成绩：

陕西科技大学
SHAANXI UNIVERSITY OF SCIENCE & TECHNOLOGY

人 工 智 能 导 论

课 程 报 告

题目：基于深度学习的自然语言处理研究进展综述

姓　　名：马凌峰

学　　号：202307020122

教　　师：陈海丰

班　　级：计算机 232

所在院系：电子信息与人工智能学院

文章编号：1002-185X2023-0702-0122

基于深度学习的自然语言处理研究进展综述

马凌峰 [1]

(陕西科技大学学报，陕西 西安，710021)

摘　要: 本报告旨在深入探讨基于深度学习的自然语言处理（NLP）领域的最新研究进展。报告首先界定 NLP 的核心任务、面临的挑战及关键评测数据集，随后详细阐述深度学习在 NLP 中的发展历程、核心模型架构（如 Transformer 及其变体、BERT、GPT 系列、T5 等预训练语言模型），以及迁移学习、注意力机制、上下文学习和多模态学习等关键学习范式。最后，报告对比了主流模型的性能，并对未来研究方向进行了展望。本报告为综述类，旨在为人工智能领域的学生和研究人员提供一个全面、深入的 NLP 深度学习技术现状概览。分析表明：Transformer 架构及其衍生的大规模预训练语言模型已成为 NLP 领域的主导范式，显著提升了多项任务的性能；迁移学习和上下文学习极大地提高了模型效率和泛化能力；多模态融合是未来的重要方向，但数据稀缺、模型偏见和可解释性等挑战依然存在，亟待解决。

关键词: 自然语言处理；深度学习；预训练语言模型；Transformer；伦理挑战

A Review of Research Progress in Natural Language Processing Based on Deep Learning

LINGFENG MA[1]

(1.Shaanxi University of Science & Technology Journal, Xi'an, Shaanxi, 710021, China)

Abstact : This report provides a comprehensive overview of the latest advancements in deep learning-based Natural Language Processing (NLP). It begins by defining NLP's core tasks, challenges, and key evaluation datasets. Subsequently, it details the evolution of deep learning in NLP, core model architectures (such as Transformer and its variants, including pre-trained language models like BERT, GPT series, and T5), and essential learning paradigms like transfer learning, attention mechanisms, in-context learning, and multimodal learning. Finally, the report compares the performance of mainstream models and outlines future research directions. This review-type report aims to offer a thorough and forward-looking overview of the current state of deep learning NLP technologies for students and researchers in artificial intelligence. Analysis indicates that the Transformer architecture and its derived large-scale pre-trained language models have become the dominant paradigm in NLP, significantly enhancing performance across various tasks. Transfer learning and in-context learning have greatly improved model efficiency and generalization capabilities. Multimodal integration is a crucial future direction, yet challenges such as data scarcity, model bias, and interpretability persist and require urgent attention.

Key words：Natural Language Processing；Deep Learning；Transformer；Ethical Challenges

1 引言

1.1 背景与意义

近年来，“以人为本，服务于人”的理念在人工智能研究中得到越来越广泛的关注。自然语言处理（NLP）作为人工智能的一个核心分支，赋予计算机理解、解释和生成人类语言的能力。随着大数据时代的到来和计算能力的飞速提升，深度学习技术在计算机视觉等领域取得了巨大成功。这种成功也迅速扩展到 NLP 领域，彻底改变了传统 NLP 方法，使其从基于规则和统计的方法转向了以神经网络为核心的范式。

深度学习在计算机视觉领域的成功为 NLP 领域提供了方法论上的借鉴和信心，加速了其发展。计算机视觉（CV）和 NLP 作为人工智能的两大核心领域，在深度学习时代呈现出方法论上的相互借鉴和融合趋势。深度学习提供了一种强大的、端到端的特征学习能力，能够从原始数据中自动提取高层次、抽象的特征，这对于处理图像和文本这种高维、非结构化数据都具有普适性。这种方法论的跨领域迁移是人工智能领域快速发展的重要驱动力。

NLP 技术的进步在医疗健康、金融、教育、交通等多个行业带来了革命性的应用，例如智能客服、机器翻译、情感分析、内容生成等。这些应用极大地提升了效率，优化了用户体验，并催生了新的商业模式。

1.2 报告结构与目标

本报告将按照既定框架，从 NLP 的定义、挑战、数据集入手，逐步深入到深度学习在 NLP 中的方法进展，包括其发展历程、核心模型架构和关键学习范式。随后，将对相关方法的性能进行对比分析，最后总结当前研究现状并展望未来发展趋势。本报告旨在为读者提供一个全面、系统且具有前瞻性的深度学习 NLP 领域综述，帮助理解该领域的核心概念、最新技术突破、面临的挑战以及未来的发展方向。

2 问题定义

2.1 自然语言处理的定义与核心任务

自然语言处理（NLP）是人工智能的一个分支，旨在使计算机能够理解、解释、操作并生成人类语言。它结合了计算语言学与统计建模、机器学习和深度学习技术。NLP 的核心任务包括：文本预处理、句法分析、语义分析、信息抽取（如命名实体识别）、情感分析、文本分类、机器翻译、问答系统、文本摘要和自然语言生成（NLG）等。

深度学习的引入使得 NLP 任务从离散的、基于规则的子任务组合转变为更统一、端到端的模型，尤其是在语言生成和复杂语义理解方面。模型能够直接从原始文本中学习复杂的特征表示，并统一处理多个任务，极大地提高了模型的泛化能力和性能，尤其是在生成式任务上，使得 AI 能够创造出更自然、更连贯的文本。

表 1　实验转化率因素极差分析

任务类别	深度学习应用示例
文本预处理	基于神经网络的分词和词向量嵌入
句法分析	基于 RNN/Transformer 的依存句法分析
语义分析	Word2Vec、BERT 等词嵌入和上下文表示
信息抽取	基于 BERT/Transformer 的命名实体识别（NER）
情感分析	基于 LSTM/Transformer 的情感分类
文本分类	基于 BERT/Transformer 的文本分类
机器翻译	Transformer 架构的神经机器翻译（NMT）
问答系统	BERT、GPT 系列在 SQuAD 等数据集上的应用
文本摘要	T5 模型、GPT 系列生成式摘要

自然语言生成	GPT 系列、T5 等大型语言模型生成文章、对话

2．2 基于深度学习的 NLP 面临的挑战

NLP 领域面临的挑战是技术、社会、伦理多维度交织的复杂系统问题，尤其体现在数据、偏见和可解释性上。

- **数据稀缺性与低资源语言：** 尽管深度学习需要大量数据，但高质量、标注好的数据获取成本高昂。对于全球数千种低资源语言，数字内容和标注数据极度匮乏，这限制了 NLP 模型在这些语言上的发展和应用，导致性能差距显著。
- **语言的歧义性与上下文理解：** 人类语言固有的歧义性（词语或句子在不同语境下有不同含义）是 NLP 的根本挑战。模型需要深层次的上下文理解、常识知识和推理能力来解决这些歧义，而这远超简单的模式匹配。
- **模型偏见与公平性：**训练数据中存在的历史、社会和文化偏见会被模型学习并放大，导致 AI 系统在招聘、贷款、医疗诊断等敏感应用中产生歧视性结果。确保模型的公平性是一个重要的伦理挑战。
- **模型可解释性与透明度：** 深度学习模型，特别是大型语言模型（LLMs），通常被视为“黑箱”，其决策过程不透明。在医疗、法律等高风险领域，缺乏可解释性会降低用户信任，并阻碍对模型错误原因的诊断和修正。
- **计算资源与可扩展性：**训练和部署大型深度学习 NLP 模型需要巨大的计算资源（GPU/TPU）、能源和时间，这使得 AI 开发成本高昂且困难。
- **鲁棒性与泛化能力：**尽管模型在基准测试上表现出色，但在面对真实世界中多样化、非结构化、甚至对抗性攻击的数据时，其鲁棒性和泛化能力仍有待提高。模型可能过度依赖数据集中的虚假关联而非真正的语言理解 。

2．3 评测数据集与基准

评测数据集和基准（Benchmarks）在 NLP 研究中扮演着至关重要的角色，它们为不同模型提供了标准化、可量化的性能比较平台，推动了领域的发展。基准测试的演进，例如从 GLUE 到 SuperGLUE 的演进，体现了模型能力与评估方法之间的动态博弈。GLUE 的出现标准化了语言理解任务的评估，使得不同模型可以进行公平比较。

然而，随着模型规模和复杂性的增加，模型开始在 GLUE 上取得接近甚至超越人类的性能，这促使研究者开发了 SuperGLUE，它包含了更困难、更需要深层推理和常识知识的任务。MMLU 则进一步将评估范围扩展到跨学科的知识广度，反映了对通用人工智能（AGI）能力的追求。

表 2：主流 NLP 基准测试数据集概览

基准名称	主要任务类型	特点/挑战
GLUE	通用语言理解（多任务，如情感分析、文本分类、自然语言推理）	标准化评估，推动早期模型发展，但可能存在过拟合风险
SuperGLUE	更复杂的通用语言理解（多任务，如因果推理、阅读理解、词语上下文理解）	任务更具挑战性，要求深层推理和常识，旨在克服 GLUE 的局限
SQuAD	机器阅读理解（抽取式问答）	真实世界问题，答案在文本中，SQuAD 2.0 引入无法回答的问题
WMT	机器翻译（多语言对）	人工评估，关注翻译质量和流

		畅度，涉及低资源语言挑战
MMLU	多学科知识理解与推理	跨学科，评估知识广度和深度，接近人类专家水平

3 核心能力与方法进展

3．1　深度学习在 NLP 中的发展历程

NLP 的发展并非简单的技术迭代，而是一个深刻的范式转变，其核心在于对“语言理解”的模拟方式的演进。

- **早期探索 (2000s-2010s 初)：** 2000 年代，NLP 开始集成更复杂的算法，如支持向量机（SVM）和隐马尔可夫模型（HMM）。随着深度学习的兴起，循环神经网络（RNN）及其变体（如长短期记忆网络 LSTM）被引入 NLP，它们在处理序列数据方面表现出色，能够捕捉语言中的长期依赖关系。Word2Vec 等词嵌入技术（2013 年）的出现，将词语映射到低维向量空间，捕获了词语的语义和上下文信息，为深度学习在 NLP 中的应用奠定了基础。
- **Transformer 时代 (2017 至今)：** 2017 年，Vaswani 等人提出了 Transformer 架构，其核心是自注意力机制（Self-Attention），彻底改变了 NLP 领域。Transformer 摒弃了 RNN 的序列处理方式，实现了并行计算，极大地提升了训练效率和处理长序列的能力。它在机器翻译等任务上取得了突破性进展，并成为后续所有大型预训练语言模型（PLMs）的基础架构。
- **预训练语言模型（PLMs）的崛起 (2018 至今)：** 2018 年，Google 发布了 BERT（Bidirectional Encoder Representations from Transformers），通过“掩码语言模型”和“下一句预测”等无监督任务在大规模语料上进行预训练，实现了对上下文的双向理解。BERT 的出现开启了“预训练-微调”的范式，即先在海量无标注文本上预训练通用语言表示，再针对特定下游任务进行微调。OpenAI 的 GPT 系列模型（GPT-1, GPT-2, GPT-3, GPT-4 等）则专注于生成式任务，采用 Decoder-only 架构，参数量从百万级迅速增长到千亿级，展现了惊人的文本生成能力和少样本学习能力。Google 的 T5（Text-to-Text Transfer Transformer）将所有 NLP 任务统一为“文本到文本”的格式，简化了模型架构，并在多任务学习中表现出色 。

3．2　核心模型架构

Transformer 模型于 2017 年提出，完全依赖于自注意力（Self-Attention）机制来计算输入和输出的表示，无需循环或卷积。它由编码器（Encoder）和解码器（Decoder）组成，每个部分都包含多头自注意力层和前馈网络。自注意力机制允许模型在处理序列中的每个词时，同时关注序列中的所有其他词，并根据其相关性分配不同的权重，从而有效捕捉长距离依赖。多头注意力机制则允许模型从不同的表示子空间中学习信息，增强了模型的表达能力。相比 RNN 和 LSTM，Transformer 能更好地处理长距离依赖，并且通过并行计算显著提高了训练速度。

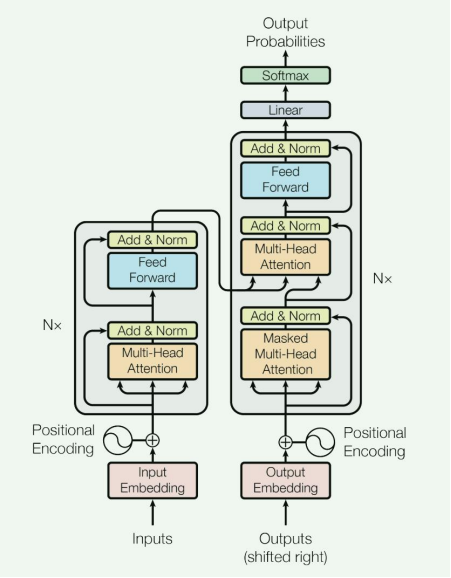

图 1　Transformer 架构图

3．2．2　预训练语言模型(BERT, GPT 系列, T5 等)

预训练语言模型（PLMs）是 Transformer 架构

在 NLP 领域最成功的应用之一，它们通过在大规模无标注文本上进行预训练，学习通用的语言表示，然后通过微调适应各种下游任务。

- **BERT (Bidirectional Encoder Representations from Transformers)：** 由 Google 于 2018 年推出，是首个通过双向上下文理解进行预训练的模型。BERT 通过“掩码语言模型”和“下一句预测”等无监督任务进行预训练，使其能够捕捉词元在上下文中的双向关系，从而在问答、情感分析、文本分类等多种 NLP 任务中取得了显著性能提升。

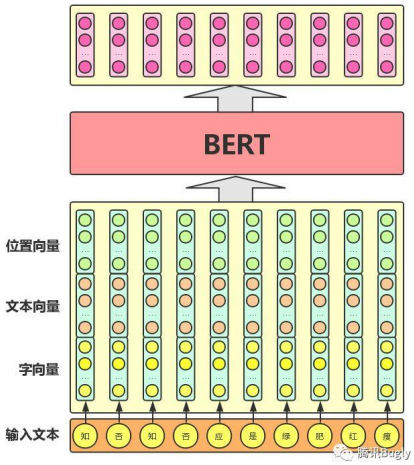

图 2　BERT 预训练任务示意图

- **GPT 系列 (Generative Pre-trained Transformer)：** 由 OpenAI 开发，专注于文本生成任务，采用 Decoder-only 的 Transformer 架构。GPT-3（2020 年）拥有 1750 亿参数，展现了惊人的少样本学习（Few-shot Learning）能力。GPT-4（2023 年）进一步提升了上下文理解和文本生成能力，在复杂推理和创造性任务中表现出色，并加强了偏见缓解和伦理安全措施。
- **T5 (Text-to-Text Transfer Transformer)：** 由 Google 于 2019 年提出，其核心思想是将所有 NLP 任务统一为“文本到文本”的格式，无论是翻译、摘要、问答还是分类，都被视为输入文本到输出文本的转换问题。T5 采用 Encoder-Decoder 架构，并在大规模语料上进行“去噪”预训练。

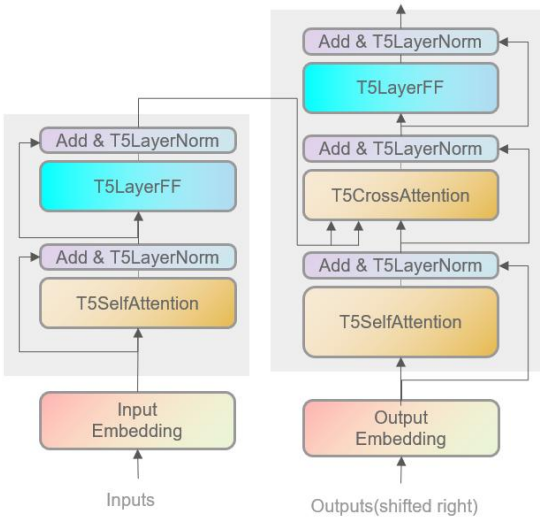

图 3　T5 模型架构图

3．3　关键学习范式

3．3．1 迁移学习与微调

迁移学习是深度学习在 NLP 领域取得突破的关键范式之一。它允许模型将在一个任务上学到的知识应用于另一个相关任务，从而显著减少目标任务所需的数据量和训练时间。在 NLP 中，这通常意味着使用在大规模文本语料库上预训练的语言模型（如 BERT、GPT、T5）作为基础，然后针对特定的下游任务进行微调。这种“预训练-微调”范式显著提升了 NLP 任务的性能，降低了对大量标注数据的需求，并增强了模型的泛化能力。

3．3．2 注意力机制与关系学习

注意力机制是 Transformer 架构的核心，也是深度学习 NLP 模型能够有效捕捉词语间复杂关系的关键。它允许模型在处理序列中的每个元素时，动态地分配不同的“注意力权重”给序列中的其他元素，从而聚焦于最相关的信息。自注意力机制和多头注意力机制是其主要类型，它们有效捕捉长距离依赖并增强模型表达能力。注意力机制的引入显著提升了模型的性能，并增强了模型的可解释性。

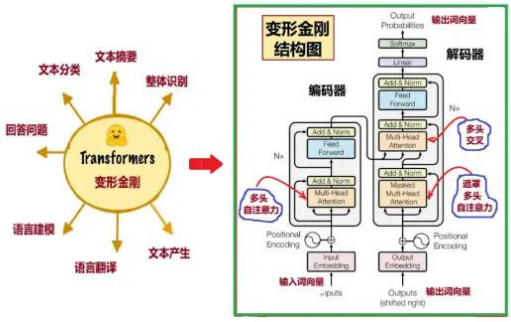

图4　Transformer 模型及其中的注意力机制

3．3．3 上下文学习与提示工程

上下文学习（In-Context Learning, ICL）是一种新兴的范式，允许大型语言模型（LLMs）通过在输入提示（Prompt）中提供少量示例来学习新任务，而无需进行模型参数的更新（即无需微调）[49]。这种能力是大型预训练模型在足够多样性和规模的预训练数据上训练后出现的一种“涌现现象”。提示工程是优化提示语言以引导 LLM 生成所需输出的过程，有效的提示工程能够显著提升模型性能。

3．3．4 多模态学习

多模态学习是 NLP 领域的一个重要发展方向，它旨在使模型能够处理和理解来自不同模态的数据，如文本、图像、音频和视频，从而更全面地理解人类交流和世界。通过结合不同模态的信息，模型可以捕捉单一模态无法感知的线索，例如，仅凭文本可能难以理解的讽刺意味，通过结合图像可以被多模态模型准确感知。多模态 NLP 的应用包括情感识别、机器翻译、图像字幕生成和视觉问答（VQA）等。

4 AI 性能与应用成果

4．1　性能评估指标

在 NLP 领域，模型的性能评估至关重要，它为研究者提供了衡量进展和比较不同方法优劣的标准化方式。常用的评估指标因任务类型而异：

- **分类任务：** 准确率（Accuracy）、精确率（Precision）、召回率（Recall）和 F1 分数（F1-Score）。F1 分数在类别不平衡时特别有用，通常 F1 分数达到 0.85 以上被认为是高质量系统。
- **问答任务（如 SQuAD）：** 精确匹配（Exact Match, EM）和 F1 分数。
- **机器翻译任务（如 WMT）：** BLEU 分数（Bilingual Evaluation Understudy）和 COMET/ChrF。
- **通用语言理解基准（如 GLUE, SuperGLUE, MMLU）：** 这些基准本身包含多个子任务，最终通常会提供一个综合分数来衡量模型在多任务上的平均表现。

4．2　主流模型在基准测试上的表现

近年来，基于 Transformer 架构的预训练语言模型（PLMs）在各种 NLP 基准测试中取得了显著的性能突破，不断刷新着排行榜的记录。

- **GLUE 和 SuperGLUE：** BERT、RoBERTa、ALBERT、ELECTRA 等模型在这些通用语言理解基准上取得了从早期到先进的性能飞跃。SuperGLUE 由于其更具挑战性的任务，对模型的深层推理和常识能力提出了更高要求。
- **SQuAD：** 在问答任务 SQuAD 数据集上，基于 BERT 及其变体（如 ALBERT、ELECTRA）的模型，通过结合各种改进技术，取得了非常高的精确匹配（EM）和 F1 分数。例如，IE-Net 等集成模型在 SQuAD 2.0 上 F1 分数超过 93%。
- **WMT：** 在机器翻译领域，Transformer 架构本身就带来了革命性的进步，而后续的神经机器翻译（NMT）模型，如 T5 系列、ALMA 模型等，在 WMT 竞赛中持续取得领先地位。例如，ALMA-R 模型在 WMT'21、WMT'22 和 WMT'23 测试数据集上能够匹配或超越 WMT 竞赛获胜者和 GPT-4 的性能。
- **MMLU：** 对于评估模型在多学科知识理解和推理方面的能力，MMLU 基准显示，GPT-4 等大型语言模型取得了令人印象深刻的成绩，其准确率已经接近甚至超越人类专家水平。

表 3：代表性预训练语言模型在基准测试上的表现示例

模型系列	核心架构	典型应用	典型基准测试表现 (F1/BLEU/准确率)
BERT	Encoder-only Transformer	问答、文本分类、命名实体识别	SQuAD 2.0 F1 > 92% (变体)
GPT	Decoder-only Transformer	文本生成、摘要、对话	MMLU 准确率 ~86.4% (GPT-4)
T5	Encoder-Decoder Transformer	统一文本到文本任务（翻译、摘要、问答）	WMT 机器翻译 (高分)
ALMA	Decoder-only LLM (基于 Transformer)	机器翻译	WMT'21, '22, '23 匹配或超越 GPT-4

4．3 应用普及与社会影响

AI 正迅速从实验室走向日常应用，渗透到医疗、交通等多个领域。2023 年，美国食品药品监督管理局（FDA）批准了 223 款 AI 医疗设备，显示出 AI 在医疗领域的快速普及。在交通领域，自动驾驶汽车已不再是实验性技术：Waymo 每周提供超过 15 万次自动驾驶服务，而百度旗下的 Apollo Go 自动驾驶出租车队也已在中国多个城市投入运营。

负责任 AI（RAI）生态系统正在发展，但其发展并不均衡。AI 相关事件急剧增加，然而，主要工业模型开发者中，标准化的 RAI 评估仍然罕见。尽管如此，HELM Safety、AIR-Bench 和 FACTS 等新基准为评估 AI 的事实性和安全性提供了有前景的工具。各国政府也加大了对 AI 的投资和监管力度，2024 年，美国联邦机构出台了 59 项与 AI 相关的法规，是 2023 年的两倍多。全球范围内，自 2023 年以来，75 个国家立法提及 AI 的次数增加了 21.3%，自 2016 年以来更是增长了九倍。

5 总结与展望

5．1 报告主要内容总结

本报告全面回顾了基于深度学习的自然语言处理领域的研究进展。报告首先阐明了 NLP 的核心任务，并深入分析了当前深度学习 NLP 模型面临的挑战，包括数据稀缺性、语言歧义性、模型偏见、可解释性、计算资源限制以及鲁棒性问题。随后，报告详细介绍了该领域的方法进展，从深度学习在 NLP 中的发展历程，到 Transformer 及其变体、BERT、GPT 系列、T5 等核心模型架构的原理与影响。此外，报告还探讨了迁移学习与微调、注意力机制与关系学习、上下文学习与提示工程以及多模态学习等关键学习范式。最后，通过对比主流模型在各项基准测试上的表现，指出了当前技术的成就与仍需克服的可靠性挑战。

5．2 未来发展趋势与研究方向

尽管深度学习在 NLP 领域取得了显著进展，但其精度和实际应用需求之间仍存在差距。未来的研究可从以下几个方面进一步探索：

- **解决标签稀缺性与低资源语言问题：** 探索更有效的数据增强、跨语言迁移学习和无监督学习方法，弥补低资源语言的数据鸿沟，促进 NLP 技术的全球普惠性。
- **缓解模型偏见与提升公平性：** 发展更透明、可解释的 NLP 模型，不仅能提高用户信任，还能帮助研究者诊断和缓解模型中的偏见来源。
- **增强模型鲁棒性与泛化能力：** 开发更先进的对抗性训练方法和评估基准，以提升模型在面对真实世界复杂和对抗性输入时的鲁棒性。
- **构建高质量数据集：**鉴于当前数据集样本规模小、多样性低、标签稀缺且不均衡等不足，未来可以构建一个规模大、样本多样性丰富、标注全面的非受控环境数据集。鉴于人工标注成本高昂，将采用主动学习等方法，从少量人工标注数据开始，迭代训练模型并选择信息最丰富、包含低频 AU 的未标注样本进行人工标注，从而优化标注成本。
- **通用人工智能（AGI）的探索：** 随着 LLMs 在多任务、多模态和复杂推理方面的能力不断增强，NLP 将继续在通用人工智能的实现路径上扮演关键角色，例如通过更深层次的跨模态理解和更强的常识推理能力。

6 结论

人工智能正处于一个前所未有的快速发展阶段，其技术突破和应用普及正在深刻改变全球经济和社会格局。从复杂基准测试性能的显著提升，到生成式 AI 和多模态 AI 的广泛应用，AI 已从理论研究走向各行各业的实际部署，成为推动生产力增长和解决复杂问题的核心力量。

然而，AI 的快速发展并非没有挑战。数据稀缺性、特征捕捉难度、标签不均衡性等技术障碍，以及模型可解释性不足、训练数据偏见、伦理风险和就业替代等伦理和社会问题，都构成了 AI 持续健康发展的关键制约。这些挑战要求研究界和产业界不仅要追求技术前沿，更要关注 AI 的负责任发展。

展望未来，AI 将朝着更自主、更智能、更普适的方向演进，自主 AI 代理和无处不在的生成式 AI 将成为主流。同时，AI 在特定领域的数据挑战将通过多策略融合、鲁棒性提升和特征解耦等方式得到解决。为确保 AI 的长期福祉，持续的跨学科研究、建立健全的治理框架以及推动全球范围内的伦理共识至关重要。只有通过技术创新与负责任发展并重，AI 才能真正实现其造福人类的巨大潜力。

参 考 文 献

〔1〕邵志文, 周勇, 谭鑫, 等. 基于深度学习的表情动作单元识别综述[J]. 电子学报, 2022, 50(8): 2003-2017.

〔2〕Vaswani A, Shazeer N, Parmar N, 等. Attention is all you need〔C〕//Advances in Neural Information Processing Systems. 2017: 5998-6008.

〔3〕Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). BERT: Pre - training of Deep Bidirectional Transformers for Language Understanding.

〔4〕Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives.

〔5〕Goyal, A., & Khot, T. (2021). Multimodal Machine Learning: A Survey and Taxonomy

〔6〕Wang, A., Singh, A., Michael, J., Hill, F., Levy, O., & Bowman, S. R. (2018). GLUE: A Multi - Task Benchmark and Analysis Platform for Natural Language Understanding.

〔7〕Hendrycks, D., & Gimpel, K. (2021). Measuring Massive Multitask Language Understanding.

〔8〕Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier.

〔9〕Rajpurkar, P., Zhang, J., Lopyrev, K., & Liang, P. (2016). SQuAD: 100, 000+ Questions for Machine Comprehension of Text.

〔10〕Hendrycks, D., & Gimpel, K. (2021). Measuring Massive Multitask Language Understanding.