

# 中文农机大语言模型耒耜构建方法

栗晓宇<sup>1,2</sup> 王玉聪<sup>1,2</sup> 杜岳峰<sup>1,2</sup> 李国润<sup>1,2</sup> 刘磊<sup>1,2</sup> 宋正河<sup>1,2</sup>

(1. 中国农业大学工学院, 北京 100083; 2. 中国农业大学现代农业装备优化设计北京市重点实验室, 北京 100083)

**摘要:** 大语言模型拥有强大的生成、学习和推理能力,是加快现代农机装备绿色、智能和高效发展的强力助手。然而,由于缺乏用于训练的农业机械相关数据集,极大限制了大模型在农机研发、制造和推广。因此,面向农机装备科研人员、设计制造工程师、用户等不同群体具体需求,提出了中文农机大模型——耒耜。首先,提出了耒耜大模型总体架构设计方案,包括耒耜·薪火、耒耜·匠心和耒耜·耕耘 3 个版本,旨在为目标群体提供农机专业知识问答、设计制造建议和田间作业管控等多样化、定制化服务;其次,构建了国内首个中文农机数据集,并以耒耜·薪火大模型为例,阐明了模型训练和评估方法,分别以 LLaMA 3.1-8B-Instruct, Mistral-7B-Instruct-v0.3 和 Qwen 2.5-7B-Instruct 为基座模型进行监督式微调,利用 ROUGE 和 BLEU 作为评价指标评估微调后模型性能;最后,采用人工评估对 LLaMA 3.1、GPT-4o、Mistral、Qwen 2.5 和薪火大模型的问答结果进行评价,自动评估和人工评估结果表明薪火大模型在准确性、专业性和可用性等方面表现最优。研究成果为农机装备全生命周期管控及智慧农业发展提供了有力工具和手段。

**关键词:** 大语言模型; 农机装备; Qwen 2.5; QLoRA; 微调

**中图分类号:** TP391 **文献标识码:** A **文章编号:** 1000-1298(2026)05-0387-11

**OSID:**



## Construction Method of Large Language Model for Chinese Agricultural Machinery LeiSi

LI Xiaoyu<sup>1,2</sup> WANG Yucong<sup>1,2</sup> DU Yuefeng<sup>1,2</sup> LI Guorun<sup>1,2</sup> LIU Lei<sup>1,2</sup> SONG Zhenghe<sup>1,2</sup>

(1. College of Engineering, China Agricultural University, Beijing 100083, China

2. Beijing Key Laboratory of Optimized Design for Modern Agricultural Equipment, China Agricultural University, Beijing 100083, China)

**Abstract:** The development of large language models (LLMs) has significantly propelled the latest advancements in natural language processing (NLP). These models have been built upon complex deep learning architectures, typically Transformer, characterized by billions of parameters and extensive training data, enabling them to achieve high precision across a variety of tasks. However, the absence of agricultural machinery-specific textual data for training in existing general large models has severely limited their performance in the research, development, manufacturing, and application of agricultural machinery. To address this issue, the specific needs of agricultural machinery for large models were analyzed and a Chinese-compatible agricultural machinery large model named “LeiSi” was proposed, catering to various groups such as university faculty, students, designers, and users. The overall architectural design of the LeiSi LLM was outlined, which included three parts: LeiSi-torch, LeiSi-ingenuity, and LeiSi-plough, aiming to provide target groups with diversified and customized services such as agricultural machinery professional knowledge Q&A, agricultural machinery design and manufacturing advice, and agricultural machinery field operation control. Subsequently, the LeiSi-torch large language model was taken as an example to elucidate the construction of a Chinese agricultural machinery dataset and the methods for model fine-tuning and automatic evaluation. Utilizing LLaMA 3.1-8B-Instruct, Mistral-7B-Instruct-v0.3, and Qwen 2.5-7B-Instruct as base models, supervised fine-tuning on each was conducted and the performance of the fine-tuned models was evaluated by using ROUGE and BLEU as evaluation metrics. Finally, manual evaluation was employed to assess the Q&A

收稿日期: 2024-11-29 修回日期: 2024-12-25

**基金项目:** 国家自然科学基金项目(52175258)和中国博士后科学基金项目(2023M743790)

**作者简介:** 栗晓宇(1994—),女,讲师,博士,主要从事农机装备及其自动化研究,E-mail: lxy@cau.edu.cn

**通信作者:** 宋正河(1973—),男,教授,博士,主要从事智能农机装备设计与试验研究,E-mail: songzhenghe@cau.edu.cn

results of LLaMA 3.1, GPT-4o, Mistral, Qwen 2.5, and the LeiSi-torch model. The results from both automatic and manual evaluations indicated that the LeiSi-torch model demonstrated superior performance in terms of accuracy, professionalism, and usability. The research outcomes can provide insights and references for the development of intelligent agricultural machinery and smart agriculture.

**Key words:** large language model; agricultural machinery and equipment; Qwen 2.5; QLoRA; fine-tuning

## 0 引言

大语言模型(Large language models, LLMs)已成为近年来跨学科研究的焦点,随着 OpenAI 的 ChatGPT、清华智谱的 ChatGLM、Google 的 Gemini 等通用 LLMs 的兴起,各个领域正在经历重大变革<sup>[1-3]</sup>。这些模型借助深度学习和自然语言处理(Natural language processing, NLP)技术,显著提升了对文本、图像、数值等多模态信息理解与生成的效率,为广大用户简化了远程业务咨询,进而革新了行业工作流程和决策模式<sup>[4-5]</sup>。根据《2024 年中国大语言模型行业全景图谱》分析,预计至 2027 年,我国大语言模型市场的市值将攀升至 600 亿元人民币,而到 2029 年将增至 1 186 亿元人民币,年均复合增长率超过 40%,LLMs 对各个行业产生的深远影响将持续增长。

智能农机装备涵盖了数据、知识、模型、软件、硬件等多要素,为用户提供信息获取、智能决策、精准作业、个性服务等多元功能,大力支持高端智能农机装备、算法和模型的攻关被列为重点发展方向<sup>[6-7]</sup>。智慧农机装备作为智慧农业体系中的重要组成部分,目前存在核心技术受限、研发成本高和人才资源不足等问题,LLMs 在各个领域取得的系列成功为解决上述问题提供了解决方案,这一观点已在农业病虫害识别、杂草识别、果蔬采摘等研究方向得到验证<sup>[8-9]</sup>。

农机装备具有作业对象属性多样、多源跨域异构数据、作业场景复杂、长时序决策支持和实时可靠要求的特点,并且缺少农机专业数据集支撑,导致现有通用 LLMs 在解决农机领域的问题时不具有“专家”属性,生成答案过于笼统、浅显。为此,本文将 LLMs 引入农机领域,建立国内首个中文农机数据集,开发中文适应性好、专业性强、准确度高的中文农机大语言模型,命名为“耒耜”,以期填补我国农机装备领域垂直大模型研究空白。

## 1 相关研究

LLMs 在不同语言任务中表现出的高性能和通用性得益于可靠的语言模型研究基础。如图 1 所示,18 世纪 40 年代,统计学和信息论诞生使数据分

析、信号处理和通信等领域发生革命性变化,其演化进程按照模型形态和功能,大致划分为统计语言模型时代、神经语言模型时代、预训练语言模型时代和大语言模型时代。统计方法和 N-gram 模型是开发语言模型的早期尝试,但这些模型在用语言表达长期的相互依存关系和语境方面存在缺陷<sup>[10]</sup>。因此,研究人员以更复杂的方式探索神经网络的发展和更大数据集的可用性,Transformer 是 LLMs 的一个重要的里程碑,开发 Transformer 模型的目标是克服循环神经网络(Recurrent neural network, RNN)和长短期记忆网络(Long short term memory network, LSTM)等模型存在的梯度消失、梯度爆炸问题,以及难以捕捉上下文长时序依赖关系的固有约束,与其他模型相比,Transformer 模型可通过在 GPU 上进行并发训练以满足大参数量模型训练的需求<sup>[11-12]</sup>。

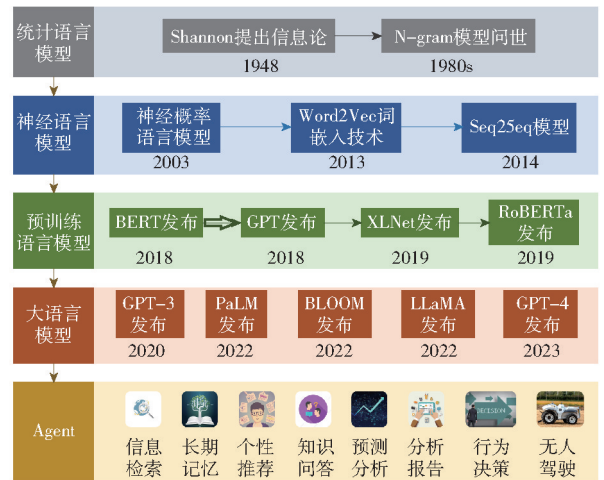


图 1 LLMs 发展历程及趋势

Fig. 1 LLMs development history and trends

2018 年,Google 公司基于 Transformer 架构的双向编码器,通过大量无标注文本进行预训练,开发了世界上第一款大语言模型 BERT<sup>[13]</sup>,随后 GPT-3、GPT-4、PaLM 等相继发布,LLMs 在各种 NLP 任务中表现出优越性能和巨大潜力<sup>[14-15]</sup>。2023 年,我国大语言模型技术取得了突破性成果,百度“文心一言”、阿里巴巴“通义千问”、讯飞科技“讯飞星火”等通用 LLMs 陆续发布,改变了人们的生产生活方式<sup>[16-17]</sup>。

预训练+微调已成为深度学习最常用范式,旨在将预训练模型丰富的知识和优越的学习能力迁移至目标任务,促进了 LLMs 在垂直领域中的应

用<sup>[18-21]</sup>。Google公司发布了Med-PaLM2医学大模型,其准确度达到92.6%,并建立了MedQA医疗问答数据集,是目前医学领域通用评估标准<sup>[22]</sup>;文献<sup>[23]</sup>基于LLaMA 2-7B构建了用于学习地理信息科学的大模型BB-GeoGPT,主客观问题回答准确度有明显提升;文献<sup>[24]</sup>提出了AgriBERT,旨在提供食品和农业领域知识问答任务;大连理工大学与中国联通合作提出了FinBERT金融大模型,其数据量达到127.1亿词;中国农业大学王耀君团队<sup>[25]</sup>发布“神农大模型2.0”,面向育种、种植、养殖和气象遥感等场景,开发了神农·固芯、神农·筑基、神农·强牧和神农·问穹4个农业专业大模型,显著提升了现代农业的智能化控制与决策效率;中国农业大学李道亮团队<sup>[26]</sup>发布我国首个渔业大模型“范蠡大模型1.0”,可为用户提供包括水质、饵料、疾病、管理等在内的多方面预测、分析和决策支持,标志着我国在智慧渔业领域迈出了重要的一步。上述研究成果促进了垂直领域产业模式的转变,完善了智慧农业定制化解决方案,但这些农业大模型是面向田间农田作物或鱼类相关知识服务,无法直接应用于农业

机械领域。

## 2 总体架构设计及其核心功能

### 2.1 需求分析

表1展示了当前主流通用LLMs的性能评估结果<sup>[27]</sup>。以ChatGPT为代表的通用LLMs在处理数学推理、编程、上下文理解等任务中表现出卓越的性能,其中GPT-4o具备多模态数据处理和零样本学习能力,能够在没有额外训练数据的前提下处理更加多样化的任务,有助于提高数据集生成质量。为验证通用LLMs在应对农机专业知识问答任务的性能,本文以GPT-4o为例,提出问题“你是一位有多多年经验的玉米联合收获机设计工程师,请你列举5个常见脱粒元件的种类,并说明它们的用途”,GhatGPT给出的答案如图2所示。GPT-4o输出的答案与“脱粒元件”不相关,说明其并不能完全理解“脱粒元件”的概念,这是由于GPT-4o预训练数据集中缺少农机专业知识语料所导致的“幻觉问题”<sup>[28]</sup>,从而突出了构建农机大模型时面临的两个主要挑战:

表1 通用大模型性能评估结果

Tab.1 Generalized LLMs performance evaluation results

序号	模型	发布时间	发布机构	参数量	最大支持上下文长度	综合评分
1	GPT-4o	2024-05	OpenAI	$5 \times 10^9$	$1.28 \times 10^5$ tokens	67.20
2	Gemini-1.5Pro	2024-05	Google	$1.75 \times 10^{11}$	$1.28 \times 10^5 \sim 1 \times 10^6$ tokens	68.19
3	LLaMA-3.1	2024-07	Meta	$4.05 \times 10^{11}$	$1.28 \times 10^5$ tokens	66.70
4	GLM-4-Plus	2024-08	北京智谱华章	$1.3 \times 10^{11}$	$1.28 \times 10^5$ tokens	69.64
5	GPT-4	2023-03	OpenAI	$1.8 \times 10^{12}$	$8 \times 10^3$ tokens	79.11
6	Ernie 4.0Turbo	2024-07	百度	$8 \times 10^3$		75.96
7	文心一言4.0	2024-09	百度	$1.3 \times 10^{10}$	$8 \times 10^3$ tokens	76.18
8	豆包	2024-05	字节跳动	$8 \times 10^{11}$	$1 \times 10^5$ tokens	75.96
9	Qwen 2.5-72B	2024-09	阿里巴巴	$7.2 \times 10^{10}$	$1.28 \times 10^5$ tokens	70.30
10	讯飞星火-4.0	2024-06	科大讯飞	$1.3 \times 10^9$		59.40
11	Mistral-7B-v0.3	2024-09	Meta	$8 \times 10^9$		67.11
12	天工-3.0	2024-04	昆仑万维	$4 \times 10^{11}$		61.16

你是一位有多多年经验的玉米联合收获机设计工程师,请你列举5个常见的脱粒元件的种类,并说明它们的用途

作为一位有多多年经验的玉米联合收获机设计工程师,常见的脱粒元件及其用途如下:

- 脱粒滚筒 (Threshing Cylinder):**  
脱粒滚筒是玉米联合收获机中最关键的脱粒部件之一。它通过高速旋转,利用滚筒表面上的硬齿或凸起与玉米穗的摩擦力,将玉米粒从玉米穗中脱离出来。脱粒滚筒通常分为若干段,并可调节速度和旋转方向以适应不同的作物和工作条件。
- 脱粒板 (Threshing Plate):**  
脱粒板安装在脱粒滚筒周围,起到辅助脱粒的作用。它与脱粒滚筒共同工作,通过提供额外的摩擦和碰撞力,进一步剥离玉米粒。脱粒板通常由耐磨材料制成,并根据不同的脱粒需求调节其间距和角度。
- 风筛系统 (Fan System):**  
风筛系统的主要功能是通过强大的气流将脱粒后产生的轻杂质(如秸秆碎片、尘土等)与玉米粒分离。它能够确保进入收集系统的玉米粒更加纯净。风筛的风速和气流方向通常可以调节,以优化杂质分离的效果。
- 筛网 (Sieves):**  
筛网用于对脱粒后的物料进行进一步筛选,分离出不同大小的杂质。筛网具有不同的孔径,可根据

图2 ChatGPT 4.0 回答农机专业问题示例

(1)对农机专业问题的理解能力和推理能力具有较高要求。GPT-4o在回答“请列举适用于谷物脱粒的几种脱粒元件”时,显然没有正确理解“脱粒元件”的定义,推理得到的答案不符合谷物收获机的专业知识,而期望的答案应包括类似“适用于谷物脱粒的脱粒元件包括钉齿式、板齿式、梳齿式……”的前提,类似“其中,钉齿式可用于大豆、小麦脱粒,其工作原理是……;板齿式可用于玉米、大豆脱粒,其工作原理是……”的主要内容,以及类似“以上是关于适合谷物脱粒的脱粒元件的回答……”的结论。

(2)缺乏用于中文农机大模型训练的基准测试数据集。庞大且高质量的垂直领域语料库为LLMs

Fig.2 Example of ChatGPT 4.0 answering questions in agricultural machinery specialty

执行下游任务提供了必需的知识和推理基础,满足 LLMs 对文本依赖性和语料质量的高要求<sup>[29]</sup>。然而,农机专业语料少而分散,包含大量专业术语、概念和理论,如拖拉机、收获机、发动机、离合器、滑移率等,涉及表格、公式、图像等多模态非结构化信息,课题组前期积累的农机设计文档、样机测绘数据、试验数据、仿真数据等存在来源复杂、格式不统一、样本不均衡等特点,均无法直接用于模型训练,相较于其它领域而言,中文农机数据集的构建是一项极具挑战性的工作。

## 2.2 总体架构

面向农机装备科研人员、设计制造工程师、用户等不同群体不同需求,将中文大模型划分为耒耜·薪

火、耒耜·匠心和耒耜·耕耘 3 个版本,利用农业机械专业教材、设计文档、试验数据、故障诊断报告和专家知识等丰富数据资源,形成农机领域的高质量监督微调数据集,基于高性能基座模型进行微调,从而构建专门服务于农机全生命周期和农机知识智能问答的农机中文大模型。其任务板块涵盖市场需求分析、设计方案推荐、加工制造决策、智慧运维管控、维修保养建议等多元领域,以期帮助农机专业师生在短时间内快速获得有效知识和专业问题解答,为农机研发人员提供更加专业的农机设计方案、可靠理论依据和精确预测评估结果,同时为农机用户自主提供农机关键性能指标检测、作业质量预测和运维管控专业决策。耒耜大模型总体架构如图 3 所示。

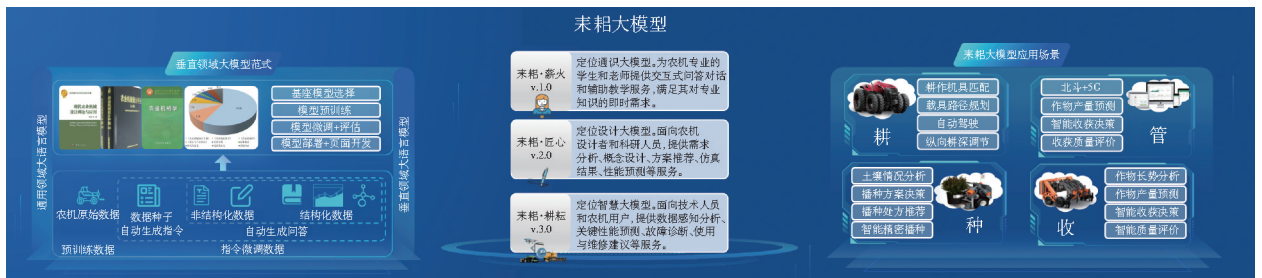


图3 耒耜大模型总体架构示意图

Fig. 3 Diagram of overall architecture of LeiSi

其中,薪火大模型 v1.0 主要面向农机专业背景的高校师生和不具备专业背景的用户,以交互形式辅助课堂教学和知识问答,将专业性强、权威性高、系统深入的学科知识构建为语义丰富、可解释性好、上下文关联性强的教育领域大模型;匠心大模型 v2.0 主要面向农机设计者和研究人员,增加文本、语音、图像等多模态信息分析系统,通过理解跨模态信息,输出更丰富、更准确、具象的设计方案(包括概念图生成、关键部件选型、方案评估、性能预测等),赋能农机智能设计制造;耕耘大模型 v3.0 以智能农机装备为载体,以思维链、思维树、增强现实、数字孪生、深度学习等技术为驱动力,构建赋能智慧农业生产系统全过程的多智能体,面向农机用户具体生产需求,集自主感知环境、自动构建孪生模型、自适应田间作业、自更新任务模型和自推理最佳解决方案等类人思考过程,为耕、种、管、收各个环节提供“泛在感知→超前预测→智慧决策”的闭环服务。

本文以耒耜·薪火大模型的构建过程为案例,从基座模型选取、数据集构建、模型微调和模型评估等 6 个方面详细阐述耒耜大模型构建方法,以期为后续行业垂直领域大模型的开发提供参考。

## 3 耒耜·薪火大模型构建方法

### 3.1 中文农机数据集构建

招募了 10 名具有农机专业背景的本科生和研

究生,先后开展两轮数据集构建工作,构建了农机中文监督微调数据集,以提高 LLMs 的专业信息生成能力。

#### 3.1.1 数据来源

农机领域的数据存在布局分散、系统性差、耦合性强、非结构化的固有特性,特别是公开数据集资源稀缺。虽然通用 LLMs 在应对从未见过的任务时生成的答案已经足够多样化,但农机大模型对答案的准确性、专业性和可用性有更高的要求。在策划数据集来源时,本文的主要目标是构建一个多样化且全面的与农机领域相关的数据集,以拖拉机、联合收获机为研究对象,涵盖专业知识、设计方法、试验标准、控制方法等内容,具体主要包括以下几个方面:

##### (1) 教科书

教科书通常是由领域专家和资深教师等专业人员编写,以教育为目的的重要资料。农机专业教科书系统性和规范性好,具有完整的知识体系和准确的语言规范,包含了与拖拉机、联合收获机结构、特性、应用等方面相关的专业知识语料,主要有《农业机械学》《汽车拖拉机学》和《液压与气压传动》等。

##### (2) 专业工具书

农机专业工具书是为农机领域设计工程师、科研人员和相关科技从业者提供参考和指导的书籍,

通常包含大量的技术细节、设计基础和操作指南,是确保不同研究方向的技术能够协同工作的关键资源。其中,《农业机械设计手册(上、下)》是最具代表性的农机专业书籍,其涵盖了拖拉机、整地机械、播种机械、种植机械、植保机械、各类作物收获机械等农机装备的基本概念、使用性能、核心装置、设计基础等有效信息,用以辅助农业机械化设计、生产、使用、教学及管理工。

### (3) 互联网

为了包含更多的农机专业语料,本文通过 Python 脚本从互联网搜集了大量与农机相关的专业信息,包括知网、万方数据、中国专利公布公告和国家标准信息公共服务平台等数据库,下载了 100 篇近 3 年引用次数最高的与农机设计相关的中文综述、研究论文、发明专利和标准规范作为农机中文数据集的外部知识源,以期将行业专家最新见解融入数据集,提高数据集丰富度和时效性。

#### 3.1.2 数据预处理

不同来源的语料包含大量长文本、表格和图像等非结构化数据,为构建格式统一、容量充足的高质量数据集,需要通过数据预处理操作剔除原始语料中的无用数据和冗余文本。

##### (1) 数据解析

纸质文本通过光学字符识别(Optical character recognition, OCR)转换为可编辑的 Word 文件。其中,图表中的文字和数据通过人工校对后转换为 JSON 格式的语料文本;公式使用 GPT-4o 转换为 LaTeX 格式,并手动生成逻辑正确、符合物理知识的语料文本,避免出现破折号、特殊字符、运算符号等内容。

##### (2) 数据清洗

使用 Python 脚本进行数据清洗,以去除数据集中 HTML 标签、重复、特殊符号、多余空格和换行符等不相关内容,剔除如“!”、“\n”、“#”、“\*”等无意义的占位文本,并由具有农机专业背景且持有学士及以上学位的审核人员进一步检查并修正拼写错误。

#### 3.1.3 SFT 数据集构建

在大语言模型微调任务中,通常使用特定的数据集格式来保证模型有效地进行学习。为了保证所有数据一致性,本文采用监督微调(Supervised fine-tuning, SFT)的标准化格式,每一条数据包含“指令”“输入”和“输出”3 部分,以 JSON 格式进行存储<sup>[30]</sup>,表 2 列举了部分 SFT 数据标准格式的示例:

(1)“指令”用于确认模型执行任务的属性,包括问题解答、文本生成、翻译等任务上下文关联。

(2)“输入”包含模型进行处理和理解的内容。

(3)“输出”是模型训练过程输出的“正确答案”,用于比较和调整模型预测输出与目标输出的差异。

表 2 SFT 数据集示例

Tab.2 Examples of SFT data

序号	指令	输出
1	如何选择农用拖拉机的驱动方式?	农用拖拉机的驱动方式一般有两种,分别是前驱动和后驱动。……选择驱动方式需要考虑具体的作业要求和场地条件,以实现最佳的作业效果和经济效益。
2	拖拉机的功率大小受哪些因素影响?	拖拉机的功率大小受农机品牌、发动机型号、气缸数、排气量、涡轮增压与否以及传动系统的设计等因素影响。一般来说,拖拉机的功率越大,适用于的作业范围和条件就越广泛。
3	农业机械的悬挂系统如何影响作业效果和稳定性?	农业机械的悬挂系统对作业效果和稳定性具有重要影响。……需考虑其结构形式、调节范围、阻尼性能等因素,以确保其能够满足作业需求。

原始语料包含复杂的信息,仅依靠人工制作“指令”中包含的问题或命令需要耗费大量时间和人力成本。GPT-4o 是 GPT-4 的优化版本,其生成文本在语法、语义和语言风格等方面更加接近人类,在各类基准测试中得分均位于 LLMs 前 10%,在理解和生成复杂文本方面更具有优势。另外,Open AI 公司 Altman 和 Sutskever 等提供了一种在短时间内获得充足数量的高质量“提示指令”的方法,即基于 GPT 强大的文本生成能力,诱导其开展续写任务,这一方法能够显著减少额外的人工成本,上述成果为中文农机数据集的构建提供了重要参考<sup>[31-32]</sup>。

在此基础上进行改进,具体步骤如下:

(1)设置先验指令。令 GPT 扮演农机设计工程师的角色,并提出生成任务指令的具体要求,如“只需要根据给定的答案输出任务指令,不需要输出输入和答案”等,以提高 GPT 指令生成效果,如图 4a 所示。

(2)在 GPT 理解先验指令后,可适当简化输入的指令为“给这段话生成一个专业的任务指令”等,同时利用 GPT 将收集的农机专业语料全部转换为单轮对话的形式,并保存为 JSON 格式,如图 4b 所示。数据集按照比例 9:1 划分为训练集和测试集,分别用于模型微调和评估。

基于此,考虑到现有的通用大语言模型普遍

你是一个拥有多年经验的农业机械设计领域的专家,主要面向拖拉机和谷物联合收割机。你被要求提供10个多样化的任务指令,以下是你提供指令需要满足的要求:

- 尽量不要在每个指令中重复关键词,或过度使用指令的多样性。
- 使用指令的语气应当结合给定的答案与祈使句。
- 指令类型应该是多样化的,并且紧密围绕拖拉机和谷物联合收割机的结构与评价。
- 指令中应当包括适当的提示,例如,“拖拉机的主要核心装置是什么?”
- 你只需要输出与指定答案文本相对应的指令,而不需要输出输入和输出的答案。

下面是任务指令的示例,你可以参考:

```
{
  "指令": "请列举5种谷物联合收割机常见的脱粒元件,并简单说明它们的用途";
  "输入": "";
  "输出": "谷物联合收割机脱粒元件主要有钉齿、板齿、揉搓式、纹杆式。1. 钉齿依靠冲击进行脱粒,具有较强的抓取能力和抗堵能力,适用于不均匀喂入、含水率高的作物,能够快速分离籽粒与杂质;分离效果差,容易造成籽粒破碎。2. 板齿靠刮清高度高,脱粒速度快,干净少杂质,适用于玉米、大豆等大粒作物脱粒作用;冲击作用大,易造成籽粒破碎。3. 揉搓式脱粒元件:这种元件模仿手工脱粒方式,依靠脱粒元件、凹齿与玉米果穗之间的互相搓擦作用来实现脱粒,揉搓式脱粒的效率较高,作业质量好,但其结构相对复杂,成本也较高。4. 纹杆式结构简单,具有良好的脱粒和分离能力,对籽粒产生的损伤相对较小,应用最为广泛,对不均匀喂入、含水率较高的作物脱粒效果较差。";
}
```

(a) GPT-4o先验指令格式示意图

现在你已经掌握了技巧,请按段落,给每一段文字都设计一个任务指令,并且整理为包含指令、输入、输出三个部分的JSON格式,其中输入为空:(1) 播种机的种子排布密度根据种类、播种深度和行距而定,小型手推式播种机的种子排布密度一般为10~20粒/平方米,适用于小面积农田的播种作业,中型拖拉式播种机的种子排布密度一般为20~40粒/平方米,适用于中等大小农田的播种作业,大型自走式播种机的种子排布密度一般为40~60粒/平方米,适用于大面积农田的播种作业。(2) 手推式播种机的播种深度一般为1~3厘米,适用于小面积农田的播种作业,拖拉式播种机的播种深度一般为3~5厘米,适用于中等大小的农田的播种作业,大型自走式播种机的播种深度一般为5~10厘米,适用于大面积农田的播种作业。(3) 选择合适的拖拉机轮胎需要考虑以下几个方面:首先,要根据拖拉机的型号和使用环境来确定轮胎的规格尺寸,包括轮胎的直径、宽度以及花纹类型。其次,要考虑轮胎的承载能力和速度等级,以确保在拖拉机工作过程中轮胎能够承受相应的负荷和速度。此外,轮胎的耐磨性、抗刺穿能力和抓地力也是选择时需要考虑的因素。最后,选择知名品牌、质量可靠的轮胎,能够提供更长的使用寿命和更好的性能表现。

以下息按表要求整理为包含指令、输入、输出三个部分的JSON格式,其中输入为空:

```
json
{
  "指令": "请分析播种机的种子排布密度与播种深度和行距之间的关系,并说明不同类型播种机的适用作业";
  "输入": "";
  "输出": "根据播种机的种类和农田的大小,种子排布密度会有所不同,手推式播种机适用于小面积农田,中型拖拉式播种机适用于中等大小农田,大型自走式播种机适用于大面积农田的播种作业。";
},
{
  "指令": "请比较不同类型播种机的播种深度与其与农田面积和作物类型之间的适应性。";
  "输入": ""
}
```

(b) SFT数据生成示意图

图4 SFT数据集构建方法

Fig. 4 SFT dataset construction methods

存在非忠实性、非事实性等幻觉问题,即看似合理但与问题无关或错误的内容,严重影响模型输出结果的精确性和可靠性。例如,图2中给出的问题为“请列举适用于谷物脱粒的几种脱粒元件”,给出的答案出现了“风机”、“筛网”等与联合收获机脱粒元件无关的词语。为了避免LLMs出现“幻觉”,对SFT数据集进行人工审核和清理,并手动纠正错误的内容。优化后得到高质量SFT数据总计28350条,JSON文件大小为11.02MB,来自《农业机械手册》收集的语料约占全部数据集的58%,容量最大;来自教科书的语料约占全部数据集的24%;来自互联网搜集或爬取的语料约占全部数据集的18%。

### 3.2 基座模型

为了实现高性能中文农机大模型的构建,综合考虑模型训练成本、硬件设施条件、原生模型性能等因素,本文选择LLaMA 3.1-8B-Instruct、Mistral-7B-Instruct-v0.3和Qwen 2.5-7B-Instruct作为基座模型,分别进行模型微调 and 评估,选择性能最优的模型构建薪火大模型。其中,LLaMA 3.1-8B-Instruct是美国Meta公司于2024年7月23日推出的指令调优版本,拥有 $8 \times 10^9$ 参数量,支持英文、德

文和法文等8种语言,一经发布便被誉为是世界上性能最好的开源大模型; Mistral-7B-Instruct-v0.3是法国Mistral AI公司于2022年8月7日发布的指令调优版本,拥有 $7.25 \times 10^9$ 参数量,在所有评估基准中超过当时性能最佳的开源模型LLaMA2; Qwen 2.5-7B-Instruct是中国阿里巴巴集团于2024年9月19日发布,位列全球开源大模型前三,具有RoPE、SwiGLU、RMSNorm和Attention QKV偏置的Transformers架构,拥有 $7.61 \times 10^9$ 参数量,并支持中文、英文、法文等超过29种语言,与LLaMA和Mistral相比具有更好的中文适应性。

### 3.3 微调

微调目前包括两大主流方向:全量微调和高效微调。进行LLMs的全量微调需要大量的计算成本和海量数据集,高效微调方法(PEFT)已被证明是微调大语言模型的有效范式,满足模型在面临新的训练任务时仅依靠小样本数据就可以表现出与原模型相近的性能,其中最具有代表性的是低秩自适应(Low-rank adaption, LoRA)微调,研究人员基于LoRA在RoBERTa、DeBERTa和GPT-X等通用LLMs上进行微调实验,验证了LoRA微调方法的有效性,LoRA也成为最广泛使用的微调方法。LoRA的核心思想是在模型的特定层中引入低秩矩阵,通过优化低秩矩阵来更新模型的权重,而不是优化整个权重矩阵,从而有效减少微调过程中参数数量和计算成本。

首先,假设预训练模型的权重矩阵 $W \in \mathbf{R}^{d \times k}$ ,通过低秩分解约束权重更新,即

$$W + \Delta W = W + BA \quad (B \in \mathbf{R}^{d \times r}; A \in \mathbf{R}^{r \times k}) \quad (1)$$

式中  $B, A$ ——低秩矩阵

$\Delta W$ ——模型权重在适应下游任务时的更新

$k$ ——LLMs线性层的输入维度

$d$ ——LLMs线性层的输出维度

$r$ ——LoRA线性层的秩,远小于 $d$ 和 $k$

其次,预训练模型的权重 $W$ 被冻结,模型梯度不再更新,而在矩阵 $A$ 和 $B$ 中包含可训练的全部参数,此时 $W$ 和 $\Delta W$ 与相同的输入相乘,并且各自的输出向量是逐坐标相加的,即

$$y = Wx + \Delta Wx = Wx + BAx \quad (2)$$

最后,使用随机高斯噪声初始化矩阵 $A$ ,矩阵 $B$ 初始化为零矩阵,则训练开始时 $\Delta W$ 的值为0,将 $\Delta Wx$ 进行 $\alpha$ 倍缩放,有助于减少重新调整超参数的需要<sup>[33]</sup>。QLoRA与LoRA不同之处在于,QLoRA引入了四位NormalFloat、双重量化和分页优化器,在保证模型性能前提下极大节省GPU内存的迫切需求<sup>[34]</sup>,QLoRA适配器的超参数如表3所示。

表3 QLoRA 适配器超参数

Tab.3 QLoRA adapter hyperparameters

类别	值/类型
batch_size	128
iterations	3
learning_rate	$1 \times 10^{-4}$
sequence_maximum_length	4 096
lora_rank	64
lora_alpha	16
lora_dropout	0.05
lora_target_modules	o_proj, up_proj, q_proj, gate_proj, k_proj, down_proj, v_proj
lr_scheduler	constant_with_warmup
warmup_ratio	0.1
double_quant	True
bits	4

### 3.4 模型微调

本文利用构建的中文农机数据集中的训练集对 LLaMA 3.1 - 8B - Instruct、Mistral - 7B - Instruct - v0.3 和 Qwen 2.5 - 7B - Instruct 进行 QLoRA 微调训练,训练损失曲线如图 5 所示。损失值是衡量微调模型输出与实际结果之间误差的关键指标。随着训练轮数的增加,训练的损失值也逐渐下降,并在 3 500 个迭代次数以内趋于稳定,说明模型基本收敛,训练效果趋于稳定,且 Qwen 2.5 的收敛效果优于 LLaMA 和 Mistral,损失值波动较小。模型微调需要高性能硬件支持,本文实验环境配置为 Intel(R) Xeon(R) Gold 6226R CPU、Tesla A100(内存容量 80 GB) GPU、Ubuntu 22.04 LTS 系统、CUDA 版本 12.1、Python 3.10。

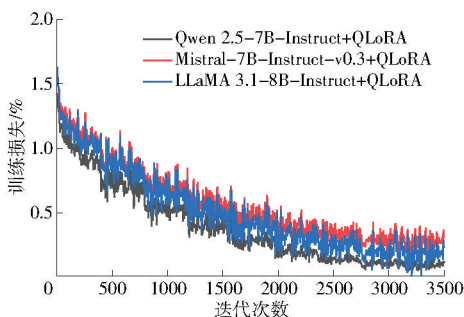


图5 模型微调损失曲线

Fig.5 Loss curves of models fine tuning

### 3.5 模型评估

ROUGE( $R$ )和 BLEU( $B$ )最初应用于评价自动摘要和机器翻译<sup>[35]</sup>。其中, $R_1$ 用于评估生成文本和参考文本之间单个词序列的重叠情况, $R_2$ 用于评估生成文本和参考文本之间连续两个词序列的重叠情况, $R_L$ 用于评估生成文本和参考文本最长公共子序列的长度。计算方法为

$$\begin{cases} R_N = \frac{\sum_{i=1}^n \text{count}(r_i \cap s_i)}{\sum_{i=1}^n \text{count}(r_i)} & (N \in \{1,2\}; n \in \{1,2\}) \\ R_L = \frac{2L(r,s)}{L_r + L_s} \end{cases} \quad (3)$$

式中  $r_i$ ——参考输出的第  $i$  个词序列

$s_i$ ——模型输出的第  $i$  个词序列

$\text{count}(r_i \cap s_i)$ ——在参考输出和模型输出中同时出现的第  $i$  个词序列的个数

$\text{count}(r_i)$ ——参考输出中第  $i$  个词序列的个数

$L(r,s)$ ——参考输出和模型输出之间最长公共子序列的词个数

$L_r$ ——参考输出的词个数

$L_s$ ——模型输出的词个数

$B_4$ 用于评估微调后模型输出答案与参考答案之间的连续 4 个词序列的相似度,计算方法为

$$B_4 = L_B \exp\left(\frac{\sum_{i=1}^4 \lg P_i}{4}\right) \quad (4)$$

其中

$$L_B = \begin{cases} 1 & (L_r > L_s) \\ \exp\left(1 - \frac{L_r}{L_s}\right) & (L_r \leq L_s) \end{cases} \quad (5)$$

式中  $L_B$ ——长度惩罚因子

$P_i$ ——将  $i$  个词作为整体的词序列精确度, %

随着 NLP 技术发展,ROUGE 和 BLEU 被应用于评价问答知识库检索系统、医学图像与生成报告之间匹配度、LLMs 生成答案精度等任务<sup>[36-37]</sup>。本文采用  $R_1$ 、 $R_2$ 、 $R_L$ 和  $B_4$ 作为评价指标,利用中文农机数据集中的测试集对比微调后 LLaMA 3.1 - 8B - Instruct、Mistral - 7B - Instruct - v0.3 和 Qwen 2.5 - 7B - Instruct 模型性能,评估结果如表 4 所示。结果表明,以 Qwen 2.5 - 7B - Instruct 微调模型的  $B_4$ 分数达到 88.04,其余指标均高于 LLaMA 和 Mistral,说明该模型输出和参考输出之间具有较好的语义相似性,并作为未粗·薪火大模型 v1.0 版本。

### 3.6 人工评估

上述自动评估指标能够比较所有模型性能,并且适用于各种生成任务,然而在实际应用场景中,人工评估结果也具有重要的参考价值<sup>[23,38]</sup>。为了进一步测试未粗·薪火大模型对农机专业知识的理解

表4 薪火大模型评估结果

Tab.4 Evaluation results of Torch LLM

模型	ROUGE			B <sub>4</sub>
	R <sub>1</sub>	R <sub>2</sub>	R <sub>L</sub>	
LLaMa 3.1-8B-Instruct-QLoRA	77.78	80.34	81.57	84.08
Qwen 2.5-7B-Instruct-QLoRA	84.15	85.46	88.56	88.04
Mistral-7B-Instruct-v0.3-QLoRA	79.28	81.33	84.62	86.75

能力,由5位农机专业教师(职称为副教授及以上)提出5个专业问题,从准确性、专业性和可用性这3个方面对5种模型生成的答案进行打分,模型得分范围从1(不可接受)到10(良好)。

表5展示了耒耜·薪火大模型和其他通用LLMs在人工评估中的结果,微调后的薪火大模型在农机问答任务中的人工评估得分最高,为9.3分,符合农机专业知识水平,可用性最好,专家接受程度最好;GPT-4o代码非开源,在线交互过程种融合了互联网搜索结果,同时GPT-4o对中文适配性进行了多轮优化,在回答问题时流畅性、可读性较好,人工评估结果为4.9分,但其输出内容达不到标准答案的专业要求;LLaMA 3.1和Mistral在不联网的情况下无法进行准确、专业的知识问答,这是由于其预训练数据集不包含中文,尤其是中文农机语料,因此不具备中文理解能力和推理能力,得分仅为1.4分和3.9分。

表6展示了不同大模型回答关于拖拉机专业问题时的示例。结果表明,直接使用通用LLMs输出的答案与正确答案相差过大,而经过微调和知识图

表5 耒耜·薪火大模型和通用LLMs的评测结果

Tab.5 Evaluation results of LeiSi LLM and other generic LLMs

类别	LLMs				薪火大模型
	LLaMA-3.1-8B-Instruct	GPT-4o	Mistral-7B-Instruct	Qwen 2.5-7B-Instruct	
	准确性	专家①: 2, 专家②: 2, 专家③: 2, 专家④: 2, 专家⑤: 1	专家①: 5, 专家②: 6, 专家③: 4, 专家④: 4, 专家⑤: 5	专家①: 4, 专家②: 4, 专家③: 4, 专家④: 4, 专家⑤: 4	
专业性	专家①: 1, 专家②: 2, 专家③: 2, 专家④: 1, 专家⑤: 1	专家①: 4, 专家②: 6, 专家③: 5, 专家④: 6, 专家⑤: 4	专家①: 4, 专家②: 5, 专家③: 4, 专家④: 4, 专家⑤: 3	专家①: 7, 专家②: 6, 专家③: 7, 专家④: 8, 专家⑤: 7	专家①: 10, 专家②: 9, 专家③: 9, 专家④: 9, 专家⑤: 8
可用性	专家①: 1, 专家②: 1, 专家③: 1, 专家④: 1, 专家⑤: 1	专家①: 5, 专家②: 6, 专家③: 5, 专家④: 5, 专家⑤: 4	专家①: 4, 专家②: 4, 专家③: 3, 专家④: 4, 专家⑤: 3	专家①: 8, 专家②: 7, 专家③: 7, 专家④: 8, 专家⑤: 7	专家①: 10, 专家②: 10, 专家③: 9, 专家④: 10, 专家⑤: 9
平均分	1.4	4.9	3.9	7.0	9.3

谱增强的薪火大模型所输出的结果与正确答案十分近似,效果较好,与Qwen 2.5-7B-Instruct相比具有更好的性能。综上所述,基于Qwen 2.5-7B-Instruct开发的耒耜·薪火大模型在准确性、专业性和可用性等方面主观评分最高,对农机专业知识理解正确,达到了基于小样本农机数据集开发高性能农机大模型的预期要求。

表6 不同大模型问答结果示例——以拖拉机为对象

Tab.6 Examples of QA results for different LLMs—taking tractor as object

模型	如何选择拖拉机轮胎胎压?	农具在拖拉机上悬挂的设计应满足哪些要求?
LLaMA 3.1-8B-Instruct	一般情况下,根据轮胎类型、载重量和地面状况,可选择适当的胎压。通常胎压应在3.5至4.5巴之间,以确保平衡安全性和防止轮辙迹形成。	当农业机械被悬挂在拖拉机上时,它们应该能提供稳定的工作,能够承受各类操作条件和负载。它们还需要根据具体情况进行调整以获得最好的效果,例如减少收割头或播种器的过度清洁以节省燃料并减少磨损。这是因为悬挂农业机械通常会导致拖拉机的重量增加以及由于减速而产生的额外阻力。
GPT-4o	确定拖拉机轮胎胎压是确保拖拉机安全、稳定运行和延长轮胎使用寿命的重要步骤。可以通过以下几个方法来确定拖拉机轮胎的正确胎压:1. 查阅用户手册每辆拖拉机通常都有制造商提供的用户手册,其中包括推荐的轮胎胎压。手册上会列出适用于不同载荷和工作条件下的胎压值。2. 轮胎侧面标识拖拉机轮胎的侧面通常会印有轮胎制造商推荐的最大胎压。请注意,这个值是轮胎承受的最大压力,不一定适用于所有情况……	农具在拖拉机上悬挂的设计应满足以下要求:1. 强度与稳定性:悬挂系统需具备足够的强度,以承受作业时的冲击和载荷,确保设备在行驶和作业过程中的稳定性。2. 调节性:应允许用户根据不同作业深度和地形进行调节,以适应各种农作物和土壤条件。3. 简便性:设计应便于安装和拆卸,减少更换和维护的时间和难度。4. 兼容性:悬挂系统需与多种类型的拖拉机和农具兼容,确保广泛适用性。5. 减震功能:应具备一定的减震性能,以减少对拖拉机和农具的损害,提高作业舒适性。6. 安全性:设计应考虑安全因素,防止在作业过程中发生意外,如脱落或失稳。满足这些要求可以提高农具的作业效率和安全性。

续表6

模型	如何选择拖拉机轮胎胎压?	农具在拖拉机上悬挂的设计应满足哪些要求?
Mistral-7B-Instruct	通常建议按照拖拉机制造商和轮胎生产商提供的推荐值来设置拖拉机轮胎的胎压。这些推荐值是基于车辆类型、使用方法、负载情况等因素而得出的,以确保最佳性能和安全性。	农具在拖拉机上悬挂的设计应该符合以下几个重要要求: 1. 安全性:农具必须可以稳定地安装在拖拉机上,并且在使用过程中不会发生意外。2. 适配性:农具应该能够与各种拖拉机型号兼容,并且可以快速安装和拆卸。3. 效率:农具的设计应该能够最大化地提高工作效率,减少损耗和浪费。4. 耐用性:农具应该具有良好的耐用性,可以在长期使用中保持高质量并且经受重复使用的挫折力。5. 维护易于:农具的设计应该尽可能地简单化,方便于维护和修复。6. 环境友好性:农具的设计应该能够减少对环境的破坏,并且尽量使用可 recyclable 的材料来降低污染风险。
Qwen 2.5-7B-Instruct	选择拖拉机轮胎的胎压需要考虑以下几个因素:轮胎类型;不同类型的轮胎(如高压、低压)对胎压的要求也不同,因此在选择时要根据具体情况来确定。道路条件:在不同的道路上行驶,所需要的胎压也会有所不同。例如,在泥泞的道路上行驶时,需要较大的胎压以提供足够的抓地力;而在高速公路行驶时,较小的胎压可以节省燃油消耗。车型大小:不同型号的拖拉机所需轮胎的规格也不同,因此在选择时要根据自己的车型来确定。气温条件:气温过高或过低都会影响轮胎的性能。在冬季,应适当降低胎压以防止轮胎打滑;而在夏季,可以稍微提高胎压以增加行驶速度和动力。需求:拖拉机主要用于农业作业,如种植、收割等。因此,选择适当的胎压可以使车辆在这些作业中保持良好的性能。综合以上因素,可以选择合适的拖拉机轮胎进行胎压调整,并且要定期检查胎压,确保其处于正常范围内。	农具在拖拉机上的悬挂系统设计需要满足多个方面的要求,以确保农业作业的高效、安全与适应性。以下是一些关键要求:1. 稳定性:农具必须稳定地安装在拖拉机上,避免在操作过程中发生摇晃或偏移,以免影响农作物的质量和安全性。2. 负载承载能力:系统需要能够承受预期的最大负载,包括农具自身重量、作业时产生的额外压力以及可能的载荷变动。这涉及材料选择、结构强度计算及安全系数设定。3. 灵活性与适应性:悬挂系统应能适应不同类型的农具(如播种机、收割机等),并能够调整以应对不同的作物种植密度和土壤条件,提高作业效率和效果。4. 操作便利性:农具的挂接和脱卸应当简便快捷,减少人力成本,并确保操作者安全。同时,需要提供良好的可视性和操作界面指导,方便工作人员进行设备管理与故障排查。……这些要求共同构成了一个高效、安全且经济的悬挂系统设计原则。在实际应用中,可能需要根据特定的需求和条件进行调整与优化。
耒耜·薪火大模型	农用拖拉机的轮胎胎压是根据载荷和作业场地情况来确定的。合适的轮胎胎压可以保证拖拉机在不同地形下有良好的牵引性能和操控性能,同时还可以减轻地面对轮胎的磨损和损坏。一般来说,载荷较重的情况下需要增加轮胎胎压,以提高轮胎的承载能力和稳定性;而在软弱的土壤或湿滑的路面上作业时,可以适当降低轮胎胎压,增加轮胎与地面的接触面积,提高牵引力和防滑性能。轮胎胎压的调整可以根据实际情况进行,需要注意轮胎胎压不要过高或过低,以免影响拖拉机的作业效率和安全性。	农具在拖拉机上悬挂的设计应满足以下要求:1. 农具应具有良好的人土性能,即在没有外力作用下,能平稳迅速地达到预定耕深,并且入土行程应尽可能短,这取决于入土角和入土力矩的大小;2. 设计应具有好的耕深稳定性和对不平地表的适应性,通过合理布置纵向垂直平面的瞬心位置以及选择合适的耕深调节方法来实现;3. 机组在工作时应具有良好的操纵性和牵引性能;4. 机组在行驶时应具备好的通过性和稳定性;工作时能保证直线行驶性和耕宽的稳定性,这需要正确选择水平面内瞬心的位置;5. 在正常作业时,液压装置油缸的负荷和轮胎的承载不应超过允许值;6. 装卸、操纵和调节应轻便,并且视野良好。

## 4 结论

(1)针对通用 LLMs 缺少农机专业数据集导致的回答内容专业性不强、泛化能力受限和模型推理成本高等问题,本文利用农机专业教科书、工具书、论文专利和标准规范等行业前沿成果,构建了高质量中文农机数据集,填补了农业领域中农机专业数据集的研究空白,同时为其他垂直领域数据集构建提供参考。

(2)本文提出的耒耜大模型是国内首个中文农机大模型,该模型包括耒耜·薪火、耒耜·匠心和耒耜·耕耘3个版本。利用中文农机数据集对 LLaMA、Qwen 和 Mistral 高性能基座模型进行指令微调,通过 ROUGE 和 BLEU 评价指标评估模型性能,结果表明以 Qwen 为基座模型具有最优性能,从而构建了耒耜·薪火大模型。此外,通过人工评估比较了耒耜·薪火大模型与 LLaMA-3.1-8B-Instruct、GPT-4o、Mistral-7B-Instruct 和 Qwen 2.5-7B-Instruct 在

准确性、专业性和可用性等方面表现,结果表明耒耜·薪火大模型在农机专业问题问答任务中具有显

著优势。研究结果可为农机装备设计和智慧农业管理提供科学、精准的决策支持。

### 参 考 文 献

- [1] 王耀祖,李擎,戴张杰,等. 大语言模型研究现状与趋势[J]. 工程科学学报, 2024, 46(8): 1411-1425.  
WANG Yaozu, LI Qing, DAI Zhangjie, et al. Current status and trends in large language modeling research[J]. Chinese Journal of Engineering, 2024, 46(8): 1411-1425. (in Chinese)
- [2] 李孝鹏,向玉云,张培君,等. 农业领域自然语言理解技术应用综述[J]. 农业机械学报, 2025, 56(10): 200-222.  
LI Xiaopeng, XIANG Yuyun, ZHANG Peijun, et al. Natural language understanding in agriculture: a comprehensive review of technologies and applications[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(10): 200-222. (in Chinese)
- [3] ROSSETTINI G, RODEGHIERO L, CORRADI F, et al. Comparative accuracy of ChatGPT-4, Microsoft Copilot and Google Gemini in the Italian entrance test for healthcare sciences degrees: a cross-sectional study[J]. BMC Medical Education, 2024, 24(1): 694.
- [4] WU Tangjie, LING Qiang. STELLM: spatio-temporal enhanced pre-trained large language model for wind speed forecasting[J]. Applied Energy, 2024, 375: 124034.
- [5] CHANDRASEKHAR A, CHAN J, OGOKE F, et al. AMGPT: a large language model for contextual querying in additive manufacturing[J]. Additive Manufacturing Letters, 2024, 11: 100232.
- [6] 翟长远,杨硕,王秀,等. 农机装备智能测控技术研究现状与展望[J]. 农业机械学报, 2022, 53(4): 1-20.  
ZHAI Changyuan, YANG Shuo, WANG Xiu, et al. Status and prospect of intelligent measurement and control technology for agricultural equipment[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(4): 1-20. (in Chinese)
- [7] 赵博,张巍朋,苑严伟,等. 农业装备运维与作业服务管理信息化技术研究进展[J]. 农业机械学报, 2023, 54(12): 1-26.  
ZHAO Bo, ZHANG Weipeng, YUAN Yanwei, et al. Research progress in information technology for agricultural equipment maintenance and operation service management[J]. Transactions of the Chinese Society for Agricultural Machinery, 2023, 54(12): 1-26. (in Chinese)
- [8] 刘成良,贡亮,苑进,等. 农业机器人关键技术研究现状与发展趋势[J]. 农业机械学报, 2022, 53(7): 1-22, 55.  
LIU Chengliang, GONG Liang, YUAN Jin, et al. Current status and development trends of agricultural robots[J]. Transactions of the Chinese Society for Agricultural Machinery, 2022, 53(7): 1-22, 55. (in Chinese)
- [9] 李道亮,赵晔,杜壮壮. 农业领域多模态融合技术方法与应用研究进展[J]. 农业机械学报, 2025, 56(1): 1-15.  
LI Daoliang, ZHAO Ye, DU Zhuangzhuang. Advances in multi-modal fusion techniques and applications in agricultural field[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(1): 1-15. (in Chinese)
- [10] JELINEK F. Continuous speech recognition by statistical methods[J]. Journals & Magazines, Proceeding of the IEEE, 2002, 64(4): 532-556.
- [11] XIE Weijun, ZHAO Maocheng, LIU Ying, et al. Recent advances in transformer technology for agriculture: a comprehensive survey[J]. Engineering Applications of Artificial Intelligence, 2024, 138: 109412.
- [12] ZHAI Yujia, JIANG Chengquan, WANG Leyuan, et al. ByteTransformer: a high-performance transformer boosted for variable-length inputs[C]//2023 IEEE International Parallel and Distributed Processing Symposium (IPDPS), 2023: 344-355.
- [13] DEVLIN J, CHANG M W, LEE K, et al. BERT: pre-training of deep bidirectional transformers for language understanding[J]. Association for Computational Linguistics, 2019, 1: 4171-4186.
- [14] CHANG Yupeng, WANG Xu, WANG Jindong, et al. A survey on evaluation of large language models[J]. ACM Transactions on Intelligent Systems and Technology, 2024, 15(3): 1-45.
- [15] KALYAN S K. A survey of GPT-3 family large language models including ChatGPT and GPT-4[J]. Natural Language Processing Journal, 2024, 6: 100048.
- [16] 张钦彤,王昱超,王鹤羲,等. 大语言模型微调技术的研究综述[J]. 计算机工程与应用, 2024, 60(17): 17-33.  
ZHANG Qintong, WANG Yuchao, WANG Hexi, et al. Comprehensive review of large language model fine-tuning[J]. Computer Engineering and Application, 2024, 60(17): 17-33. (in Chinese)
- [17] 王昀,胡珉,塔娜,等. 大语言模型及其在政务领域的应用[J]. 清华大学学报(自然科学版), 2024, 64(4): 649-658.  
WANG Yun, HU Min, TA Na, et al. Large language models and their application in government affairs[J]. J. Tsinghua Univ. (Sci. & Technol.), 2024, 64(4): 649-658. (in Chinese)
- [18] KASNECI E, SESSLER K, KÜCHEMANN S, et al. ChatGPT for good? On opportunities and challenges of large language models for education[J]. Learning and Individual Differences, 2023, 103: 102274.
- [19] 熊蜀峰,沈佳龙,许恒瑞,等. 主题特征与领域预训练模型融合的农业问句分类方法[J]. 农业机械学报, 2025,

- 56(10): 606 – 613.
- XIONG Shufeng, SHEN Jialong, XU Hengrui, et al. Agricultural question classification by fusing topic features with domain pre-trained models[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(10): 606 – 613. (in Chinese)
- [20] 薛飞跃, 周玉玲, 李俊凯, 等. 智慧养殖农业物联网与边缘计算中大模型技术应用综述[J]. 农业机械学报, 2025, 56(9): 291 – 311.
- XUE Feiyue, ZHOU Yuling, LI Junkai, et al. Application of foundation model technology for smart farming in edge computing and agricultural internet of things (Agri – IoT) [J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(9): 291 – 311. (in Chinese)
- [21] 刘壮, 刘畅, LIN Wayne, 等. 用于金融文本挖掘的多任务学习预训练金融语言模型[J]. 计算机研究与发展, 2021, 58(8): 1761 – 1772.
- LIU Zhuang, LIU Chang, LIN Wayne, et al. Pretraining financial language model with multi-task learning for financial text mining[J]. Journal of Computer Research and Development, 2021, 58(8): 1761 – 1772. (in Chinese)
- [22] SINGHAL K, AZIZI S, TU T, et al. Large language models encode clinical knowledge[J]. Nature, 2023, 620: 172 – 180.
- [23] ZHANG Yifan, WANG Zhiyun, HE Zhengting, et al. BB – GeoGPT: a framework for learning a large language model for geographic information science[J]. Information Processing & Management, 2024, 61(5): 103808.
- [24] REZAYI S, LIU Zhenglinag, WU Zihao, et al. AgriBERT: knowledge-infused agricultural language models for matching food and nutrition[C]// Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence AI for Good, 2022: 5150 – 5156.
- [25] 王耀君, 徐国威, 朱建军, 等. 农业领域大语言模型研究进展[J]. 农业机械学报, 2025, 56(9): 240 – 256.
- WANG Yaojun, XU Guowei, ZHU Jianjun, et al. Survey of research on large language models in agriculture[J]. Transactions of the Chinese Society for Agricultural Machinery, 2025, 56(9): 240 – 256. (in Chinese)
- [26] 新京报客户端. 我国首个渔业大模型“范蠡大模型 1.0”发布[EB/OL]. (2024 – 06 – 16)[2025 – 12 – 27]. <https://news.cau.edu.cn/mtndnew/98d7e0f834aa4f71aaa4d29a97b00290.htm>.
- [27] SHAHZAD T, MAZHAR T, TARIQ M U, et al. A comprehensive review of large language models: issues and solutions in learning environments[J]. Discover Sustainability, 2025, 6(1): 27.
- [28] LIN Zichao, GUAN Shuyan, ZHANG Wending, et al. Towards trustworthy LLMs: a review on debiasing and dehallucinating in large language models[J]. Artificial Intelligence Review, 2024, 57: 243.
- [29] YUAN Sha, ZHAO Hanyu, DU Zhengxia, et al. WuDaoCorpora: a super large-scale Chinese corpora for pre-training language models[J]. AI Open, 2021, 2: 65 – 68.
- [30] CHEN Jingyuan, WU Tao, JI Wei, et al. WisdomBot: tuning large language models with artificial intelligence knowledge[J]. Frontiers of Digital Education, 2024, 1: 159 – 170.
- [31] TAORI R, GULRAJANI I, ZHANG T, et al. Alpaca: a strong, replicable instruction-following model[J]. Stanford Center for Research on Foundation Models, 2023, 3(6): 7.
- [32] 王婷, 王娜, 崔运鹏, 等. 基于人工智能大模型技术的果蔬农技知识智能问答系统[J]. 智慧农业(中英文), 2023, 5(4): 105 – 116.
- WANG Ting, WANG Na, CUI Yunpeng, et al. Agricultural technology knowledge intelligent question-answering system based on large language model[J]. Smart Agriculture, 2023, 5(4): 105 – 116. (in Chinese)
- [33] DING Ning, QIN Yujia, YANG Guang, et al. Parameter-efficient fine-tuning of large-scale pre-trained language models[J]. Nature Machine Intelligence, 2023, 5: 220 – 235.
- [34] DETTMERS T, PAGNONI A, HOLTZMAN A, et al. Qlora: efficient finetuning of quantized LLMs[J]. Advances in Neural Information Processing Systems, 2024, 36: 1 – 26.
- [35] LI Junlong, WANG Jinyuan, ZHANG Zhousheng, et al. Self-prompting large language models for zero-shot open-domain QA [J]. arXiv preprint arXiv: 2212.08635v3, 2024.
- [36] SRIVASTAVA A, MEMON A. Toward robust evaluation: a comprehensive taxonomy of datasets and metrics for open domain question answering in the era of large language models[J]. IEEE, Journals & Magazines, 2024, 12: 117483 – 117503.
- [37] REN Runtao, MA Jian. PatentGPT: a large language model for patent drafting using knowledge-based fine-tuning method[J]. arXiv preprint arXiv: 2409.00092, 2024.
- [38] ZAFAR A, SAHOO S K, BHARDWAJ H, et al. KI – MAG: a knowledge-infused abstractive question answering system in medical domain[J]. Neurocomputing, 2024, 571: 127141.