的赋能,已经从人工智能算法的优化改进和适应性改造,转变为大规模的数据训练,换句话说,情报领域的高质量的行业数据,决定了人工智能对情报工作的赋能效果,从而决定了情报行业的智能程度,在这种情况下,情报界必须要回答这样的问题:情报领域的核心资源和基础资源是什么?这些资源如何将情报工作的特色反映到基础模型中,使得基础模型具备情报工作要求的智能?虽然情报机构采集并拥有了大量的科技文献,但这些文献并非情报领域的核心资源,原因在于:第一,从资源特点角度看,这些科技文献记载的科学结论,反映的是科学研究的过程和结果,而不是情报工作的过程和结果;第二,从资源归属角度看,情报机构并不拥有这些文献的知识产权,在现有的法律框架下,情报机构并不能直接将这些文献作为情报领域人工智能基础模型训练的核心数据。那么,情报领域的核心资源和基础资源究竟是什么?笔者认为主要有两类,一类是经过情报机构自己加工的元数据资源,不包括原始的、来源于第三方的元数据资源(如套录的元数据);另一类是情报研究成果,这类成果是情报工作的生成内容,反映了情报工作者的思维和情报工作的特色。这两类资源才是情报行业的核心资源,它们由情报工作者创建并拥有自主知识产权。从现状看,目前情报行业这两类资源的积累距离基础模型训练的需要还有相当的差距,尤其是情报研究成果类的资源,由于各情报单位本身的性质、任务不同,承接的情报课题也各有侧重,这就导致每个独立情报机构的此类资源仅仅是众多情报研究成果的很少的一个子集,并且彼此之间由于情报课题来源不同而不能共建共享。情报行业这种资源分散、规模化不强、标注数据数量少的特点,导致训练模型的效果和智能程度有限,进而限制了基础模型时代人工智能在情报领域的普及和深入应用。

为了解决该问题,可以考虑多种解决途径。例如,一方面,由国家科技图书文献中心(NSTL)、中国高等教育文献保障系统(CALIS)以及国家图书馆牵头,联合并协调全国的图书情报机构,建立完备和规模化的情报行业数据集,这有助于提高训练模

型的质量和深度。另一方面,借助图书情报界良好的共建共享机制和传统,加强不同情报机构之间的合作,实现联合标注数据的规模化,从而增加训练模型的效果和覆盖面。这种联合有助于整合各方资源,形成更为综合和丰富的情报数据集,为人工智能在情报领域的应用提供更为强大的支持。

### 3.2 大模型与小模型协调问题

所谓的大模型与小模型,区别主要体现在模型的通用性与参数量两方面,一个模型是否为大模型,主要是看其是否具有通用性,即基于对该模型的微调能否处理众多下游的异质任务。随着人工智能技术的不断发展,特别是近年来,对基础大模型的应用越来越多,基础大模型也成为人工智能应用的热点。然而,大模型的繁荣并不代表小模型失去了其独特的价值和广泛的应用场景。随着人们对大模型认识的深入,人工智能领域对小模型的价值愈发重视,对于情报工作这种专业性和专深性都比较强的研究领域,对小模型的开发以及小模型与大模型的协调更是值得深入分析和研究的问题。

首先,小模型在解决问题的有效性上展现出独特的优势。在某些特定的领域或场景中,使用小模型比使用大模型更为有效。举例而言,在推荐系统中,一个以用户购买行为数据为基础的小模型往往比同类大模型更具实效性。这是因为,在预测用户购买行为这一特定场景中,大模型需要海量的训练数据才能很好地工作,但在实际使用大模型的过程中,用户购买行为数据这类与预测准确率高度相关的高质量数据,却占训练大模型所需数据总量的很少比例,无法真正对大模型参数的调整方向发挥有效影响。其次,小模型在数据隐私和安全方面具备更大的优势。例如,在处理医疗记录或个人身份信息敏感数据的情境下,小模型由于具备相对简易的参数和结构,在实现相同任务效能基础上,其对隐私数据的特征学习会更少,从而更易通过安全性和隐私性检查,因此,在需要处理敏感数据的环境中相对大模型更为适用。此外,小模型在解释性和可理解性方面显著胜过大模型。小模型相对简单的结构使得小模型更容易理解和解释,从而在需要解释复杂决策或结果的场景中更为有用。例如在法律或金融领域,面向具体个体的决策结果往往非常需要逻辑性较强的解释,使用小模型更容易达到这一要求。

小模型的上述特点与情报研究领域的关注点非常契合。情报分析与研究与大数据分析不同,不仅强调相关性,还注重因果解释。由于大模型越来越

越强的不透明性以及不可解释性等特点,比较难以作为情报服务的结果提供可信的依据,这就使得没有解释说明的情报结果难以令人信服。当然,纵然小模型在特定情报场景和任务中具有不可替代的作用,但大模型对情报工作的价值也不能否定,在通用人工智能领域中如此,在情报领域中更是如此。人工智能对情报工作的赋能应该更加重视大模型和小模型二者在情报工作中的定位,并使得二者各司其职,相互协调,发挥各自所长,共同解决问题。由此,笔者认为,为了解决情报工作中大模型与小模型的协调问题,情报领域应正确认识大模型与小模型的特长差异,积极推动情报领域的 AI Agent 研究与实践。AI Agent 是一种基于大模型技术的智能体,旨在让自然人用户通过自然语言与机器交互,使机器高度自动化地执行和处理专业、繁重的工作任务,从而实现对人力资源的高效释放<sup>[60]</sup>。这一概念强调了大模型在解决特定任务和目标方面的独立思考和行动能力,相较于传统的大模型,AI Agent 更注重任务导向和目标驱动。具体来说,利用 AI Agent 可以实现对大模型与小模型二者的协调,一方面,AI Agent 使用自然语言作为主要的交互方式,使情报工作人员能够以更自然、直观的方式与智能系统进行沟通,这种交互方式提高了用户体验,并减少了对特定技能的依赖,降低了复杂情报处理任务的门槛。另一方面,AI Agent 的核心特征在于其能够独立思考并通过调用工具逐步完成情报用户给定的目标。相较于单纯使用大模型或小模型,AI Agent 不但采纳了大模型对用户意图理解更加准确的优势,也融合了小模型在特定垂直场景更加专业和精准的优势,真正体现了情报工作追求的智能化与精准化。

### 3.3 人工智能赋能情报工作的目标和情报服务的边界

基础模型阶段情报服务所面临的一个关键问题是厘清情报工作本质和边界。情报服务经历了文献服务、信息服务、知识服务,最终演变为智慧服务,这在情报界已经有了一定的共识<sup>[61]</sup>。同时,在情报界还有一个共识,情报本身就是人类智慧的体现,智慧情报服务中的“智慧”在本质上并不是指情报本身的智慧,而是指如何利用机器或人工智能技术通过与人的协作提升情报服务过程的智能程度,提升情报服务的效率和精准度<sup>[62]</sup>。在当前基础模型的环境下,需要深入思考情报服务的内在机制和目标。

人工智能的目标是创建从数据中获取价值的

自我学习系统,并使得系统可以应用这些知识以类似人类思考的方式解决新问题<sup>[63]</sup>,通俗地讲,就是用机器代替人的工作。从这个角度出发,我们有必要厘清人工智能技术赋能情报工作到底要代替谁的工作,智能情报服务要代替科学家或者是要代替情报用户本身的工作吗?答案应当是否定的。智能情报工作并不是要代替科学家或情报用户本身的工作,不是要代替、也不可能代替科学家或情报用户来阅读文献、厘清研究思路、开展科学研究等等原本由科学家或情报用户自己完成的工作。今天,情报工作“耳目”“尖兵”“参谋”的职能没有变,人工智能对情报工作的赋能,正是要加强这些职能,使得“准、快、精、全”情报准则得以更加充分地发挥,所以,人工智能赋能情报工作的目的是提升情报人员的能力,它所要强化的是情报工作人员的能力,而不是代替情报用户本身的工作,这两者是有本质区别的。从目前的研究和实践来看,对于这个问题,情报界并没有非常清楚地予以回答。

因此,当前要深入研究情报工作和情报服务的特点,尤其搞清楚情报工作应用基础模型的内在机制和目标显得至关重要,当前的研究在这方面还存在模糊不清的问题。例如,情报收集是情报工作的核心之一,情报信息的真实性和及时性是确保决策者获得准确、可信信息的关键因素,在面对复杂、动态的情报环境时,及时获取真实信息对于制定有效决策至关重要。然而,基于基础模型的情报采集并不能始终满足真实性和及时性的要求,因为,基础大模型主要依赖于历史数据,对于实时事件的准确捕捉和验证存在一定困难,这主要体现在缺乏实时数据、容易产生虚假信息两方面。这就决定了基于人工智能技术的情报收集需要进行实地验证和深入调查,以确保信息的真实性和及时性(这本身就是经典的情报工作任务之一,目前还无法由人工智能技术所替代)。此外,情报界的参考咨询工作,有自身明确的服务范围,咨询工作不能也不可能代替科学家进行科学研究,这就决定了我们参考咨询类的基础模型有别于一般意义上的通用知识模型。类似地,情报分析服务是一种决策服务和预测服务,并非它是它所服务的那个领域的科学研究,单纯的科技文献模型并不能解决决策服务和预测服务的问题。所以,深入研究人工智能在情报领域的应用,明确其在情报领域中应该发挥的作用和应用边界,是当前值得关注的问题。

### 3.4 人工智能的伦理风险与治理

当前及未来相当长一段时间,随着人工智能的

广泛应用,情报工作中伦理风险与政策治理的问题愈加突出。2019年4月,欧盟发布了《可信赖AI的伦理准则》<sup>[64]</sup>,为实现可信赖人工智能全生命周期提出了七个关键要素,涵盖了人的能动性和监督、技术鲁棒性和安全性、隐私和数据管理、透明性、多样性非歧视性和公平性、社会和环境福祉以及问责。其中,人的能动性和监督强调了对人工智能系统的监督和干预,以确保其符合道德和法律标准;技术鲁棒性和安全性要求人工智能系统在应对各种挑战时表现出稳健性,避免对用户和社会带来潜在威胁;隐私和数据管理则关注对用户数据的合理收集、使用和保护,以维护个体隐私权。2023年7月,国家网信办等七部门发布了《生成式人工智能服务管理暂行办法》<sup>[65]</sup>,强调了安全与依法治理原则,通过分类分级监管,实现对生成式人工智能服务的有效治理,明确了对人工智能服务的合规性要求,强调了安全与法治的重要性。2023年11月,首届全球人工智能安全峰会发表了《布莱切利宣言》<sup>[66]</sup>,该宣言肯定了人工智能对人类发展创造的机遇,但包括中国在内的签署成员国也一致认为人工智能对人类社会有潜在的灾难性风险,各国应加强技术的风险管控,就此,中国也围绕人工智能发展、安全、治理三个角度系统地提出并阐述了治理方案。

情报工作是根据特定场景的特定任务,对大量信息进行采集并加以深层次的处理,从而满足情报用户需求、帮助情报用户解决特定问题的过程。在当今数据量大、类型多样、变化性强的时代,借助人工智能技术强化情报工作已是必然趋势。然而,人工智能还存在着许多问题,如机器幻觉、算法偏见、算法不稳健等。首先,大量实践经验表明,基于基础模型的人工智能在内容生成过程中存在前后矛盾、答非所问、胡乱编造的机器幻觉现象,这使得情报人员无法完全依赖人工智能展开分析,并对分析结果存在疑虑;其次,人工智能中存在一定的算法偏见,如招聘过程中的性别偏见、犯罪预测中的种族偏见、人脸识别中的应用偏见<sup>[67]</sup>。这些偏见会加剧信息呈现不平等、信息分布不均衡和新型数字不平等<sup>[68]</sup>,并诱发用户的算法厌恶<sup>[69]</sup>;最后,人工智能算法并不稳健,DARPA在“AI Forward”的项目中指出,一个识别交通标识的智能算法在底层训练集、算法架构、物理特征、抽象特征等方面能被轻易攻击从而破坏其稳健性。我国的相关研究也指出,机器学习和用户行为中存在一定的偏差问题,这些偏差会构成一个循环往复的过程并被不断放大,导

致模型算法的效力下降甚至失效<sup>[70]</sup>。对情报工作而言,人工智能技术的上述问题,恰恰是情报工作的大敌。当前,情报界已经认识到情报工作不能完全依赖人工智能技术,人工智能技术中对情报工作的不可信因素越来越突出,人工智能的这些问题在某种程度上切断了人工智能赋能情报工作的路径,因此在进行情报分析之前有必要对人工智能技术进行可信验证<sup>[71]</sup>。

伦理风险与政策治理是一个复杂而长期的挑战。在人工智能发展的基础模型阶段,情报领域尤其需要关注赋能技术的鲁棒性、透明性、多样性、非歧视性以及公平性。有学者指出,即使基础模型发挥越来越大的作用,但人工智能的主导者终究是人类,因此人工智能治理也应以人的发展为目标,具体实践上应从行动者权责关系入手,即针对人类的治理,应致力提升人本身的道德自觉、伦理责任、人工智能素养;针对人工智能的治理,应坚持数字包容与算法正义<sup>[72]</sup>。也有研究指出人工智能伦理风险的产生路径主要包括技术内生型风险和应用型风险两类<sup>[73]</sup>。在情报领域,笔者认为,治理的关注点应该聚焦于应用型风险防控上,这与基础模型时代情报工作的核心定位、情报工作人员的职能相适应。此外,社会和环境福祉以及问责机制的建立也是需要不断完善的方向。在伦理风险与政策治理方面,各国情报界也应加强国际合作,共同制定更加全面和具有前瞻性的标准,以应对人工智能发展给情报领域带来的伦理挑战。

### 3.5 基础模型的通用尺度与应对策略

关于基础模型的通用尺度以及性能提升路径,也是情报工作面临的一个关键问题。大模型强调通用性,然而在一些高度专业化的领域,如法律和医疗,通用性的基础大模型难以取得理想效果,这主要表现在迭代速度无法跟上领域知识变化的步伐上,而这些领域对结果的要求非常高,这必然会影响到人工智能对这些领域的情报工作的赋能效果。

面对这一矛盾,基础模型的发展和通用尺度涉及多个值得深思的问题。通用是指任务通用还是领域通用?是否能够跨越多个领域,以及如何跨越领域?当通用性效果不理想时,又如何提升?就此,研究人员提出了三种可能的途径:第一种途径,通过极致提升基础大模型参数、训练数据的数量和质量,并在不同领域开发AI Agent,实现跨领域、跨任务的性能提升。持续极致提升基础大模型的参数规模和训练数据的数量与质量,这涉及参数优化和数据集的精心构建,以确保模型能够更好地捕捉

领域内的复杂关系和特定概念。此外,通过在不同领域开发跨领域的 AI Agent,可以实现在专业领域的跨任务性能提升,这种方法促进了横向知识迁移,使模型在一个领域学到的知识能够在另一个领域得到有效应用。第二种途径,可以专注于训练垂直领域的基础大模型,使其在垂直领域具备跨任务的高性能。这需要模型的定制化,以更好地适应领域内的任务和语境。在这个过程中,任务的共享与优化是关键,通过底层任务相关知识的共享,实现跨任务的高性能,包括任务联合训练和迁移学习。同时,引入领域专业人员的知识,与人工智能研究人员协同工作,有助于确保模型更好地理解领域内的专业术语、文本语境和任务需求。第三种途径,结合跨领域基础大模型和领域知识图谱,通过基础模型提升知识图谱构建效能,并将知识图谱视为知识库与大模型融合,以提高在专业领域的性能表现。而这种融合也存在不同的思路,一是基础模型赋能知识图谱的构建<sup>[74]</sup>,提升知识图谱的效能,包括实体抽取、关系抽取、事件抽取等;二是将知识图谱视为基础模型的外源知识库供基础模型调用<sup>[75]</sup>。总体来说,这些路径的选择取决于具体应用场景、资源投入和问题需求,不同的方法可能会在不同的领域和任务中取得最佳性能。

这三条路径为基础模型在通用性与专业领域性能提升方面提供了有益的思路。通过这些路径的深入研究和实践,可以更好地应对基础模型在情报工作中面临的复杂问题,实现更为全面和灵活的应用。第一种途径强调通过极致提升基础大模型的参数、训练数据以及开发跨领域 AI Agent 来实现性能提升,第二种途径专注于训练情报领域的垂直基础大模型。然而,在情报领域,基础资源建设尚不足,这使得完全依赖这两种途径可能不够现实,且极致追求提升基础模型的通用性能并非情报工作者的特长与职责。因此,我们可以考虑将关注点转向第三种途径,并适当结合第一种和第二种途径的思想和方法。第三种途径强调了跨领域基础大模型与领域知识图谱的融合。在情报领域,已经有了丰富的知识图谱研究,这使得将知识图谱应用于基础模型成为一个更为可行的选项,我们可以结合 AI Agent 和垂直训练的思想,利用情报领域已经建立的知识图谱和构建知识图谱的经验,探索将知识图谱作为基础模型的外源知识库的可能性。这不仅可以提高模型在专业领域的性能,还能够充分利用情报领域已有的知识资源。

为了推动这些研究方向,建议自然科学基金和

社会科学基金立项资助以上述途径为内容的研究,以便深入比较三种路径在情报领域的实际效果。通过这样的研究探索,使得情报工作者找到最适合情报领域的人工智能赋能道路,为情报工作提供更为高效和经济的解决方案。

#### 4 结语

在人工智能技术飞速发展的新时期,本研究回顾了人工智能的发展历程,探讨了各阶段人工智能技术对情报工作的赋能方式。尽管大模型的崛起为情报信息采集、处理及服务提供了先进工具,但也带来了一系列新的挑战。情报工作者在新时期需要不断适应高度技术化的环境,这就要求我们未雨绸缪,紧密结合情报工作的特点,对可能存在的问题进行深入的思考,充分发挥人工智能赋能情报工作的优势,推动情报工作的发展。

#### 参考文献:

- [1]白如江,陈鑫,任前前.基于供需理论的生成式人工智能赋能情报工作范式模型构建与应用研究[J].情报理论与实践,2024,47(1):75-83.
- [2]何立民.人工智能系统智能生成机理探索之六:从弱人工智能、强人工智能到超人工智能[J].单片机与嵌入式系统应用,2020,20(8):87-89.
- [3]Shoval P. Principles, procedures and rules in an expert system for information retrieval[J]. Information Processing and Management, 1985, 21(6):475-487.
- [4]陈光祚.专家系统在情报检索中的应用[J].图书情报知识,1986(2):28-32.
- [5]黄祥喜,刘卫国.智能情报检索系统的结构和问题[J].情报学报,1987,6(1):69-76.
- [6]汪红秋.反馈情报检索专家系统[J].情报学报,1988,7(5):345-352.
- [7]Cortes C, Vapnik V. Support - vector networks [J]. Machine Learning, 1995, 20(3):273-297.
- [8]Joachims T. Text categorization with support vector machines; Learning with many relevant features [C]//Proceedings of the European Conference on Machine Learning. Heidelberg: Springer, 1998:137-142.
- [9]Dumais S, Chen H. Hierarchical classification of web content [C]//Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. Athens, Greece. New York: ACM, 2000:256-263.
- [10]李晓黎. WEB 信息检索与分类中的数据采掘研究 [D].北京:中国科学院研究生院(计算技术研究所),2001.
- [11]Tian W X, Zhu F X. Learning to rank using semantic features in document retrieval [C]//2009WRI Global Congress on Intelligent Systems. Xiamen, China. IEEE, 2009:500-504.
- [12]Fall C J, Törösvári A, Benzineb K, et al. Automated cat-

egorization in the international patent classification [C]//ACM SIGIR Forum. New York: ACM, 2003, 37(1): 10-25.

[13] 李程雄, 丁月华, 文贵华. SVM-KNN 组合改进算法在专利文本分类中的应用[J]. 计算机工程与应用, 2006, 42(20): 193-195, 212.

[14] 符川川, 陈国华, 袁勤俭. 基于机器学习的专利质量分析与分类预测研究——以区块链技术专利为例[J]. 现代情报, 2021, 41(7): 110-120.

[15] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3: 993-1022.

[16] AlSumait L, Barbarú D, Domeniconi C. On-line LDA: Adaptive topic models for mining text streams with applications to topic detection and tracking [C]//2008 Eighth IEEE International Conference on Data Mining. Pisa, Italy. Piscataway: IEEE, 2008: 3-12.

[17] Kongthon A, Haruechaiyasak C, Thaiprayoon S. Enhancing the literature review using author-topic profiling [C]// International Conference on Asian Digital Libraries. Heidelberg: Springer, 2008: 335-338.

[18] Boelli L, Ertekin S, Zhou D, et al. Finding topic trends in digital libraries [C]// Proceedings of the 9th ACM/IEEECS Joint Conference on Digital Libraries. Austin, TX, USA. New York: ACM, 2009: 69-72.

[19] Jin X, Zhou Y Z, Mobasher B. A maximum entropy web recommendation system: Combining collaborative and content features [C]// Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining. Chicago, Illinois, USA. New York: ACM, 2005: 612-617.

[20] Harvey M, Baillie M, Ruthven I, et al. Tripartite hidden topic models for personalised tag suggestion [C]// European Conference on Information Retrieval. Heidelberg: Springer, 2010: 432-443.

[21] 郑祥云, 陈志刚, 黄瑞, 等. 基于主题模型的个性化图书推荐算法[J]. 计算机应用, 2015, 35(9): 2569-2573.

[22] 李晓敏, 王昊, 李跃艳. 基于细粒度语义实体的学术论文推荐研究[J]. 情报科学, 2022, 40(4): 156-165.

[23] 王井. 一种基于订阅记录的图书协同过滤推荐方法研究[J]. 情报科学, 2020, 38(3): 54-59, 77.

[24] Geetha G, Safa M, Fancy C, et al. A hybrid approach using collaborative filtering and content based filtering for recommender system [C]// National Conference on Mathematical Techniques and its Applications (NCMTA 18). Bristol: IOP Publishing, 2018, 1000: 012101.

[25] Jiang S H, Qian X M, Shen J L, et al. Author topic model-based collaborative filtering for personalized POI recommendations [J]. IEEE Transactions on Multimedia, 2015, 17(6): 907-918.

[26] Hinton G E, Osindero S, Teh Y W. A fast learning algorithm for deep belief nets [J]. Neural Computation, 2006, 18(7): 1527-1554.

[27] Szarvas M, Yoshizawa A, Yamamoto M, et al. Pedestrian detection with convolutional neural networks [C]// IEEE Proceedings on Intelligent Vehicles Symposium. Las Vegas, NV, USA. Piscat-

away: IEEE, 2005: 224-229.

[28] Zhang P, Niu X, Dou Y, et al. Airport detection from remote sensing images using transferable convolutional neural networks [C]// 2016 International Joint Conference on Neural Networks (IJCNN). Vancouver, BC, Canada. Piscataway: IEEE, 2016: 2590-2595.

[29] Fujikawa Y, Li H Y, Yue X B, et al. Recognition of oracle bone inscriptions by using two deep learning models [J]. International Journal of Digital Humanities, 2023, 5(2): 65-79.

[30] Kaastra I, Boyd M. Designing a neural network for forecasting financial and economic time series [J]. Neurocomputing, 1996, 10(3): 215-236.

[31] Connor J T, Martin R D, Atlas L E. Recurrent neural networks and robust time series prediction [J]. IEEE Transactions on Neural Networks, 1994, 5(2): 240-254.

[32] Ho S L, Xie M, Goh T N. A comparative study of neural network and Box - Jenkins ARIMA modeling in time series prediction [J]. Computers and Industrial Engineering, 2002, 42(2/3/4): 371-375.

[33] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need [C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach, California, USA. New York: ACM, 2017: 6000-6010.

[34] Zamora - Resendiz R, Crivelli S. Structural learning of proteins using graph convolutional neural networks [J]. BioRxiv, 2019: 610444. Doi: <https://doi.org/10.1101/610444>.

[35] Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold [J]. Nature, 2021, 596(7873): 583-589.

[36] Wang C J, Chen Y Y, Zhang S Q, et al. Stock market index prediction using deep transformer model [J]. Expert Systems with Applications, 2022, 208: 118128.

[37] Zhang Q Y, Qin C, Zhang Y F, et al. Transformer-based attention network for stock movement prediction [J]. Expert Systems with Applications, 2022, 202: 117239.

[38] McMahan H B, Moore E, Ramage D, et al. Communication-efficient learning of deep networks from decentralized data [EB/OL]. [2023-11-20]. <http://arxiv.org/abs/1602.05629v4>.

[39] Sun C, Ippell L, Wouters B, et al. Analyzing partitioned FAIR health data responsibly [EB/OL]. [2023-11-20]. <http://arxiv.org/abs/1812.00991v1>.

[40] Bommasani R, Hudson D A, Adeli E, et al. On the opportunities and risks of foundation models [EB/OL]. [2023-11-20]. <http://arxiv.org/abs/2108.07258v3>.

[41] Intelligence community data strategy for 2023 - 2025 [EB/OL]. [2023-11-25]. <https://www.odni.gov/index.php/newsroom/reports-publications/reports-publications-2023/3708-ic-data-strategy-2023-2025>.

[42] Devlin J, Chang M W, Lee K, et al. BERT: Pre-training of deep bidirectional transformers for language understanding [EB/OL]. [2023-11-27]. <http://arxiv.org/abs/1810.04805v2>.

[43] Xue K, Zhou Y M, Ma Z Y, et al. Fine-tuning BERT for

joint entity and relation extraction in Chinese medical text [ C ] // 2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). San Diego, CA, USA. Piscataway: IEEE, 2019: 892-897.

[ 44 ] Sun C, Yang Z H. Transfer learning in biomedical named entity recognition: An evaluation of BERT in the Pharma-CoNER task [ C ] // Proceedings of the 5th Workshop on BioNLP Open Shared Tasks. Hong Kong, China. Stroudsburg, PA, USA: Association for Computational Linguistics, 2019: 100-104.

[ 45 ] Zhai Z N, Nguyen D Q, Akhondi S A, et al. Improving chemical named entity recognition in patents with contextualized word embeddings [ EB/OL ]. [ 2023-11-27 ]. <http://arxiv.org/abs/1907.02679v1>.

[ 46 ] Li F, Jin Y H, Liu W S, et al. Fine-tuning bidirectional encoder representations from transformers (BERT)-based models on large-scale electronic health record notes: An empirical study [ J ]. JMIR Medical Informatics, 2019, 7 ( 3 ): E14830.

[ 47 ] Rodríguez Medina S. Multi-label text classification with transfer learning for policy documents: The case of the sustainable development goals [ D ]. Sweden: Uppsala University, 2019.

[ 48 ] Lee J S, Hsiang J. PatentBERT: Patent Classification with Fine-Tuning a pre-trained BERT Model [ EB/OL ]. [ 2023-11-27 ]. <http://arxiv.org/abs/1906.02124v2>.

[ 49 ] Meindla B, Ottb I, Zierahn U. Binary patent classification methods for few annotated samples [ C/OL ] // Proceedings of the 1st Workshop on Patent Text Mining and Semantic Technologies, 2019. [ 2023-11-25 ]. <https://doi.org/10.34726/pst2019.4>.

[ 50 ] Lu X L, Ni B. BERT-CNN: A hierarchical patent classifier based on a pre-trained language model [ EB/OL ]. [ 2023-11-27 ]. <http://arxiv.org/abs/1911.06241v1>.

[ 51 ] Raffel C, Shazeer N, Roberts A, et al. Exploring the limits of transfer learning with a unified text-to-text transformer [ EB/OL ]. [ 2023-11-30 ]. <http://arxiv.org/abs/1910.10683v4>.

[ 52 ] Laugier L, Pavlopoulos J, Sorensen J, et al. Civil rephrases of toxic texts with self-supervised transformers [ EB/OL ]. [ 2023-11-30 ]. <http://arxiv.org/abs/2102.05456v2>.

[ 53 ] Pezik P, Mikołajczyk A, Wawrzyński A, et al. Keyword extraction from short texts with a text-to-text transfer transformer [ C ] // Asian Conference on Intelligent Information and Database Systems. Singapore: Springer, 2022: 530-542.

[ 54 ] Hassanzadeh M. Intelligent agents and facilities for knowledge management: ChatGPT and beyond [ J ]. Sciences and Techniques of Information Management, 2022, 8 ( 4 ): 7-22.

[ 55 ] GPT\_academic [ EB/OL ]. [ 2024-01-10 ]. [https://github.com/binary-husky/gpt\\_academic](https://github.com/binary-husky/gpt_academic).

[ 56 ] Yang L Y, Chen H Y, Li Z, et al. Give us the facts: Enhancing large language models with knowledge graphs for fact-aware language modeling [ EB/OL ]. [ 2023-11-30 ]. <http://arxiv.org/abs/2306.11489v1>.

[ 57 ] Zhang J W. Graph-ToolFormer: To empower LLMs with graph reasoning ability via prompt augmented by ChatGPT [ EB/

OL ]. [ 2023-11-30 ]. <http://arxiv.org/abs/2304.11116v3>.

[ 58 ] Zha D C, Bhat Z P, Lai K H, et al. Data-centric artificial intelligence: A survey [ EB/OL ]. [ 2023-12-05 ]. <http://arxiv.org/abs/2303.10158v3>.

[ 59 ] Villalobos P, Sevilla J, Heim L, et al. Will we run out of data? An analysis of the limits of scaling datasets in machine learning [ EB/OL ]. [ 2023-12-05 ]. <http://arxiv.org/abs/2211.04325v1>.

[ 60 ] Xu M R, Niyato D, Kang J W, et al. When large language model agents meet 6G networks: Perception, grounding, and alignment [ EB/OL ]. [ 2024-02-21 ]. <http://arxiv.org/abs/2401.07764v2>.

[ 61 ] 罗立群, 李广建. 智慧情报服务与知识融合 [ J ]. 情报资料工作, 2019, 40 ( 2 ): 87-94.

[ 62 ] 王延飞, 赵柯然, 何芳. 重视智能技术 凝练情报智慧——情报、智能、智慧关系辨析 [ J ]. 情报理论与实践, 2016, 39 ( 2 ): 1-4.

[ 63 ] 钟义信. 人工智能: 概念·方法·机遇 [ J ]. 科学通报, 2017, 62 ( 22 ): 2473-2479.

[ 64 ] Ethics guidelines for trustworthy AI [ EB/OL ]. [ 2023-12-09 ]. <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>.

[ 65 ] 生成式人工智能服务管理暂行办法 [ EB/OL ]. [ 2023-12-09 ]. [https://www.gov.cn/zhengce/zhengceku/202307/content\\_6891752.htm](https://www.gov.cn/zhengce/zhengceku/202307/content_6891752.htm).

[ 66 ] The Bletchley Declaration [ EB/OL ]. [ 2023-12-09 ]. <https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration>.

[ 67 ] 王小光. 人脸识别技术应用与侦查情报分析的偏差控制 [ J ]. 情报杂志, 2023, 42 ( 4 ): 50-56.

[ 68 ] 贾诗威, 闫慧. 算法偏见概念、哲理基础与后果的系统回顾 [ J ]. 中国图书馆学报, 2022, 48 ( 6 ): 57-76.

[ 69 ] 李游, 梁哲浩, 常亚平. 用户对人工智能产品的算法厌恶研究述评及展望 [ J ]. 管理学报, 2022, 19 ( 11 ): 1725-1732.

[ 70 ] 郭迅华, 吴鼎, 卫强, 等. 机器学习与用户行为中的偏差问题: 知偏识正的洞察 [ J ]. 管理世界, 2023, 39 ( 5 ): 145-159, 199, 160-162.

[ 71 ] 丁晓蔚, 苏新宁. 基于区块链可信大数据人工智能的金融安全情报分析 [ J ]. 情报学报, 2019, 38 ( 12 ): 1297-1309.

[ 72 ] 李韬, 周瑞春. 生成式人工智能的社会伦理风险及其治理——基于行动者网络理论的探讨 [ J ]. 中国特色社会主义研究, 2023 ( 6 ): 58-66, 75.

[ 73 ] 赵志耘, 徐峰, 高芳, 等. 关于人工智能伦理风险的若干认识 [ J ]. 中国软科学, 2021 ( 6 ): 1-12.

[ 74 ] Zhu Y Q, Wang X H, Chen J, et al. LLMs for knowledge graph construction and reasoning: Recent capabilities and future opportunities [ EB/OL ]. [ 2023-12-22 ]. <http://arxiv.org/abs/2305.13168v1>.

[ 75 ] Wen Y L, Wang Z F, Sun J M. MindMap: Knowledge graph prompting Sparks graph of thoughts in large language models [ EB/OL ]. [ 2023-12-22 ]. <http://arxiv.org/abs/2308.09729v4>.