引用格式： 方思怡.大模型赋能标准数字化应用的路径思考与发展建议[J].标准科学, 2025(6): 29-36.
FANG Siyi. Thoughts and Development Suggestions on the Path of Empowering Standard Digital Applications with Large Language Models [J].Standard Science, 2025 (6): 29-36.

大模型赋能标准数字化应用的路径思考与发展建议

方思怡

（上海市质量和标准化研究院）

摘　要：【目的】大模型技术能有效推进标准数字化的深入发展，对标准数字化转型有重要的意义。【方法】通过文献分析、文本挖掘、定性分析等方法，探讨大模型在标准数字化领域的应用前景，总结大模型在标准数字化领域的应用现状，并基于小样本国家标准数据集初步探索大模型在特定标准数字化场景中的应用效果。【结果】提出大模型赋能不同层级标准数字化应用的技术路线图，并针对大模型的潜在问题给出了大模型在标准数字化应用中的发展建议。【结论】从大模型的角度出发，为标准数字化的深入发展提供一定的技术性参考。

关键词： 大模型；标准数字化；标准语料；标准智能体；人工智能

DOI编码： 10.3969/j.issn.1674-5698.2025.06.004

Thoughts and Development Suggestions on the Path of Empowering Standard Digital Applications with Large Language Models

FANG Siyi

(Shanghai Institute of Quality and Standardization)

Abstract: [Objective] Large language model can effectively promote the in-depth development of standard digitization, and plays a pivotal role in the transformation of standard digitization. [Methods] Through literature analysis, text mining, qualitative analysis and other methods, this paper discusses the prospect of applying large language model in the field of standard digitization, summarizes the application status of large language model in the field of standard digitization, and preliminarily explores the application effect of large language model in the specific standard digitization scene based on a small sample of national standard dataset. [Results] This paper proposes a technical roadmap for the large language model to enable the digital application of standards at different levels, and gives suggestions for the development of the large language model in the digital application of standards in view of the potential problems. [Conclusion] From the perspective of the large language model, it provides technical reference for the further development of standard digitization.

Keywords: large language model; standard digitization; standard corpus; standard intelligent agent; artificial intelligence

作者简介： 方思怡，硕士，工程师，研究方向为标准数字化、标准知识服务。

0　引 言

随着数字经济时代的到来，标准数字化转型已经成为国内外标准领域的重大战略发展方向，目前普遍将实现机器可读标准视为标准数字化转型的核心[1]。近年来，我国围绕标准数字化的顶层设计、基础建设、应用场景等方面陆续开展了一系列研究[2]。在众多信息技术中，人工智能已成为标准数字化转型的关键核心技术之一[3]。随着标准数字化技术路线不断完善，人工智能在标准数字化中的应用深度和广度也在不断拓展。作为人工智能领域的新兴技术，快速发展的大语言模型（Large language model, LLM）技术已一跃成为赋能行业发展的焦点。大语言模型，简称为“大模型”，是一种包含千亿级参数且在大规模、多模态语料库上预训练而得的大型深度学习模型[4]。它的出现标志着自然语言处理和生成进入了新阶段[5]。与以往的深度学习模型相比，大模型具有较强的涌现（Emergent）能力，其优势主要来自思维链（Chain-of-Thought, CoT）、知识蒸馏[6]、基于人类反馈的强化学习（Reinforcement Learning from Human Feedback, RLHF）等技术。自OpenAI在2022年发布ChatGPT后，国内外的大语言模型呈现迅猛发展之势，开启“百模大战”。以DeepSeek为代表的国产开源大模型在2024年底迅速崛起，带领国产大模型进入新一轮的发展历程。作为新质生产力的重要组成部分[7]，大模型在标准领域的应用已是大势所趋。如何在标准数字化转型浪潮中把握好大模型的“东风”，加快标准数字化转型的步伐，已成为当前标准数字化工作的焦点之一[8]。

本文主要探讨大模型在标准数字化领域的应用前景，总结大模型在标准数字化领域的应用现状，并以小样本国家标准数据集为例，初步分析大模型在部分标准数字化场景中的应用效果，针对大模型的潜在问题，提出大模型在标准数字化应用中的发展建议，以期能够为标准数字化转型提供一定的技术参考。

1　大模型在标准数字化领域的应用前景

标准是一种经由相关方协商一致、按照特定程序所制定的可共同和重复使用的技术性文件[9]。当前与标准存在密切关联的标准衍生数据主要有相关政策文件、专利文本、论文、法律法规、标准体系、产品信息等。由此可见，当前的标准文本及其衍生数据以文本和图片模态的数据为主。截至目前，国内外已发布一系列基础大模型和行业垂类大模型。大模型的能力图谱已经涵盖了常规数据模态的处理能力，包括文本生成、语音识别、视频生成、图像理解等，已具备实现标准数字化应用的能力基础。

从标准文本及衍生数据的机器可读水平出发，基于信息管理领域DIKW模型的4层结构[10]，将标准数字化工作由低到高依次划分为标准数据获取层、标准数据建设层、标准知识管理层和标准应用场景层，其水平分别与DIKW模型的数据（Data）、信息（Infomation）、知识（Knowledge）、智慧（Wisdom）相对应。根据目前国内外常见大模型的能力特点，围绕标准数字化的发展需求，提出大模型赋能不同层级标准数字化应用的技术路线图，见图1。

1.1　标准数据获取层

标准数据获取层处于DIKW模型的数据层级，旨在获取标准文本数据、直接来自标准文本的标准衍生数据及标准文本以外的标准衍生数据。

在标准数据获取层，大模型能够参与标准文本及标准衍生数据的获取，通过多模态数据的处理能力解决当前部分标准在高质量语料数据获取上存在的问题。与国外标准相比，目前我国的国家标准和行业标准大多以纸质文本和扫描件PDF文本等非结构化的形式流通，机器可读等级较低。为了获取上述格式的标准文本内容，通常需要采用光学字符识别（Optical Character Recognition, OCR）工具将非结构化标准文本转化为机器可读取和可操作的电子数据形式[11]。标准中的技术信息通常以公式、指标数值等细粒度的数据形式出现，

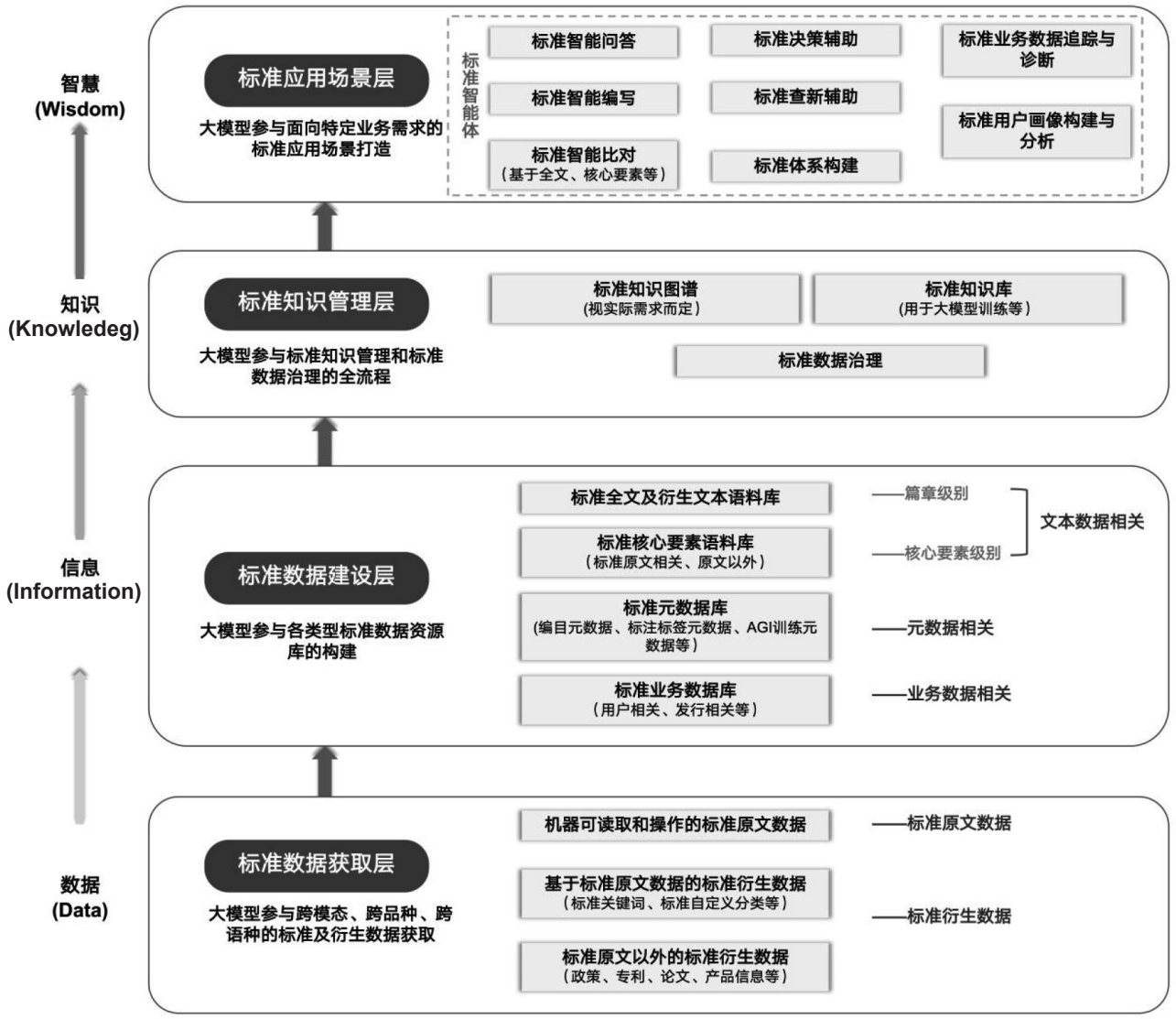

图1 大模型赋能标准数字化的技术路线图

部分标准的插图也蕴含关键的技术信息，但此类图像数据长期以来大多未能得到知识化加工。传统的OCR技术在自动识别上述类型的数据时准确度不够高，泛化能力不足，已无法充分满足扫描件PDF文本的数据获取需求。与以往的OCR工具相比，基于大规模预训练而来的大模型具有更强的泛化能力，在跨语种和多模态的复杂文档识别上表现更优异。将特定的大模型应用于非结构化标准文本，将有效解决现阶段机器可读能力低水平标准的痛点，从而提升标准语料数据的获取质量，为标准数字化工作奠定更为坚实的数据基础。

1.2 标准数据建设层

标准数据建设层处于DIKW模型的“信息”层级，其目的在于打造和储存源自标准文本及标准衍生数据的语料库。

在标准数据建设层，以文本挖掘与生成能力见长的大模型能够优化和加快标准语料数据库的构建。近年来，大模型赋能的语料库建设已在图情领域取得一定的实践成果[12]。就标准文本的功能属性而言，标准是科技文献的一大分支，与图情领

域密不可分。本文根据标准语料数据的类型，将常见的标准数据资源划分为标准全文语料库、标准核心要素语料库、标准元数据语料库和标准业务数据语料库。其中，标准核心要素语料库包括直接和间接来自标准原文数据的核心要素；标准元数据语料库涵盖了标准编目、标注、训练数据等方面的元数据；标准业务语料库与标准应用相关，以标准业务的用户数据为主。

对于标准核心要素语料库的构建而言，在大模型时代之前，直接来自标准原文数据的核心要素通常有3种收集渠道：（1）标准题录数据，以标准号、标准名称、标准实施时间等不深入涉及标准技术信息的核心要素为主；（2）通过基于规则和深度学习相结合的命名实体识别（Naming Entity Recognition, NER）技术抽取标准文本中的术语、指标、范围、规范性引用文件等核心要素，其中标准指标在标准文本中的分布位置和构成形式较为复杂，是标准核心要素语料库构建的一大难点；（3）采用传统机器学习方法获取基于标准原文数据的核心要素，这一类要素以标准主题关键词为典型代表，可采用文本挖掘中的潜在狄利克雷分配（Latent Dirichlet Allocation, LDA）模型获得。相比以往的深度学习模型和传统机器学习模型，大模型在长文本的自然语言处理上优势显著，主动学习能力更强，可通过微调迅速适应全新的领域[13]。标准是一种横跨不同专业领域的技术性文本，大模型的上述优势能节省标准命名实体识别在跨专业领域上的训练成本，其自然语言生成能力也能在标准主题关键词等生成式的核心要素获取上得到充分应用。

1.3 标准知识管理层

标准知识管理层处于DIKW模型的知识层级，旨在建立不同标准语料之间的关联性，构建标准知识图谱，将不同类型的标准语料进一步转化为机器可理解的标准综合知识库，并开展标准知识管理与数据治理。其中，标准知识图谱是对标准知识进行重组并建立关联性关系的新型结构化知识库[14]，主要涵盖来自标准文本和标准文本衍生数据的知识；而标准综合知识库则是在标准知识图谱的基础上与大模型技术深度结合后优化而成的知识库[15]。

在标准知识管理层，大模型能参与标准知识图谱的构建与应用，整合源自标准文本及衍生数据的标准知识，形成更为丰富、全面的标准知识网络，加强标准综合知识库的建设，提升标准知识图谱的应用效能。具体而言，标准知识图谱的构建流程通常包括知识抽取、知识表征、知识融合和知识推理[16]。与基于规则的自然语言处理技术和以往的深度学习模型相比，大模型在语义理解、内容生成上具有较强的通用能力。近来科技情报领域的知识融合研究显示了大模型在知识融合上的优势[17]。在标准知识融合中采用大模型技术能提升标准知识融合的效率。大模型与知识图谱的有机结合也逐渐成为构建高质量知识库的全新方式。将大模型与标准知识图谱深度结合，能有效降低大模型的“幻觉”现象，以共同协作的方式打造的标准综合知识库也可作为优化大模型性能的重要输入，从而进一步提升大模型在标准数字化中的应用效果。

1.4 标准应用场景层

标准应用场景层处于DIKW模型的智慧层级。当前大模型技术主要以人工智能体（AI Agent）为载体实现落地应用。随着人工智能体的框架愈发成熟，知识智能体化的趋势愈发明显。基于大模型的人工智能体已成为大模型近来的重要发展方向。它以大模型为核心控制器，通过整合规划、记忆等不同模块的组件[18]，基于自主规划的指令完成任务。人工智能体的一大优势在于能够将复杂场景简单化，将复杂的应用场景分解为可复用、可推广的简单子任务。

在标准应用场景层，大模型将通过人工智能体的方式，面向标准业务打造具体的标准应用场景，主要包括标准智能编写、标准智能翻译、基于标准全文或核心要素的标准智能比对、标准体系智能构建、标准决策辅助、标准查新辅助、标准舆情智能追踪与分析、标准业务数据智能分析与诊断、标准用户画像自动构建与分析等。

2 大模型在标准数字化领域的应用现状与探索

2.1 大模型在标准数字化领域的应用现状

鉴于标准是具有版权保护的技术性文本[19]，目前标准数据资源尚未被国内外大多数大模型纳入语料训练的范畴。当前大模型技术在标准数字化领域的应用尚处于起步阶段，聚焦标准领域的行业垂类大模型数量较少，大多采用直接调用或微调大模型的应用方式。

在行业垂类大模型的构建与应用方面，作为国内知名的知识服务提供机构，同方知网于2024年4月正式发布了中华知识大模型2.0版本（简称为华知大模型2.0），并与中国标准出版社联合制定了面向标准领域的中华标准大模型。截至2025年2月，中华标准大模型的功能尚未全部上线。根据其官网信息可知，该模型以华知大模型为底座，其数据资源涵盖大量国家标准、行业标准、地方标准、团体标准及部分规程规范，以标准知识问答为核心能力，打造标准文档辅助阅读、标准文件智能写作、智能翻译、标准比对等功能应用。

在直接调用或微调大模型开展相关应用方面，郑佳明等[14]提出了适用于船舶标准领域的大模型与知识图谱的融合应用方法。该方法能有效发挥大模型辅助知识图谱构建和知识图谱辅助大模型研发的双向增强作用，为其他领域的标准数字化建设提供了一定的技术参考。王立玺等[8]总结了中国电子技术标准化研究院在开展大模型应用实践方面的初步成果，包括知识标注、检索增强、智能问答、内容生成、阅读辅助理解、技术要素比对与分析等方面。从中可见，大模型在标准细粒度知识抽取与标注、标准内容语义理解与摘要生成等任务上取得明显成效。

2.2 基于大模型的国家标准数字化应用的初步探索

标准的中文名称是揭示标准主题的标准核心要素，通常涉及标准中的重要术语，与标准编写的目的、范围也存在一定的相关性。本文聚焦强制性国家标准的中文名称，开展大模型在国家标准数字化应用中的初步探索。

为了能够深入比较不同大模型和传统机器学习模型的应用效果，采用DeepSeek-R1模型、DeepSeek-V3模型、讯飞星火Spark-Max模型及传统机器学习下的LDA模型。在上述4种模型中，DeepSeek-R1模型在后训练阶段采用强化学习技术，以推理能力见长；DeepSeek-V3模型在百科知识和长文本处理上表现较好；Spark-Max模型适用于对知识专业性要求较高的知识服务应用场景；传统的LDA模型则是自然语言处理中的经典方法，其通过一系列概率分布判断词语与主题的相关性[20-21]，常被用来揭示文本的核心内容。

以上海标准文献馆“标准文献发行服务系统”中食品安全与化学领域的100篇现行强制性国家标准为样本（截至2025年2月），采用大模型技术基于强制性国家标准的标准中文名称获取5个中文关键词和5个英文关键词。考虑到当前的LDA模型不具备大模型通常具有的文本翻译、内容推理等能力，采用LDA模型基于强制性国家标准的标准中文名称获取中文关键词。为了进一步比较不同技术方法的语义加工效果，采用jieba分词工具，对100篇现行强制性国家标准的标准中文名称进行词语切分和数据处理后获得对应的词语序列。上述4种模型和jieba分词工具的结果各计100条。采用人工方式逐一判断上述运行结果的合理性与准确性。表1为4种模型和jieba分词工具运行结果的部分实例，展示了不同模型和工具对于GB 29940—2013的处理效果。

通过表1可以发现，对于GB 29940—2013《食品安全国家标准 食品添加剂 柠檬酸亚锡二钠》，3种大模型均能准确提取标准中文名称的有效语义单元，但语义理解和联想广度存在差异：DeepSeek-R1能识别出柠檬酸亚锡二钠在食品安全中的基本功能及其化学属性，得出柠檬酸亚锡二钠是食品防腐剂、隶属于亚锡化合物的结论并基于此输出中、英文关键词，所给结果与标准中文名称存在较强的语义关联性；相比之下，DeepSeek-V3的输出结果虽然在语义理解上较为准

表1 DeepSeek-R1、DeepSeek-V3、Spark-Max、LDA模型和jieba分词工具的运行结果实例

标准号与标准中文名称	模型或工具的类型	输入	输出
GB 29940—2013《食品安全国家标准 食品添加剂 柠檬酸亚锡二钠》	DeepSeek-R1	输入标准中文名称，并在提示词中要求大模型根据强制性国家标准的标准中文名称获得与之相关的中文关键词和英文关键词各5个	中文关键词：食品安全国家标准，食品添加剂，柠檬酸亚锡二钠，食品防腐剂，亚锡化合物 英文关键词：National food safety standards, Food additives, Disodium stannous citrate, Food preservatives, Tin compounds
	DeepSeek-V3		中文关键词：食品安全，国家标准，食品添加剂，柠檬酸亚锡二钠，食品标准 英文关键词：Food Safety, National Standard, Food Additive, Sodium Stannous Citrate, Food Regulation
	Spark-Max		以下是为您提供的中文关键词和英文关键词： ### 中文关键词 食品安全国家标准，食品添加剂，柠檬酸亚锡二钠，食品加工，质量检测 ### 英文关键词 National food safety standard,Food additive,Disodium stannous citric acid,Food processing,Quality inspection 解释：“食品安全国家标准”明确指出了这是关于食品方面的国家安全标准，所以“National food safety standard”与之对应；“食品添加剂”是该标准涉及的主要类别，“Food additive”是其英文表述；“柠檬酸亚锡二钠”是具体涉及的物质，其英文名称是“Disodium stannous citric acid”；“食品加工”体现了与该物质在食品生产环节中的关联，“Food processing”是相应英文；“质量检测”表明需要对该物质进行质量方面的检验，“Quality inspection”是对应的英文表达
	LDA	输入标准中文名称	亚锡，柠檬酸，添加剂，钠，食品，食品安全
	jieba	输入标准中文名称	食品安全，国家标准，食品，添加剂，柠檬酸，亚锡，二钠

确，但语义联想的广度不及DeepSeek-R1；Spark-Max能在准确提取标准中文名称有效语义单元的基础上进一步给出模型自身对关键词的理解，但其输出内容与标准中文名称的语义关联紧密性不如DeepSeek-R1。与大模型相比，LDA模型与jieba分词工具在语义单元识别的准确性上存在一定偏差，也无法提供语义联想和推理层面的有效结果。

上述模型和工具的人工评估结果显示，DeepSeek-R1仅在少数回答中表现出了幻觉现象和内容错误，不仅能准确识别标准中文名称的有效语义单元，也能根据标准中文名称的语义内容进行联想，获取与之相关的领域专业知识和标准文本信息，在部分情况下也可根据自身回答提供关键词结果的应用参考建议。相比之下，DeepSeek-V3所输出的关键词大多直接与标准中文名称相关，Spark-Max能根据标准中文名称进行一定的语义联想，但联想的广度和深度不及DeepSeek-R1。整体而言，大模型对于标准中文名称的语义理解准确性高于LDA模型和jieba分词工具，体现出明显的文本加工优势。

3　大模型赋能标准数字化的潜在风险与发展建议

3.1　大模型赋能标准数字化的潜在风险

3.1.1 数据治理风险

尽管大模型技术在语义理解、内容生成等方面体现出卓越的处理能力，但数据治理风险是大模型在标准数字化应用过程中无法回避的潜在问题。数据治理风险主要包括数据质量风险和数据安全风

险[22]。近来的研究显示，现阶段的大模型在应用过程中普遍存在时效性、稳定性、可解释性、可靠性等方面的不足[23]，其隐患主要来自数据合规、算法合规、隐私保护、幻觉问题等方面[4]，存在侵权、数据泄露、数据偏差等现象[24]。在初步探索中也发现了DeepSeek-R1的回答存在低比例的幻觉问题，在后续工作中将继续加以重视，提高数据结果的可信度。

3.1.2 版权风险

随着人工智能时代的到来，数据兼具数据资源与训练数据的双重价值，训练数据的版权信息披露已成为人工智能法治问题的热点[25]。标准是受著作权保护的技术性文本，大模型赋能标准数字化将引发一定的版权风险。标准数字化工作者应当重点关注数字时代二次创作的合理使用方式[26]，确保标准数据的使用合法合规。

3.2 大模型赋能标准数字化的发展建议

本研究基于大模型技术的整体应用现状，从标准数字化的发展需求入手，提出以下发展建议。

3.2.1 夯实标准语料基础

语料数据、算法和算力是人工智能的“三驾马车”。随着大模型发展逐渐步入模型的后训练时代，语料数据被视为决定模型性能上限的关键因素。高质量的语料库建设与应用已成为我国近年来人工智能领域的重要方向。对于标准数字化而言，不论是构建标准垂类大模型，还是调用基础大模型加以微调，都需要高质量标准语料库的支撑。现阶段的标准数字化工作在数据基础方面还存在较大的提升空间，后续工作应当重点聚焦标准语料库的建设，根据标准数字化任务的具体要求打造大模型所需的标准知识库，从而有助于大模型给出针对性的回答，降低其出现幻觉问题的可能性。

3.2.2 加强标准提示词工程建设

与以往的深度学习模型和传统机器学习模型不同，大模型需要提供提示词（Prompt）作为任务的输入文本或指令[27]。提示词通常是问句、上下文信息、指令说明等形式。提示词工程（Prompt Engineering）的质量与大模型的回答准确性密切相关。为了提高大模型赋能标准数字化的应用效果，应当制定契合特定标准数字化应用场景和大模型自身特性的提示词数据集，在标准领域形成可复制、可推广的提示词生成方法。

3.2.3 在实践中择优选取基准型

当前国内外大模型类型众多。尽管大模型在语义理解、文本生成等方面表现优异，但其与已有的深度学习和机器学习模型并非完全的对立关系。标准数字化工作者应当意识到模型“各司其职”在标准数字化工作中的重要性，从具体应用场景的实际需求出发，通过比较不同模型在特定任务中的表现结果，择优选取该应用场景的基准型。

3.2.4 构建基于大模型的标准智能体

近年来，基于大模型的智能体已被证实能提升知识服务模式的智能化程度[28]，能处理复杂的任务，在多个领域得到广泛应用。标准智能体也将成为大模型赋能标准数字化应用的重要载体。与其他模式相比，基于大模型的标准智能体以大模型为“大脑”。大模型通过掌握完成标准数字化应用任务所需的工具操作方式和领域专业知识，能够快速适应标准数字化应用场景的实际需求，灵活应对标准数字化应用需求的变化。

3.2.5 制定大模型回答审查方法

鉴于目前大模型的回答尚存在“幻觉问题”，而标准数字化工作普遍对结果精确性有较高的要求，标准数字化工作者应当针对大模型的大规模批量调用结果制定科学有效的审查方法，形成大模型回答的质量评估方法，以此确保大模型结果的精确性和可信度。

4 结 语

在人工智能时代，大模型赋能标准数字化应用已成为大势所趋。依靠出色的语义理解与文本生成能力[29]，大模型有望加快机器可读标准的构建，通过在标准智能比对、标准智能编写等多个标准数字化应用场景中发挥重要的作用，提升标准知识服务的供给能力，进而推动标准向数字化、网络化和智能化发展。

参考文献

[1] 标准数字化理论研究与发展趋势洞察[J].中国标准化,2025(3):14–15.

[2] 袁文静,方洛凡.标准对话：标准数字化的阶段性目标与实践[J].中国标准化,2024(3):6–29.

[3] 程云,陈国祥,陈寒竹,等.基于人工智能的标准数字化关键技术路径探索[J].信息技术与标准化,2022(10):60–67.

[4] 刘泽垣,王鹏江,宋晓斌,等.大语言模型的幻觉问题研究综述[J].软件学报,2025,36(3):1152–1185.

[5] 张泽华,柴豪.国内外大模型在情感分析中对比与应用策略[J].重庆工商大学学报(自然科学版),1–11[2025–03–04].http://kns.cnki.net/kcms/detail/50.1155.N.20241015.1411.004.html.

[6] 邓建鹏,赵治松.DeepSeek的破局与变局：论生成式人工智能的监管方向[J].新疆师范大学学报(哲学社会科学版),2025,46(4):99–108.

[7] 钟新龙,渠延增,王聪聪,等.国内外人工智能大模型发展研究[J].软件和集成电路,2024(1):80–92.

[8] 王立玺,吕千千,孔庆炜,等.大语言模型加快标准数字化发展进程实践与思考[J].信息技术与标准化,2024(8):32–37.

[9] 标准化工作导则　第1部分：标准化文件的结构和起草规则:GB/T 1.1—2020[S].

[10] 曹亚威,陈月艳,曹倩倩.基于DIKW模型的零售业数字化转型路径研究[J].生产力研究,2023(7):101–105.

[11] 刘曙.多模态文档知识库问答研究及应用[D].上海:华东师范大学,2023.

[12] 翟蓉.大模型赋能的智能问答FAQ语料库建设实践与思考:以国家图书馆为例[J].四川图书馆学报,2025(2)：80–87.

[13] 梁佳,张丽萍,闫盛,等.基于大语言模型的命名实体识别研究进展[J].计算机科学与探索,2024,18(10):2594–2615.

[14] 郑佳明,陈家宾,胡杰鑫,等.基于大模型和知识图谱的标准领域融合应用方法研究[J].中国标准化,2023(23):39–46.

[15] 金宝生,郭越,侯守立.金融业融合大模型多模态知识图谱的构建研究[J].中国科技论文在线精品论文,2024,17(4):420–424.

[16] 方思怡.标准知识图谱的技术路径与应用场景探讨[J].中国标准化,2023(11):49–55.

[17] 杨倩,林鹤.科技情报基础服务知识融合流程自动化框架研究[J].信息与管理研究,2024,9(4):65–72.

[18] 刘炜,刘倩倩.生成式人工智能十大趋势与公共文化机构的应对策略[J].图书馆建设,2025(1):4–14.

[19] 黄波.人工智能解决标准版权问题刍议[J].中国标准化,2024(20):16–17.

[20] 王轶群,陈成鑫.基于LDA的欧盟信息迷雾政策报告主题挖掘与治理研究[J].图书情报工作,2025,69(5)：94–106.

[21] 孙亚洲,李晓松.采用LDA模型的美国《芯片与科学法案》主题挖掘及分析[J].信息工程大学学报,2025,26(1):120–126.

[22] 张欣.生成式人工智能的数据风险与治理路径[J].法律科学(西北政法大学学报),2023,41(5):42–54.

[23] 肖建力,邱雪,张扬,等.交通大模型综述[J].交通运输工程学报,2025,25(1):8–28.

[24] 黄锫.人工智能大模型训练数据的风险类型与法律规制[J].政法论丛,2025(1):23–37.

[25] 李安.人工智能训练数据的版权信息披露：理论基础与制度安排[J].比较法研究,2024(5):136–152.

[26] 铁婧可.数字时代二次创作的合理使用:宪法视角下著作权法的规范性重构[J/OL].海南大学学报(人文社会科学版),1–9[2025–03–04].https://doi.org/10.15886/j.cnki.hnus.202412.0199.

[27] 孙晨伟,侯俊利,刘祥根,等.面向工程图纸理解的大语言模型提示生成方法[J].计算机应用,2025,45(3):801–807.

[28] 孙蒙鸽,付芸,刘细文.智能体赋能科研知识服务的路径解析[J].智库理论与实践,2025,10(1):3–18.

[29] 秦小林,古徐,李弟诚,等.大语言模型综述与展望[J].计算机应用,2025,45(3):685–696.