

doi: 10.11835/j.issn.1000-582X.2025.10.009

引用格式:李桂仔,魏莉,尹彦君,等.自然语言处理技术在油气勘探开发的应用前景——以 ChatGPT 为例[J].

重庆大学学报,2025,48(10): 95-109.



自然语言处理技术在油气勘探开发的应用前景 ——以 ChatGPT 为例

李桂仔,魏莉,尹彦君,冯高城,王凤刚,张震

(中海油能源发展股份有限公司 工程技术分公司,天津 300450)

摘要:随着人工智能的快速发展,自然语言处理技术的应用逐渐扩展至各领域,如金融、医疗、教育和电子商务等,它为多元业务领域提供了更高效、智能的解决方案。研究重点讨论了 ChatGPT 在油气勘探开发领域的具体应用场景、面临的挑战以及未来发展的可能性。利用 Python 调用 ChatGPT 的 API(应用程序接口),通过实例解析,说明 ChatGPT 在油气勘探开发领域,在诸如信息检索、决策支持、客户服务等方面具有显著的优势。这些优势表现为提高工作效率、优化决策制定、提升客户服务和沟通质量,以及创新的方法解决问题。ChatGPT 在编程方面的强大优势进一步提升人工智能在油气勘探开发领域中的应用效率。ChatGPT 支持利用专业知识和数据对模型进行微调,构建专属油气勘探开发智能专家,数据的数量和质量决定模型精准度和专业性。同时,需要应对挑战,如回答的真实性、数据质量,模型的准确性以及数据安全等问题。未来,ChatGPT 的发展趋势可能表现为增强理解能力、创新能力、智能交互能力,而且 ChatGPT 可以与数据湖进行高效融合,实现在数据查询、故障预测、报告自动生成、培训以及自动化工作流程等方面的应用。

关键词:ChatGPT;自然语言处理;人工智能;油气勘探开发;数据湖

中图分类号:P631.8

文献标志码:A

文章编号:1000-582X(2025)10-095-15

The prospects of applying natural language processing technology in oil and gas exploration and development: A case study of ChatGPT

LI Guizai, WEI Li, YIN Yanjun, FENG Gaocheng, WANG Fenggang, ZHANG Zhen

(CNOOC EnerTech-Drilling & Production Co., Ltd., Tianjin 300450, P. R. China)

Abstract: With the rapid progress of artificial intelligence, the application of natural language processing(NLP) has expanded across various fields, such as finance, medical care, education, and e-commerce, offering efficient and intelligent solutions for diverse business arenas. This study primarily discusses the specific application scenarios, challenges, and potential future developments of ChatGPT in oil and gas exploration and development. Using Python to access the ChatGPT API, illustrative examples demonstrate its strengths in information retrieval, decision support, and customer service within this industry. These advantages translate into improved operational

收稿日期:2023-08-12

基金项目:海油发展重大专项(HFKJ-ZX-GJ-2023-02)。

Supported by Major Special Program of Cenertech (HFKJ-ZX-GJ-2023-02).

作者简介:李桂仔(1988—),男,硕士研究生,工程师,主要从事油气勘探开发大数据及人工智能方向研究,(E-mail)

ligz3@cnooc.com.cn。

efficiency, optimized decision-making, enhanced customer service and communication, and innovative problem-solving methods. Additionally, ChatGPT's strong programming capabilities further improve the efficiency of AI applications in this field. Fine-tuning ChatGPT with domain-specific knowledge and data enables the development of dedicated intelligent systems for oil and gas operations, where the quantity and quality of data determine the model's accuracy and expertise. Nevertheless, challenges remain, such as response reliability, data quality, model accuracy, and data security. Future trends are expected to include enhanced comprehension, improved creativity, and greater interactivity. Furthermore, ChatGPT can be integrated with data lakes to support data querying, fault prediction, automated report generation, training, and workflow automation. With continuous upgrades and user-driving optimization, NLP technologies such as ChatGPT are anticipated to play an increasingly critical role in the oil and gas sector.

Keywords: ChatGPT; natural language processing; artificial intelligence; oil and gas exploration and development; data lakes

自然语言处理(natural language processing, NLP)与人工智能(artificial intelligence, AI)2大领域密切相关,并处在快速发展阶段。NLP主要是借助机器学习中的深度学习技术(deep learning, DL)和神经网络技术(neural network, NN),使得计算机得以解读和产生人类语言(如图1所示)。NLP的应用涵盖了语音识别、自然语言理解、自然语言生成及机器翻译等多个领域。其目标是让计算机可以全面处理并应用人类语言,这包括语义解析、语法结构分析、语言生成和机器翻译等各个方面^[1-2]。

近年来,预训练语言模型(pre-trained language model, PLM)发展迅速,将自然语言处理推到了一个全新的发展阶段。比较有代表性的2个系列模型分别是BERT模型(bidirectional encoder representations from transformers)和GPT模型(generative pre-trained transformer)。BERT模型是由谷歌在2018年研究发布的一款NLP预训练模型。BERT模型为各种NLP任务带来了性能的显著提升。它的革新性在于使用Transformer的编码器作为基础架构,并通过双向和全面的词语上下文理解优化模型。GPT模型是由OpenAI于2018年发布,它使用深度学习生成类似人类的文本。经过多年的迭代升级,OpenAI于2022年11月发布了ChatGPT,是GPT-3.5架构的主力模型^[3]。2023年3月,OpenAI正式推出GPT-4,它的特点在于训练数量更大,能够支持多元输出输入形式,且在专业领域的学习能力更强^[4]。研究的相关实例主要由GPT-4生成的。

ChatGPT是一种AI工程化的成功实践,使普通人可以应用最新的自然语言生成与对话技术,完成很多之前需要较高技术门槛才能完成的工作^[5]。研究通过实例解析重点讨论了ChatGPT在油气勘探开发领域的具体应用场景、面临的挑战及未来发展的可能性。

1 ChatGPT 基本原理

ChatGPT的原理是基于一种强大的神经网络架构(Transformer模型),通过学习大规模文本数据中的模式和规律理解和生成自然语言。模型首先学习输入序列中单词间的关系,获取上下文信息。通过自注意力机制,关注到输入序列中不同位置的相关性,类似于写作时关注文中不同部分的逻辑关系。然后,模型使用这些上下文信息预测下一个单词,考虑前面已生成的单词,选择最适合的下一个单词。这个预测过程是逐步进行的,每次只生成1个单词,将其加入到已生成的内容中,成为下一步预测的一部分。通过不断重复这个

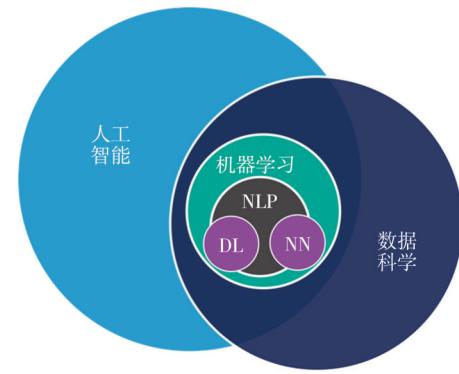


图1 自然语言处理(NLP)的技术定位
Fig.1 Technical positioning of NLP

过程,模型一步步生成全新文本,就像写作时一步步构建出完整的文章一样,如图 2 所示。总之,自回归 Transformer 模型以逐步预测的方式生成文本,利用前面已生成的内容指导后续生成过程,产生连贯、上下文相关的文本。

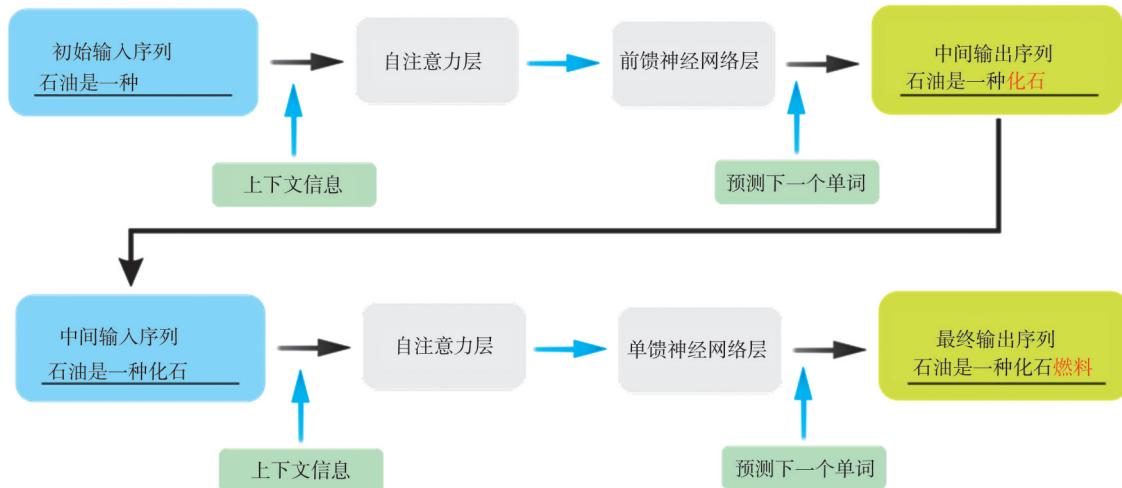


图 2 Transformer 模型概念图

Fig.2 Conceptual diagram of the transformer model

ChatGPT 利用这种自回归 Transformer 架构和大规模的文本数据进行训练,模型具有多个层次的自注意力机制,使 ChatGPT 在处理长篇文本时表现出色^[6]。训练 ChatGPT 的具体过程可以分为 3 个阶段:监督学习、训练奖励模型和强化学习(如图 3 所示)。

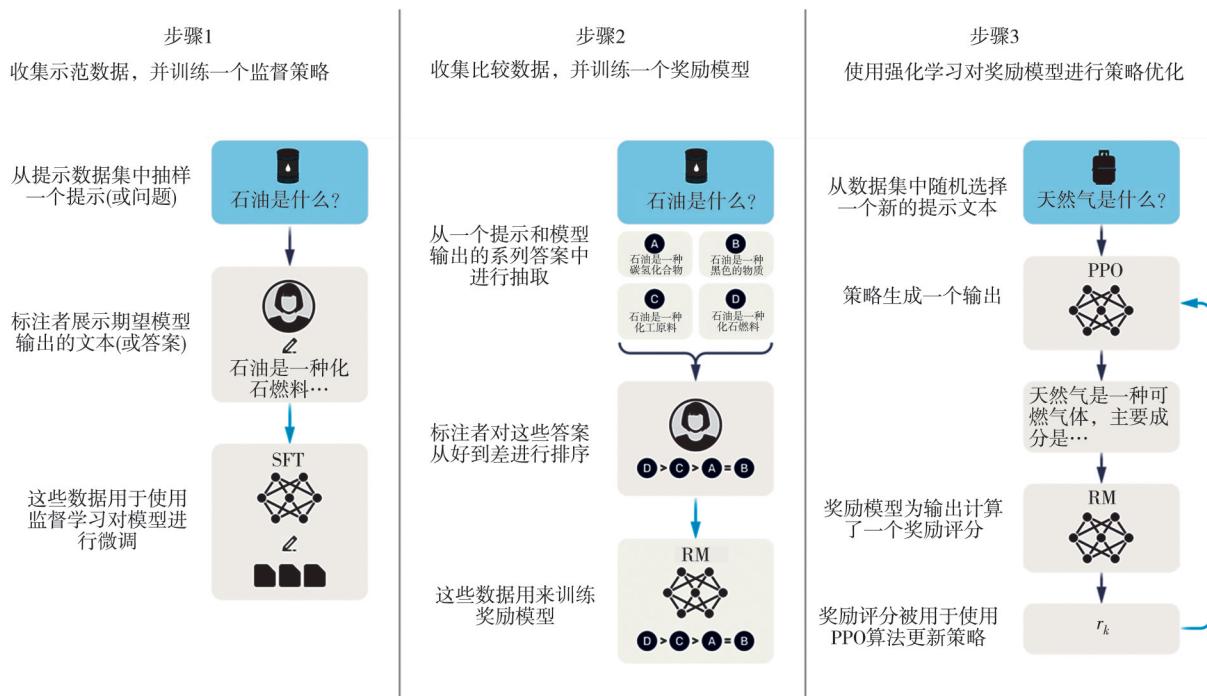


图 3 训练 ChatGPT 模型的 3 个步骤

Fig.3 Three steps for training the ChatGPT model

在监督学习阶段,ChatGPT 使用大量互联网文本数据进行训练,通过学习预测下一个字词的方式,逐步建立对语言的理解和生成能力。这个阶段的目标是使 ChatGPT 能够掌握语法、词汇和常见的语义关系。

在训练奖励模型阶段,明确奖励目标,构建奖励函数评估和奖励模型回答的质量,将奖励信号引入训练数据,用于指导模型优化。通过标记或调整数据样本,指示正确回答并作为奖励标记,使用带有奖励信号的数据重新训练模型,不断迭代优化,根据评估结果调整奖励函数和训练策略。例如,模型可能会学习到当用户提出特定问题时,某个特定的回答更被人喜欢。这个过程需要时间、计算资源和有效的奖励设计,提高模型的回答质量和准确性。在强化学习阶段,人类评估员会从数据集中随机选择一个新的提示文本评估模型给予反馈,这是采用强化学习中的一种特性,即试错法。人类评估员会对模型的回答进行评估和反馈,模型再根据反馈进行不断学习和调整。例如,模型可能会学习到,当用户在谈论某个敏感话题时,模型采取更细致和体贴的语言获得更高评分,而对用户的观点进行判断或攻击则会得到更低评分。这个阶段的目标是优化对话质量,补充和优化模型在特定任务和应用中的表现。

ChatGPT 的关键组成部分是生成模块和推理模块。生成模块负责根据输入生成有意义的回答,而推理模块用于理解上下文并生成连贯的响应。

总之,ChatGPT 是一种基于 Transformer 模型的语言生成工具,通过对大量数据进行预训练模拟人类对话,提供自然流畅的回答。ChatGPT 是根据概率结合轻微的随机因素预测单词,生成句子。尽管它有一定的局限性,但在许多应用场景中,ChatGPT 能够成为一个有用的工具,并为用户提供便利和支持^[7-9]。

2 ChatGPT 在油气勘探开发中的应用场景分析

为提升工作效率、改善策略决策和优化客户服务,油气行业不断探寻智能化的解决方案,适应复杂的挑战和不断变化的环境,ChatGPT 在油气行业的应用具有深厚的动力基础和独特优势。

2.1 提供高效的信息检索和知识管理能力

ChatGPT 能为油气勘探开发领域提升信息检索和知识管理的效能。此领域拥有丰富的数据和文献资源,包括地质勘探数据、钻井数据、油藏动态数据和工艺流程等。传统的搜索和阅读方式通常耗时费力,而 ChatGPT 则能运用自然语言处理和机器学习的技术,从大量数据中快速提取及归纳相关信息,为工程师、研究人员及决策者提供知识支持。例如,若想了解目前针对常规稠油油藏较有效的开发方式,通常需要通过搜索和阅读相关书籍或文献,现在能先使用 ChatGPT 进行初步查询(如图 4 所示)。若对某种方法产生兴趣,可深入提问后,有针对性地查阅和研究具体文献。

```

prompt = f"""
针对常规稠油油藏有哪些比较有效的开发方式,并给出实例,并给出代表文献。
"""

response = get_completion(prompt)
print(response)

```

常规稠油油藏的开发方式主要包括:蒸汽驱、聚合物驱和化学驱等方法。

1. 蒸汽驱(steam flooding): 通过注入高温高压的蒸汽,降低油藏黏度,提高原油流动性,增加采收率。实例:以加拿大阿尔伯塔省的阿萨巴斯卡原油沉积层为例,该区域的常规稠油油藏采用蒸汽驱进行开发。

代表文献: Dusseault M B, Gray M R, & Scott J V. Heavy oil and oil sand petroleum systems in alberta and beyond[J]. AAPG Memoir, 2013, 110: 867-898.

2. 聚合物驱(polymer flooding): 通过注入高分子量的聚合物溶液,降低油藏渗透率,增加驱替效果,提高采收率。实例:以委内瑞拉的奥里诺科油田为例,该油田采用聚合物驱进行常规稠油油藏的开发。

代表文献: González G, Fu Y. Polymer-improved oil recovery in a low-permeability sandstone reservoir in lake maracai bo, Venezuela [J]. Journal of Petroleum science and Engineering, 2010, 70(3-4), 272-281.

3. 化学驱(chemical flooding): 通过注入化学物质如表面活性剂、聚合物或聚合物与表面活性剂的复合物,改变岩石表面性质,提高油分离能力与油藏采收率。实例:以美国加利福尼亚州的贝克尔油田为例,该油田采用化学驱进行常规稠油油藏的开发。

代表文献: Lake L W, Russell J, Bill R. Fundamentals of Enhanced Oil Recovery[D]. Texas:Society of Petroleum Engineers, 2014.

请注意,根据实际情况和油藏特性,选择合适的开发方式十分重要,具体开发方式需要根据实地勘查和工程设计来确定。

图 4 信息检索示例

Fig.4 An example of information retrieval

其次,ChatGPT 能作为强大的决策辅助工具。油气行业常面临复杂的决策问题,包括投资决策、生产优化和风险管理等。ChatGPT 能基于大量行业数据、历史案例和模型预测结果,提供对决策的评估、对比和优化建议(如图 5 所示)。这种实时的智能决策辅助可提升决策的精准性和效率,同时降低潜在风险和损失。

```

prompt = f"""
针对潜山油藏的勘探需要注意哪些方面的不确定因素, 目前有哪些勘探实例, 请从成功和失利两个方面列举。
"""

response = get_completion(prompt)
print(response)

```

潜山油藏是在地层的深部,且经常被盖层的地层完全覆盖,导致对其的勘探有很多不确定性。以下是一些主要的不确定因素:

1. 地震数据解释的准确性:在前期勘探阶段,勘探队伍主要依赖地震数据来理解地下的地质构造。由于地震数据的质量和解释准确性这可能导致潜山油藏的位置、大小等关键特性的大量不确定性。
2. 油源岩和储层的质量:如果没有足够的优质油源岩或者储层,潜山油藏可能无法形成或者无法产出经济上有效的油气。由于对地下深部地层信息的缺乏,这些因素的存在和质量充满了不确定性。
3. 地质时间和结构变迁:油藏形成是一个需要长时间的过程,包括了多个地质时代的沉积、成岩、构造变化等过程。这些时间和过程变化可能对油气的形成和储藏产生很大影响。

由于商业秘密和相关权益的保护,没有很多公开可查的具体潜山油藏勘探实例。但大体来说:

1. 成功的实例:例如,某些海底潜山油藏在经过多次地震测层和精准探井定位后,确认了优质的储层和油源岩的存在,经济上成功地开发出大量油气。如巴西的盐下油藏。
2. 失败的实例:在某些情况下,虽然地震数据显示了可能存在的潜山油藏,但在钻井探测后,发现缺乏优质油源岩或者储层,或者油藏规模远小于预期,导致经济上的失败。例如一些深水海域的潜山油藏开发。

图 5 决策辅助分析示例

Fig.5 An example of decision support analysis

此外,ChatGPT 还可以提供更智能的客户服务和沟通。在油气行业中,企业需要与客户、供应商和合作伙伴保持频繁的沟通和协作。ChatGPT 可以作为一个虚拟助手,通过自然语言交互与对话,回答客户关于产品规格、交付状态、技术支持等方面问题,解决客户疑问并提供即时的帮助。另外,可以用 ChatGPT 来提取客户的评论中识别客户表达的情感、快速提取重要信息,优化对话内容(如图 6 所示)。这种智能化的客户服务将为企业提供更好的客户体验,提高客户满意度和忠诚度。

```

review_zh = """
我需要一台方便好用的岩心取样仪, 这台取样仪器有额外的检测功能, 价格也不算太高。 \
我很快就收到了它。在运输过程中, 我们的电线压坏了, 但是公司很乐意寄送了一个新的。 \
几天后就收到了。这款岩心取样仪很容易组装。我发现少了一个零件, 于是联系了他们的客服, 他们很快就给我寄来了缺失的零件! \
在我看来, Rockstar 是一家非常关心顾客和产品的优秀公司!
"""

prompt = f"""
识别以下评化的作者表达的情感。包含不超过5个项目。将答案格式化以逗分隔的单词列表。
"""

```

评论文本: ``` {review_zh} ```
```

```

response = get_completion(prompt)
print(response)

```

满意, 感激, 认可, 愉快, 正面

```

prompt = f"""
从评论文本中识别以下项目:
- 评论者购买的物品
- 制造该物品的公司
"""

```

评论文本: ``` {review\_zh} ```  
```

```

response = get_completion(prompt)
print(response)

```

- 评论者购买的物品: 岩心取样仪
- 制造该物品的公司: Rockstar

图 6 评论情感识别和信息提取示例

Fig.6 An example of sentiment analysis and information extraction

ChatGPT 以强大的语言理解和生成能力显著提升了各行各业的办公效率^[10-13]。油气勘探开发是一个复杂且技术密集的过程,涉及大量数据的采集、处理和分析,这些过程产生的结果需要以报告或汇总的形式进行提交和交流,这部分工作是相当耗时且需要高精度的。通过深度学习和自然语言处理技术,ChatGPT 对大量的技术数据进行理解和总结,将其中的关键信息提炼出来(如图 7 所示),根据给定的报告模板,将这些总结和关键信息以清晰、准确的方式呈现出来。

```
text = f"""
A油田含油层段位于新近系馆陶组,含油层段集中,主体区内部纵向含油层段约300 m;纵向上共划分为8个油组12个小层,小层数目多且单层厚度薄,约70%的单层厚度小于3 m。储层岩性为含砾中粗砂岩,岩屑长石砂岩,储层物性较好,馆陶组储层为高孔高渗储层。属于浅水辫状河三角洲沉积,油藏地层压力系数为1.0,地层压力梯度为0.95 MPa/100 m,地层温度梯度为2.9 °C/100 m,属正常温压系统;原油密度较大、粘度较高、含蜡量低、含硫酸盐低及凝固点低,属典型重质油,具多套油水系统的构造层状油藏。
A油田采取整体部署、开辟试验区、分阶段实施的开发策略,其中先导生产试验区(I期开发)于1988年10月底投产,II期开发于2000年投产。伴随开发的不断持续,油田先后开展综合调整方案研究、低产低效井治理、调整井研究进行挖潜。截至2008年10月,油田日产油水平为1 234 t,动用储量采出程度为20.5%,综合含水80%,采油速度0.88%。
prompt = f"""
对以下用3个反引号括起来的文本进行归纳总结,提炼重要信息,从含油层段、储层物性、温压系统、开发策略及产量情况5个方面
分条列出,每条不超过20个字。
```
{text}
```
response = get_completion(prompt)
print(response)
"""

1.含油层段: 新近系馆陶组,纵向含油层段300 m,共8个油组12个小层。
2.储层物性: 含砾中粗砂岩,岩屑长石砂岩,高孔高渗储层。
3.温压系统: 地层压力系数1.0,压力梯度0.95 MPa/100 m,温度梯度2.9 °C/100 m。
4.开发策略: 整体部署,开辟试验区,分阶段实施。
5.产量情况: 日产油水平1 234 t,采出程度20.5%,综合含水80%。

```

图 7 提炼技术报告关键信息示例

Fig.7 An example of key information extraction from the technology report

2.2 提高人工智能在油气勘探开发应用的广度和效率

人工智能和机器学习在油气勘探开发中发挥了关键角色。主要用于勘探潜力靶区预测、地震资料解释、钻井数据分析、智能注采及评估和预测油井产量等方面。例如,借助神经网络及深度学习模型处理大量复杂的地震数据,可以达到更智能化、精确地解释断层和储层预测的目标^[14-16]。机器学习还被用于优化钻井路径,减少风险和成本,通过分析多维度地质、工程数据生成最优路径^[17-18]。此外,通过分析历史油水井动静态数据,机器学习可预测未来的产量^[19-21]。总的来说,人工智能和机器学习的应用在油气勘探开发领域具有巨大潜力,提高油气勘探开发的效率和效益。然而,人工智能与油气勘探开发专业的融合程度尚需要提高^[22-28]。原因是数据科学、人工智能等计算机科学领域与勘探开发领域之间存在的专业壁垒。例如,一位优秀的数据工程师或算法工程师在对油气勘探开发的专业理解上可能不足,导致无法有效地应用其技能;反之,勘探开发专家在接触人工智能解决方案时,可能对数据科学、算法和编程等领域知识的掌握不足,导致无法有效利用人工智能工具。

在提高人工智能在油气勘探开发应用的广度和效率方面,ChatGPT 强大的自然语言处理能力,使其在提高人工智能的应用效率等方面发挥巨大作用,将人工智能高效地应用到油气勘探开发的每一个环节。

2.2.1 创新人工智能解决问题方式

ChatGPT 具备重塑人工智能问题解决框架的巨大潜力。通过深度融合人工智能和机器学习技术与油气勘探开发领域的专业知识,ChatGPT 催生了一种创新且高效的工作模式,将大数据、智能算法、软件工程的前沿技术,以及地球物理、地质学、测井和油藏工程等油气勘探和开发领域的专业知识,创新地结合在一起。例如,一名没有人工智能应用经验的地质实验工程师想将医学领域中相当常用的基于机器学习的图像识别技术^[29-30]应用到岩心描述和薄片鉴定等研究工作中。他可以向 ChatGPT 提问,得到快速准确的解决方案(如图 8 所示)。通过 ChatGPT 的引导,其他领域的技术和方法也有可能在油气勘探开发领域找到合适的应用。

```

prompt = f"""
医学领域中常用的基于机器学习的图像识别技术,如何应用到岩心描述和薄片鉴定,请提供准确的解决方案。
"""

response = get_completion(prompt)
print(response)

```

在岩心描述和薄片鉴定中,可以采用以下基于机器学习的图像识别技术:

- 1.数据收集与标注:收集大量的岩心描述和薄片鉴定的图像数据,并对其进行详细的人工标注,建立训练和测试数据集。
- 2.图像预处理:对图像进行预处理操作,如图像增强、去噪、尺寸调整等,优化图像质量,提取有用信息。
- 3.特征提取与选择:使用计算机视觉技术,提取图像中的重要特征。对岩心描述可以提取颜色、纹理、形态等特征,而对薄片鉴定可以提取晶体结构、矿物组分等特征。同时,根据问题需要选择适当的特征。
- 4.模型训练与优化:利用机器学习算法,如卷积神经网络(convolutional neural networks, CNN)、支持向量机(support vector machines, SVM)等,基于已标注的数据集进行模型训练。通过优化算法和调参技术,提高模型的准确度和泛化能力。
- 5.模型评估与应用:使用独立测试数据集对训练好的模型进行评估,并根据评估结果对模型进行改进。最终,将训练好的模型应用于岩心描述和薄片鉴定的实际应用中,实现图像的自动识别和分类。

需要注意的是,实际应用中可能会面临数据质量、样本不均衡、模型的可解释性等挑战,针对具体问题需要做相应的优化和调整。同时,合理利用人工智能技术与专业领域知识相结合,能够进一步提高岩心描述和薄片鉴定的准确性和效率。

图 8 ChatGPT 提供机器学习解决方案示例

Fig.8 An example of the machine learning solution provided by ChatGPT

2.2.2 提高人工智能的应用效率

ChatGPT 能够凭借其强大的编程能力以协助油气勘探开发专业人员高效率地应用人工智能解决问题。例如,问题背景:A 油田开发调整阶段需要对储层分布进行精细研究,常规地震属性、地震反演方法无法精细预测储层的分布。目前有该油田的地震资料(时间域)、井数据(测井资料、解释结论)、地质模型等。经过初步调研,能够利用多地震属性机器学习的方法提高储层预测的精准度,但不知如何开展相关工作。解决方案:先将问题描述给 ChatGPT,让它给出解决方案(如图 9 所示),然后参照该方案的步骤开展工作。

```

prompt = f"""
A油田开发调整阶段需要对储层分布进行精细研究,常规地震属性、地震反演方法无法精细预测储层的分布。
目前有该油田的地震资料(时间域)、井数据(测井资料、解释结构)、地质模型等。经过初步调研。
能够利用多地震属性机器学习的方法提高储层预测的精准度,但不知如何开展相关工作。
请给出方案。
"""

response = get_completion(prompt)
print(response)

```

在基于多地震属性的机器学习方法提高储层预测精度的问题上,你可以按照以下步骤进行:

- 1.**数据整理和预处理**:首先,你需要处理和清洗收集的井数据和地震数据。这包括缺失值的处理、异常值的识别和处理,以及必要的数据转换工作。
- 2.**数据集合并**:将已经清洗和处理过的地震属性和井数据进行合并,形成一个完整的数据集。
- 3.**特征工程**:对地震数据和井数据进行特征工程,这可以包括提取新的特征、对特征进行选择和转换等。地质模型的产生和处理也可在此阶段进行。
- 4.**模型选择和训练**:根据问题的特性决定使用何种机器学习模型。可能的选择包括决策树、随机森林、支持向量机、神经网络等。使用训练集进行模型训练。
- 5.**模型评估和优化**:使用测试集对模型进行评估,根据模型的表现进行相关的调整和优化。
- 6.**模型应用**:将优化后的模型应用于新的或未知的数据,预测储层的分布。
- 7.**结果分析和反馈**:分析预测结果,比较预测结果与实际发现的储层分布之间的差异,进一步优化模型。

这个方案为一个大体的流程,具体的执行细节需要根据实际的情况和数据进行调整。例如,可能需要考虑到地震属性和储层之间的复杂关系,因此需要选择能够处理复杂非线性问题的模型来进行预测。另外,不同的特征可能需要进行不同的预处理和转换操作。建议你可能需要一个跨学科的团队。包括地震学家、石油工程师和数据科学家,最大程度地利用这些数据和方法。

图 9 ChatGPT 所提供的多地震属性机器学习储层预测工作流程

Fig.9 An example of the machine learning solution provided by ChatGPT

数据准备和预处理:这一步包括测井数据、地震属性的重采样,可以在建模软件中完成。但是数据预处理所涉及的数据体较大,常规数据处理软件很难对其处理,所以需要通过 Python 才能完成,这个过程对从事

开发地质工作者来说有一定的挑战性。因此,继续通过 ChatGPT 来解决该问题。首先是数据格式化。导出的数据为单列数据,需要先转换为多列数据,且需要进行删除异常值等操作。格式化以后对数据进行训练、预测,然后再将数据体转换为初始格式,最后导入地质模型。图 10 为根据 ChatGPT 给出的代码调整后的结果。

```

import pandas as pd
import numpy as np

# 该文件为petrel中导出的属性数据, 格式为Gslib properties
with open (r"C:\Users\admin\Desktop\test", "r") as file:
    next (file) # Skip first line
    n = int (next (file). strip ()) # Get n value from second line
    # 读取列名
    with open (r"C:\Users\admin\Desktop\test", "r") as file:
        for _ in range (2): next (file)
        column_names = [next (file). strip () for _ in range (n)]
    # 读取数据并转换
    df = pd.read_csv (r"C:\Users\admin\Desktop\test", skiprows=n+2, delimiter="\s+",
                      header=None, skipinitialspace=True, usecols=range (n))
    df. columns =column_names # 将列名添加到df中
    df = to_csv (r"C:\Users\admin\Desktop\test_transformed.csv", index=False)
    # 删除标签列中的空值(-99)
    df = df[df. iloc[:, 3] != -99]

```

图 10 数据预处理代码

Fig.10 Data preprocessing code

模型选择、训练及预测:数据格式化完成后,优先选择 Xgboost 回归模型进行训练,对模型参数进行寻优(如图 11 所示)。将调整好的模型应用于待预测数据进行预测,得到预测成果。

```

# 将数据分为特征和标签
X = df. iloc[:, 4:] # 特征从第5列开始
y = df. iloc[:, 3:] # 标签是第4列
# 划分训练集和测试集
X_train, X_text, y_train, y_text = TTS (X, y, test_size=0.25)
# 初始化XGBoost回归模型
model = xgb. XGBRegressor (booster='gbtree', n_estimators=1000,
                           gamma=0, min_child_weight=1,
                           learning_rate=0.1, subsample=0.8,
                           base_score=0.5, max_delta_step=0,
                           reg_alpha=0.05, scale_pos_weight=1)

# 训练模型
model.fit (X_train, y_train)
# 预测
y_pred = model. predict (X_text)
# 计算MAE, MSE和R2值
mae = mean_absolute_error (y_text, y_pred)
mae = mean_squared_error (y_text, y_pred)
r2 = r2_score (y_text, y_pred)
print (' Mean Absolute Error: ', mae)
print (' Mean Squared Error: ', mae)
print (' R2 Score: ', r2)

```

图 11 特征选择及模型训练代码

Fig.11 Feature selection and model training code

成果导出及展示:将预测成果再通过数据处理后导入地质模型并进行展示,对比验证效果。

2.3 ChatGPT 在油气勘探开发领域的独特应用

ChatGPT 不仅在上述方面能够发挥重要作用,在油气勘探开发领域具有其独特应用。ChatGPT 目前支持用户使用私有专业领域的资料对模型进行微调,基于自己特有的业务用例和知识模型,构建专属的 ChatGPT 专家。大语言模型的微调(Fine-tuning)是一种在预训练模型的基础上对特定行业的专业知识进行深度训练的方法,其最终的效果是使模型可以准确回答特定行业的专业问题。

ChatGPT 能够进行丰富的文本分析和理解,这对于沉积环境的识别具有很大应用潜力。沉积环境研究是地质学的一个重要领域,其相关信息往往包含在大量的文本数据中,如学科专业书籍、地质调查报告、地层描述资料等。然而,这些信息的提取和整理需要专业的地质知识和大量的人工工作。此时,如果可以训练 ChatGPT 学习和理解这些专业信息,将极大减轻地质工作者的工作负担,并且可能发现被忽视的潜在信息,为沉积环境的识别提供更客观、全面的数据支持。笔者对 OpenAI 推荐的 GPT-3.5-turbo-0613 模型进行微调训练,训练了一个关于沉积学的 AI 专家(如表 1 所示)。

表 1 模型微调训练步骤、代码、及要求和注意事项

Table 1 Steps, code, data sources, and requirements for fine-tuning the model

步骤	代码	要求和注意事项
1. 准备数据	<pre> { "messages": [{"role": "system", "content": "我是一名沉积学 AI 专家"}, {"role": "user", "content": "什么是沉积相?"}, {"role": "assistant", "content": "沉积相是指某种沉积环境的沉积产物"}] } </pre>	按照标准格式创建一系列训练数据,OpenAI 建议每个训练集至少要有 10 条数据。(本次使用 385 条数据)
2. 上传数据	<pre> openai.File.create(file = open("mydata.jsonl", "rb"), purpose = "fine-tune") </pre>	通过调用 OpenAI SDK 将数据上传。
3. 创建微调作业	<pre> import os import openai openai.api_key = os.getenv("OPENAI_API_KEY") openai.FineTuningJob.create(training_file = "file", model = "GPT-3.5-turbo") </pre>	提交后,微调任务进入 OpenAI 的训练队列依次进行,训练时间取决于数据集大小(本次训练时长约 6 h)

通过使用相同的提示词(prompt)对模型进行提问,可以看出未经微调的模型所回答的内容(图 12(a))存在较明显的错误(①层介形虫化石常见,说明沉积环境为海洋环境,因此属于水下沉积),而经过微调训练的模型答案更加精确和专业(图 12(b))。

ChatGPT 可以作为策略推断和优化的工具。在油气勘探开发过程中,怎样形成一个有效的策略是一个重要的问题。ChatGPT 是一个能够进行复杂预测和推理的模型,可以输入历史勘探数据,聚焦于勘探环境、各类地质因素和预期产值等重点信息,最后输出可能的最优策略。这不仅缩短了策略分析和生成的时间,也提供了更加高效准确的决策参考。

ChatGPT 在优化增产措施方面同样有其独到之处。通过输入大量的现有案例和最新科研成果,ChatGPT 能够输出面向具体情况的解决方案或者增产策略。例如,通过 ChatGPT 的模型预测,不同类型的低产低效井能够匹配最优的增产措施,达到增产效果。

总之,在油气勘探开发的各专业都可以训练专属的 ChatGPT 模型,数据的数量和质量决定了模型的精确度与专业性。

```

prompt = f"""
某野外露头剖面,自下而上为①~④层,根据各层的特征,判断哪些层沉积于水下,哪些层沉积下陆上。
④浅灰色细砂岩,对称波痕发育,厚4 m。
③紫红色泥岩,具有块状层理,厚5 m。
②灰黑色泥岩,植物根、茎、叶化石常见,厚3 m。
①灰绿色泥岩,介形虫化石常见,厚6 m。
"""

"""
response = get_completion(prompt)
print(response)

```

根据给出的信息,可以判断以下层是水下沉积的:

④浅灰色细砂岩,对称波痕发育,厚4 m。

根据波痕的特征,可以推断这层是在水中形成的沉积物。

③紫红色泥岩,具有块状层理,厚5 m。
②灰黑色泥岩,植物根、茎、叶化石常见,厚3 m。
①灰绿色泥岩,介形虫化石常见,厚6 m。

这些层的特征,如块状层理和植物化石的存在,表明它们是在陆地上沉积的。

(a) 未微调模型问答

```

prompt = f"""
某野外露头剖面,自下而上为①~④层,根据各层的特征,判断哪些层沉积于水下,哪些层沉积下陆上。
④浅灰色细砂岩,对称波痕发育,厚4 m。
③紫红色泥岩,具有块状层理,厚5 m。
②灰黑色泥岩,植物根、茎、叶化石常见,厚3 m。
①灰绿色泥岩,介形虫化石常见,厚6 m。
"""

"""
response = get_completion(prompt)
print(response)

```

根据地层的特征,可以推测每一层的沉积环境:

①灰绿色泥岩,介形虫化石常见,厚6 m。这层地层的特征显示,植物存在但并不丰富,而且介形虫是古代海洋生物,所以这一层很可能是在水下环境中形成的。

②灰黑色泥岩,植物根、茎、叶化石常见,厚3 m。这一层有大量植物化石存在,说明这是在一个陆地环境中形成的,展示了一个沼泽或湿地环境。

③紫红色泥岩,具有块状层理,厚5 m。紫红色一般指示氧化环境,块状层通常会在干燥的陆地环境中形成,所以这一层应该是在陆地环境中形成。

④浅灰色细砂岩,对称波痕发育,厚4 m。对称的波痕通常在浅海或湖泊环境下形成,表明水流两边来回运动造成的,所以这一层地层可能是在水下形成的。

综合以上,可以判断①和④层是在水下环境中形成,则②和③层是在陆地环境中形成。

(b) 微调后模型问答

图 12 模型微调前后问答对比

Fig.12 Comparison of question and answer performance before and after model fine-tuning

3 存在的局限性、挑战和解决策略

ChatGPT作为一种强大的自然语言处理工具,对许多领域产生了深远影响。在油气勘探和开发专业中,ChatGPT能够协助进行数据分析,模型预测等多重任务。实际应用过程中,也需要面对一些局限和挑战,包括数据的可用性及质量、模型的精确度与可靠性,以及数据安全与保护等多方面问题。

3.1 存在的局限性

ChatGPT作为一种基于人工智能的对话生成模型,其在提供高度自然的机器交流体验中起到重要作用,然而,也受制于多方面存在的局限性。首当其冲的问题便是回答的真实性问题。由于ChatGPT基于的是机器学习算法,它并不能像人类一样对生成的内容进行实时审核。该模型无法判别对于某一信息的合理性,它将使用其在训练过程中学习到的信息进行自主生成,而这种生成的内容很可能并不保证准确性。在实际运用中,科研工作者需要保持极高的警惕性和审慎态度,对ChatGPT的回答进行核实,确保获取到信息的真实性。

另外,ChatGPT面临着可能的数据污染问题。这个问题源自当大量低质量或者错误的信息进入训练数据集时,会对ChatGPT的整体性能和输出质量产生负面影响^[31-32]。例如,如果大量的虚假信息和不准确论述

被包括在训练数据中,那么ChatGPT就有可能吸收和学习到这些信息,从而在生成的回答中反映出来。针对这一问题,科研人员需要使用数据检测工具筛选并核实用于训练的数据,保证生成文本的质量。

3.2 数据可用性和质量问题

数据的可用性和质量是强化学习系统如ChatGPT应用在油气勘探开发专业中面临的主要挑战之一。数据的缺乏,或者数据质量低,变异性大。在油气勘探开发中,高质量、可用的数据是关键。但是,由于地理和气候条件严酷、设备故障等各种因素,数据收集经常会遇到困难,数据质量也可能会受到影响,这对ChatGPT模型的训练和预测产生挑战。以某油田的物探项目为例,由于地质条件复杂,检波器收集到的地震数据噪声大,对ChatGPT模型的训练造成困难。解决这个问题的策略是通过预处理技术,如去噪算法,提高数据质量。此外,运用迁移学习技术,即使在数据稀少的情况下,也可以在其他相似任务中学习的知识迁移到此任务中,提高模型的预测性能。

在油气开发生产场景中,训练数据是比较重要的,而这些数据可能包括生产层的层位信息、岩性、孔隙度、渗透率、产量、含水率等。如果数据记录不全或数据误差大,ChatGPT在训练过程中获取的信息可能会产生偏差,导致预测结果的准确性减小。

因此,建立一个适当的数据收集和管理系统评估数据至关重要,并据此制定改进策略。这可以包括改进数据收集过程以减小可能的误差,或者引入适当的数据清理流程识别和消除找到的错误。目前各大油田开展的数据湖工作极大提高数据的可用性和质量。

3.3 确保模型的准确性和可靠性

模型的准确性和可靠性是另一个关键的挑战。因为在油气勘探开发中,模型预测的误差可能导致巨大经济损失,甚至可能造成安全事故。例如,在某油田开发项目中,ChatGPT用于预测石油储层的分布,如果预测准确,有助于制定准确的开发策略,大大提高开发效率和效益;反之,如果预测不准确,可能加大钻探风险,造成巨大的经济损失。为了提高模型的准确性,可以使用各种方法,如使用更复杂的模型,更多的数据,以及结合领域专家的知识对模型进行微调和优化。

3.4 数据安全和保护

在油气勘探开发中,数据安全和保护也是十分重要的问题。这不仅包括研究数据,也包括与油气企业的商业机密的相关数据。例如,地震数据、储量数据及产量数据等,包含了与油气田相关的重要信息,这些数据如果发生泄密,可能会带来安全风险。

因此,需要实施严格的数据保护策略,包括对数据进行加密,对访问进行身份验证,以及把敏感数据和其他数据分离存储等措施,保护企业的重要数据和商业机密^[33-34]。同时,随着GPT系列模型的不断开源,石油企业可以尝试训练自己的本地化模型,这样能够充分保护企业数据的安全。当然,随着国内对于自然语言处理模型的开发,也会涌现出擅长处理汉字的模型(如:百度的文心一言、科大讯飞的讯飞星火),也可以尝试将其应用到油气勘探开发领域。

4 ChatGPT在油气勘探开发领域的前景

ChatGPT与数据湖的结合。数据湖(data lake)是一种大数据存储和管理方式。其基本意义是提供一种方法来存储大量、原始或预处理的数据,这些数据可以是结构化的,半结构化的,或者未结构化的,并且可以从各种来源获取。例如,油气勘探开发专业的地震资料、测井资料等数据属于结构化数据;半结构化数据,比如各种excel文档和xml文件等;非结构化数据,比如研究报告word、PPT和pdf文档;二进制数据,比如图片和视频等^[35]。相比传统的数据库,数据湖能更灵活地处理不同类型的数据,并且能以更低的成本进行数据保存及分析^[36]。另外,它可以存储大规模的原始数据,使科研人员可以充分挖掘隐藏的信息,得出更准确的分析结果。目前GPT-4可以支持多模态数据的输入,包括一些半结构化和非结构化数据,甚至对图片和视频这样的二进制数据也能够识别和分析,并根据用户的提问,得到相关分析结果。

ChatGPT 与数据湖相结合可以应用于以下方向:

数据查询和分析: ChatGPT 可以被训练成一个用于查询和分析数据湖里的大量油气田数据的工具。用户可以使用自然语言查询,例如:“请分析 A 油田上个月的产量情况,并给出产量变化的主要因素。”然后 ChatGPT 处理这个问题,翻译成相应的数据查询,并返回人类可读的结果。

故障预测与维护: 油气生产过程中会生成大量的实时数据,这些数据可以用来预测设备出现故障的可能性。ChatGPT 可以被训练来理解这些数据,并在问题出现之前发出警告。例如,当某口油井产量出现下降的情况,模型可以根据数据湖中所有油井的历史情况,综合分析原因,并推荐解决方案。

自动生成报告: 基于数据湖中的数据,ChatGPT 可以生成关于勘探、开发、生产、效率、成本等的定制化报告。这些报告以易于理解的自然语言格式生成,让决策者更快获取重要信息。

培训和教育: 通过对具有专业知识的资料进行训练,ChatGPT 可以为新员工或不熟悉特定主题的人提供咨询和解答服务。

自动化工作流程: ChatGPT 可以帮助自动化一些日常的、基于数据湖中数据的工作流程。比如,系统可以自动检测到某个设备的性能下降,并生成请求维修的工单。

当然,要实现这些功能,需要对模型进行一定的训练和调整,适应特定的数据结构和业务需求,也需要充分考虑数据安全性和隐私保护等问题。随着用户对模型的专业性和数据安全等问题的要求,专属的本地化大语言模型的发展越来越快,如清华大学的 ChatGLM 和 Meta 的 LLaMA,通过本地化部署和训练,能够更好地理解并生成特定语言或者特定专业的内容。然而,在使用这些模型时,也需要注意一些挑战,例如数据的数量与多样性、模型的公正性与偏见问题,以及在更广泛的语境中保持其一致性和理解能力的问题。

5 结 论

根据上述对 ChatGPT 在油气勘探开发中的应用潜力、面临的挑战以及未来前景的分析,得出以下主要结论。

1) ChatGPT 是基于人工智能的语言模型,利用深度学习中的 Transformer 模型进行训练。它通过自回归的方式生成连贯、上下文相关的回答。ChatGPT 推动工作效率和决策力的提升,在应对油气行业复杂挑战中具备独特的优势。

2) ChatGPT 在油气勘探开发中提供高效的信息检索和知识管理能力、提高应用人工智能的广度和效率,创新人工智能解决问题方式、提高人工智能的应用效率。另外,ChatGPT 能够利用专业知识和数据对模型进行微调,构建油气勘探开发领域的专属应用。

3) ChatGPT 在油气勘探开发中面临的局限性和挑战包括回答的真实性问题、数据可用性和质量问题、确保模型的准确性和可靠性、数据安全和保护等。解决这些问题需要加强数据质量管理、模型优化和安全保护等方面的努力。

4) ChatGPT 的未来前景在于进一步提升对话质量和复杂问题处理能力,并结合数据湖等技术实现关键业务的自动化和智能化。数据湖与 ChatGPT 的结合能够提供高度自定义和适应性的应用,如数据查询和分析、故障预测与维护、自动生成报告、培训和教育等。

参考文献

- [1] 徐卫克. 基于人工智能的自然语言处理系统分析[J]. 网络安全技术与应用, 2023(7): 49-51.
Xu W K. Analysis of natural language processing system based on artificial intelligence[J]. Network Security Technology and Applications, 2023(7): 49-51. (in Chinese)
- [2] 周鼎凯, 张枫林, 丁治国, 等. 自然语言处理技术在博物馆领域的应用前景研究: 以 ChatGPT 为例[J]. 科学教育与博物馆, 2023, 9(3): 39-48.

- Zhou D K, Zhang F L, Ding Z G, et al. A study on the application prospects of natural language processing technology in the museum field:taking ChatGPT as an example[J]. *Science Education and Museums*, 2023, 9(3): 39-48. (in Chinese)
- [3] 孙凯丽,罗旭东,罗有容.预训练语言模型的应用综述[J].*计算机科学*,2023,50(1): 176-184.
- Sun K L, Luo X D, Luo Y R. Survey of application of pretrained language models[J]. *Computer Science*, 2023, 50(1): 176-184. (in Chinese)
- [4] OpenAI, Josh A, Steven A, et al. GPT-4 technical report[EB/OL]. (2023-03-15)[2023-07-18]. <https://arxiv.org/abs/2303.08774>.
- [5] 王树义,张庆薇. ChatGPT给科研工作者带来的机遇与挑战[J].*图书馆论坛*,2023,43(3): 109-118.
- Wang S Y, Zhang Q W. ChatGPT's opportunities and challenges for researchers[J]. *Library Tribune*, 2023, 43(3): 109-118. (in Chinese)
- [6] Ouyang L, Wu J, Jiang X, et al. Training language models to follow instructions with human feedback[EB/OL]. (2022-03-04)[2023-07-18]. <https://arxiv.org/abs/2203.02155>.
- [7] 苏静, Murtadha A. 一种基于半监督的句子情感分类模型[J].*重庆大学学报*,2024,47(12):100-113.
- Su J, Murtadha A. A semi-supervised model for sentence-level sentiment classification[J]. *Journal of Chongqing University*, 2024, 47(12): 100-113. (in Chinese)
- [8] 赵朝阳,朱贵波,王金桥. ChatGPT给语言大模型带来的启示和多模态大模型新的发展思路[J].*数据分析与知识发现*,2023,7(3): 26-35.
- Zhao C Y, Zhu G B, Wang J Q. The inspiration brought by ChatGPT to LLM and the new development ideas of multi-modal large model[J]. *Data Analysis and Knowledge Discovery*, 2023, 7(3): 26-35. (in Chinese)
- [9] 冯志伟,张灯柯,饶高琦. 从图灵测试到ChatGPT: 人机对话的里程碑及启示[J].*语言战略研究*, 2023, 8(2): 20-24.
- Feng Z W, Zhang D K, Rao G Q. From turing test to ChatGPT:A milestone of man-machine interaction and its enlightenment[J]. *Chinese Journal of Language Policy and Planning*, 2023, 8(2): 20-24. (in Chinese) 1-11.
- [10] 吕君杰,郑石桥. ChatGPT在审计中的可能应用路径[J].*商业会计*,2023(9): 47-49.
- Lv J J, Zheng S Q. Possible application paths of ChatGPT in auditing[J]. *Commercial Accounting*, 2023(9):47-49. (in Chinese)
- [11] 卢宇,余京蕾,陈鹏鹤,等.生成式人工智能的教育应用与展望:以ChatGPT系统为例[J].*中国远程教育*, 2023, 43(4): 24-31.
- Lu Y, Yu J L, Chen P H, et al. Educational applications and prospects of generative artificial intelligence: taking the ChatGPT system as an example[J]. *Chinese Journal of Distance Education*, 2023, 43(4): 24-31. (in Chinese)
- [12] 陆伟,刘家伟,马永强,等. ChatGPT为代表的大模型对信息资源管理的影响[J].*图书情报知识*,2023,40(2): 6-9.
- Lu W, Liu J W, Ma Y Q, et al. The influence of large language models represented by ChatGPT on information resources management[J]. *Documentation, Information & Knowledge*, 2023, 40(2): 6-9. (in Chinese)
- [13] 陆小华. ChatGPT等智能内容生成与新闻出版业面临的智能变革[J].*中国出版*, 2023, 550(5): 8-15.
- Lu X H. ChatGPT and the intelligent transformation faced by the news publishing industry[J]. *China Publishing Journal*, 2023, 550(5):8-15. (in Chinese)
- [14] 杨勇.胜利油田勘探开发大数据及人工智能技术应用进展[J].*油气地质与采收率*, 2022, 29(1): 1-10.
- Yang Y. Application progress of big data&AI technologies in exploration and development of Shengli Oilfield[J]. *Petroleum Geology and Recovery Efficiency*, 2022, 29(1): 1-10. (in Chinese)
- [15] 马陇飞,萧汉敏,陶敬伟,等.基于梯度提升决策树算法的岩性智能分类方法[J].*油气地质与采收率*, 2022, 29(1): 21-29.
- Ma L F, Xiao H M, Tao J W, et al. Intelligent lithology classification method based on GBDT algorithm[J]. *Petroleum Geology and Recovery Efficiency*, 2022, 29(1): 21-29. (in Chinese)
- [16] 史长林,魏莉,张剑,等.基于机器学习的储层预测方法[J].*油气地质与采收率*, 2022, 29(1): 90-97.
- Shi C L, Wei L, Zhang J, et al. Reservoir prediction method based on machine learning[J]. *Petroleum Geology and Recovery Efficiency*, 2022, 29(1): 90-97. (in Chinese)
- [17] 李阳,廉培庆,薛兆杰,等.大数据及人工智能在油气田开发中的应用现状及展望[J].*中国石油大学学报(自然科学版)*, 2020, 44(4): 1-11.
- Li Y, Lian P Q, Xue Z J, et al. Application status and prospect of big data and artificial intelligence in oil and gas field development[J]. *Journal of China University of Petroleum (Edition of Natural Science)*, 2020, 44(4):1-11. (in Chinese)

- [18] 王敏生, 光新军, 耿黎东. 人工智能在钻井工程中的应用现状与发展建议[J]. 石油钻采工艺, 2021, 43(4): 420-427.
Wang M S, Guang X J, Geng L D. Application status and development suggestions of artificial intelligence in drilling engineering[J]. Oil Drilling & Production Technology, 2021, 43(4): 420-427. (in Chinese)
- [19] 黄家宸, 张金川. 机器学习预测油气产量现状[J]. 油气藏评价与开发, 2021, 11(4): 613-620.
Huang J C, Zhang J C. Overview of oil and gas production forecasting by machine learning[J]. Petroleum Reservoir Evaluation and Development, 2021, 11(4): 613-620. (in Chinese)
- [20] 杨洋, 程悦菲, 谢英, 等. 基于时序动态分析的油井产量预测研究[J]. 西南石油大学学报(自然科学版), 2020, 42(6): 82-88.
Yang Y, Cheng Y F, Qiao Y, et al. A study on oil well production prediction based on time series dynamic analysis[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2020, 42(6): 82-88. (in Chinese)
- [21] 赵洪涛, 李金泽, 杨毅, 等. 基于长短期记忆神经网络的油井产液量和含水率预测方法[J]. 石油地质与工程, 2022, 36(5): 75-80.
Zhao H T, Li J Z, Yang Y, et al. Production and water cut performance predicting method based on long-short term memory neural network model[J]. Petroleum Geology and Engineering, 2022, 36(5): 75-80. (in Chinese)
- [22] 苏健, 刘合. 石油工程大数据应用的挑战与发展[J]. 中国石油大学学报(社会科学版), 2020, 36(3): 1-6.
Su J, Liu H. Challenges and development of big data application in petroleum engineering[J]. Journal of China University of Petroleum (Edition of Social Sciences), 2020, 36(3): 1-6. (in Chinese)
- [23] 阎红巧, 樊志强, 郝壮远. 大数据技术在油气行业事故事件致因分析中的应用[J]. 安全与环境工程, 2021, 28(6): 31-37.
Yan H Q, Fan Z Q, Hao Z Y. Application of big data technology to cause analysis of accidents and incidents in oil and gas industry[J]. Safety and Environmental Engineering, 2021, 28(6): 31-37. (in Chinese)
- [24] 闵超, 代博仁, 张馨慧, 等. 机器学习在油气行业中的应用进展综述[J]. 西南石油大学学报(自然科学版), 2020, 42(6): 1-15.
Min C, Dai B R, Zhang X H, et al. A review of the application progress of machine learning in oil and gas industry[J]. Journal of Southwest Petroleum University (Science & Technology Edition), 2020, 42(6): 1-15. (in Chinese)
- [25] 窦宏恩, 张蕾, 米兰, 等. 人工智能在全球油气工业领域的应用现状与前景展望[J]. 石油钻采工艺, 2021, 43(4): 405-419.
Dou H E, Zhang L, Mi L, et al. The application status and prospect of artificial intelligence in the global oil and gas industry[J]. Oil Drilling & Production Technology, 2021, 43(4): 405-419. (in Chinese)
- [26] 林伯韬, 郭建成. 人工智能在石油工业中的应用现状探讨[J]. 石油科学通报, 2019, 4(4): 403-413.
Lin B T, Guo J C. Discussion on current application of artificial intelligence in petroleum industry[J]. Petroleum Science Bulletin, 2019, 4(4): 403-413. (in Chinese)
- [27] 于红岩, 丁伟, 高彦芳, 等. 人工智能在提高油气田勘探开发效果中的应用[J]. 西北大学学报(自然科学版), 2022, 52(6): 1086-1099.
Yu H Y, Ding S W, Gao Y F, et al. Application of artificial intelligence in improving the effectiveness of oil and gas field exploration and development[J]. Journal of Northwest University (Natural Science Edition), 2022, 52(6): 1086-1099. (in Chinese)
- [28] 宋洪庆, 都书一, 周园春, 等. 油气资源开发的大数据智能平台及应用分析[J]. 工程科学学报, 2021, 43(2): 179-192.
Song H Q, Du S Y, Zhou Y C, et al. Big data intelligent platform and application analysis for oil and gas resource development[J]. Chinese Journal of Engineering, 2021, 43(2): 179-192. (in Chinese)
- [29] 彭涛, 肖建明, 张仕慧, 等. 基于多参数 MRI 及影像组学建立机器学习模型诊断临床显著性前列腺癌[J]. 中国医学影像技术, 2019, 35(10): 1526-1530.
Peng T, Xiao J M, Zhang S H, et al. Establishment of machine learning models for diagnosis of clinically significant prostate cancer based on multi-parameter MRI and radiomics[J]. Chinese Journal of Medical Imaging Technology, 2019, 35(10): 1526-1530. (in Chinese)
- [30] 肖焕辉, 袁程朗, 冯仕庭, 等. 基于深度学习的癌症计算机辅助分类诊断研究进展[J]. 国际医学放射学杂志, 2019, 42(1): 22-25.
Xiao H H, Yuan C L, Feng S T, et al. Research progress of computer aided diagnosis in cancer based on deep learning[J]. International Journal of Medical Radiology, 2019, 42(1): 22-25. (in Chinese)
- [31] 姜思羽, 张智恒, 姜立标, 等. 基于伪标签和迁移学习的双关语识别方法[J]. 重庆大学学报, 2024, 47(2): 51-61.
Jiang S Y, Zhang Z H, Jiang L B, et al. Pun detection based on pseudo-label and transfer learning[J]. Journal of Chongqing

- University, 2024, 47(2): 51-61. (in Chinese)
- [32] 李敏, 时瑞浩, 张莹, 等. 基于混合风格迁移的智能合约漏洞检测方法[J]. 重庆大学学报, 2024, 47(12): 70-82.
Li M, Shi R H, Zhang Y, et al. Smart contract vulnerability detection method based on MixStyle transfer[J]. Journal of Chongqing University, 2024, 47(12): 70-82. (in Chinese)
- [33] 邹旭鹏, 史纪强, 彭英, 等. 万物互联时代数据安全保护与保密问题思考[J]. 中国信息化, 2021, 323(3): 73-76.
Zou X P, Shi J Q, Peng Y, et al. Reflection on data security protection and confidentiality in the era of everything interconnected[J]. China Informatization, 2021, 323(3): 73-76. (in Chinese)
- [34] 王滔, 杨德保, 帅旗. 油田数据中心数据的安全保护策略[J]. 中国管理信息化, 2021, 24(18): 109-110.
Wang T, Yang D B, Shuai Q. Security protection strategy for data in the oilfield data center[J]. China Management Informatization, 2021, 24(18): 109-110. (in Chinese)
- [35] 张芸. 浅谈石油勘探行业数据湖建设中的数据治理问题[J]. 中国管理信息化, 2021, 24(9): 122-124.
Zhang Y. A discussion on data governance issues in the construction of data lakes in the petroleum exploration industry[J]. China Management Informatization, 2021, 24(9): 122-124. (in Chinese)
- [36] 杨剑锋, 杜金虎, 杨勇, 等. 油气行业数字化转型研究与实践[J]. 石油学报, 2021, 42(2): 248-258.
Yang J F, Du J H, Yang Y, et al. Research and practice on digital transformation of the oil and gas industry[J]. Acta Petrolei Sinica, 2021, 42(2): 248-258. (in Chinese)

(编辑 侯湘)