2025届高三部分重点中学12月联合测评

语文试题（答案在最后）

考试时间：2024年12月12日9:00~11:30　　试卷满分：150分　　考试用时：150分钟

注意事项：

1.答卷前，考生务必将自己的姓名、准考证号填写在答题卡上。

2.回答选择题时，选出每小题答案后，用铅笔把答题卡对应题目的答案标号除黑。如需改动、用橡皮擦干净后，再选涂其他答案标号。回答非选择题时，将答案写在答题卡上。写在本试卷上无效。

3.考试结束后，将本试卷和答题卡一并交回。

一、现代文阅读（34分）

（一）现代文阅读Ⅰ（本题共5小题，18分）

阅读下面的文字，完成下面小题。

材料一：

①近年来，人工智能技术与人文历史研究碰撞出不少火花，那么，AI对于古代汉语的理解能力怎么样？

②日前，北京师范大学中文信息处理与古典文献学专业的研究者们尝试训练一个“饱读诗书”的语言模型，并让AI来参与古籍整理的工作。模型学习的对象包括《四库全书》与“殆知阁”语料库，而学习的方法则来自语言智能领域最新的深层语言模型。

③2018年，谷歌公司曾推出了深层语言模型BERT，它在阅读理解等11项语言理解任务中刷新纪录。和之前的方法相比，深层语言模型一是可以吞吐超大量的数据，二是有很强的记忆和理解能力。现有的深层语言模型覆盖了英文和中文，但却不具备理解古汉语的能力。于是，研究者们希望通过上述海量的古汉语数据来让机器“感受”一下博大精深的诗书礼乐文化。在多块计算卡上并行训练了约一周时间后，古汉语BERT初出江湖，AI可以像人一样联系上下文理解字词含义，并将其以数学向量表示。

④为了检测其理解效果，研究者们引入了句读任务。在古典文史学习过程中，句读通常是必备的基本功。古文句读不仅需要考虑当前文本的语义和语境信息，还需要综合历史文化常识，对古汉语知识有较高要求。宋代大儒朱熹读韩愈文章，便有“然不知此句当如何读”之感。在句读的过程中，有三项重要的技能点：利用古汉语特有的节奏和韵律感，联系上下文语境信息推敲求解，调用文本之外的历史文化知识。

⑤虽然现在很多古代经典都出版了标点本，但其中常常包含错误，并且，在现有的古籍数据中，大部分文献仍未实现句读。据统计，“殆知阁”古代文献藏书2.0版语料库规模约33亿字，其中仅25%左右的数据包含标点。如果依靠人工继续整理这些古籍，则不知何年何月才能整理到头。如果依靠计算机，现有的技术方法却普遍只能达到60%～70%的准确率，还很难为人所用。

⑥为了让AI在理解文义的基础上具备句读功能，研究者们准备了大量带标点的数据，包括超过30万首古诗，2万余首词，800多万段古文，模型还引入了处理标签序列的机制来专门学习句读方法。由于诗词具有较为明显的格律特征，如大部分古诗为四五七言，而词牌名可以提示断句规则，为了帮助模型更好地学习语义和韵律信息，在预处理数据时保留了古诗题目，并去除词牌名。

⑦在多块计算卡上并行学习了数天句读后，模型终于“出山”。在测试环节，研究者们引入了两轮难度不同的实验。普通版测试对象为一批模型从来没见过的诗词古文数据（古诗5000首，词2000首，古文5000段），实验中，模型在诗、词、古文的句读准确率上分别超过了99%、95%和92%，较之已有的自动断句方法取得了巨大提升。升级版测试对象为已出版古籍中的句读疑难案例。根据司马朝军等学者研究，从中华书局出版的《钦定四库全书总目》《周礼正义》中抽取出了60则句读误例（排除了模型训练时见过的数据）。这两本古籍均由该领域专家完成整理和句读标点，并经多次校对，其中的误例可谓句读任务的难点所在。研究人员从司马朝军的研究中找出了《总目》中11则与句读相关的错误，发现模型可以完全做对8则。又选择了王文锦等人点校的《周礼正义》一书，将颜春峰等人整理的49则断句误例送入模型测试。其中，模型能完全正确断句27则，断句存在问题的有22则。

⑧总的来说，模型在已出版古籍的断句疑难误例上取得了很好的效果，测试共计60例（均为专家标点错误，并经多次校对未查出），而模型竟能完全正确断句35例，可以说达到了较为实用的水平。

⑨从应用角度看，这里的句读方法既可以用于大规模古籍整理中预断句工作，大大减轻专家负担，也可用于校对环节，帮助检测人工断句或标点的错误。在后续工作中，除了提升已有的句读模型，还希望将基于深层语言模型的古汉语知识表示方法应用到古文翻译、古诗文创作等其他古汉语信息处理任务中。

（摘编自胡韧奋、李绅、诸雨辰《让AI一口气读完《四库全书》会怎样？》）

材料二：

2024年8月27日，北京师范大学古汉语大语言模型“AI太炎2.0”发布会暨数智时代应用语言学学科建设路径与方法座谈会在京举行。

“AI太炎”是国家语言文字工作委员会重大项目“古籍整理智能化关键技术研究”的核心成果，是专门适用于古汉语文本理解的大语言模型。该模型是针对古汉语信息处理任务的特点，从头构建的专门适用于古汉语文本理解的大语言模型。该模型具有较强的古典文献释读能力，支持词义注释、文白翻译、句读标点、用典分析等多种具有挑战性的文言文理解任务，且兼容繁简中文输入。此外，在辅助古籍整理、辞书编纂和语言研究等方面，该模型也表现出了很大的应用潜力。

据悉，“AI太炎1.0版”于2023年11月进入内测阶段，学术科研、基础教育、编辑出版等不同行业领域的4000余名用户参与内测，为模型改进提出了宝贵建议。研究团队在此基础上对模型进行了持续迭代优化，并发布了该模型2.0版，以期更好地助力古籍整理、文言文教学和古汉语信息处理研究。

（摘编自中国教育报《古汉语大语言模型“AI太炎2.0”发布会在京举行》）

1. 下列对原文相关内容的理解和分析，不正确的一项是（　　）

A. 以往语言模型无法吞吐巨量的数据，也没有强劲的记忆和理解能力，深层语言模型在这两方面具有优势。

B. 研究者在预处理古诗文数据时保留了诗题，并去除词牌名，旨在帮助模型更顺利地学习语义和韵律信息。

C. 在第一轮的普通版实验中，AI 模型在词体的句读准确度上表现最好，而诗体、古文次之，三者差别不大。

D. 基于科研、教育等不同领域用户的参与以及他们对模型提出的改进建议，“AI 太炎 2.0 版”得以顺利发布。

2. 根据材料内容，下列说法正确的一项是（　　）

A. 研究者们之所以让语言模型解决句读问题，是为了用客观的方式测试模型的理解能力，考查其在古典文史阅读中的必备技能。

B. 现在很多古代典籍都有了标点本，但总体上错漏较多。因此现有的识别技术方法的准确率并不理想，依然很难为人们所使用。

C. 根据 AI 模型在古代典籍数据库的断句实战上的优秀表现，可以认为实验取得了很好的效果，AI 模型已经超越了专家的水平。

D. 古典文献释读是“AI 太炎”模型具有的重要功能，这一模型能应对多种具有挑战性的古文理解难题以及语言研究方面的困境。

3. 下列案例中，哪一项最不适合作为论点“语言模型具有句读三项重要的技能点”的论据（　　）

A.“AI 太炎”在处理“非曰能之愿学焉宗庙之事如会同端章甫愿为小相焉”一句句读时，能根据古代的文化常识将后半部分的“端章甫”保留而不断开。

B.“AI 太炎”在处理“昨日蒙教窃以为与君实游处相好之日久而议事每不合”一句句读时，能根据古代的书信格式要求在“教”“久”处而非在“君”处断开。

C.“AI 太炎”在处理“四面边声连角起千嶂里长烟落日孤城闭”一句句读时，能根据《渔家傲》的词牌特点在“起”“里”处断开而非在“连”“里”处断开。

D.“AI 太炎”在处理“雕园绿竹气凌彭泽之樽邺水朱华光照临川之笔”一句句读时，能根据骈文“四六成句”的特点正确地在“竹”“樽”“华”处分别断句。

4. 小刚在阅读上述材料后，绘制了一幅思维导图。请结合材料内容，下列对思维导图的解说正确的一项是（　　）

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：https://d.book118.com/468014133010007012