

目录

1 摘要	1
2 理论基础	1
2.1 大语言模型	1
2.2 下游任务	1
2.2 提示工程	2
3 实验过程	2
3.1 文本分类-情感分析	2
3.2 命名体识别	5
3.3 文本生成	7
3.3.1 文本摘要	7
3.3.2 文本续写	8
3.4 文本纠错	9
3.5 逻辑推理	11
4 总结	11
5 参考文献	12

NLP 语言大模型下游任务性能对比

——自然语言处理的深度学习报告(第五次)

崔潇雅 ZY2203502

1 摘要

本实验通过提示工程的方法比较三种大语言模型在不同下游任务上的表现，本实验使用了三种常见的大语言模型：**chatgpt3**、讯飞星火大模型、天工大模型；与 5 种下游任务：中文文本分类、命名体识别、文本生成、中文文本纠错、逻辑推理。最后使用 **chatgpt4** 分别对不同模型下游任务上的输出按照统一标准进行打分分析。结果表明，天工模型在中文分类任务、文本生成上表现更好，但在逻辑推理上，星火模型则更优。

2 理论基础

2.1 大语言模型

GPT-3 是由 OpenAI 公司于 2020 年 6 月发布的一种基于深度学习的自然语言处理模型，它是前一代模型 GPT-2 的升级版。GPT-3 使用了最新的架构和训练技巧，其参数量高达 1750 亿，是目前公认的最大的语言模型之一。它可以自动完成各种任务，包括文本生成、对话系统、语音识别等。与以往的模型相比，GPT-3 在多个领域都有出色表现，甚至可以创作类似于人类写作的文章。

星火大模型是由讯飞公司开发的一种中文预训练语言模型，其参数量高达 270 亿。该模型采用了 Transformer 网络结构，并使用了大规模语料库进行训练，取得了较好的效果。它可以支持多种自然语言处理任务，包括智能客服、机器翻译等。此外，星火大模型还具备强大的文本生成能力，可以自动生成新闻报道、推荐信、论文摘要等。

天工大模型也是由讯飞公司开发的一种中文预训练语言模型，其参数量高达 100 亿。该模型主要面向自然语言生成领域，可以支持多种应用场景，包括文章创作、自动摘要等。与其他模型相比，天工大模型具有更好的可控性和可读性，可以生成更具有条理性和逻辑性的文本。此外，天工大模型还可以基于用户输入快速生成符合需求的文本，从而提高工作效率。

2.2 下游任务

大语言模型的下游任务可以分为三类：分类任务、序列标注任务和生成任务。

文本分类是自然语言处理中最常见的任务之一，它的目标是将一段文本划分到不同的类别中。通常使用大语言模型对输入文本进行编码，得到一个固定长度的向量，然后将其输入到分类器中进行分类。常见的文本分类任务包括情感分析、新闻分类、垃圾邮件过滤、主题分类等。

序列标注任务的目标是对输入的序列进行标注，例如命名实体识别、词性标注、实体关系抽取等。通常使用大语言模型对输入序列进行编码，然后将其输入到序列标注模型中进行标注。在命名实体识别任务中，大语言模型会学习如何将文本中的实体、人名、地名等标注出来。在词性标注任务中，大语言模型会学习如何将文本中的每个单词都标注上相应的词性，例如动词、名词等。

生成任务的目标是使用大语言模型生成自然语言文本，例如文本摘要、机器

翻译、对话生成等。通常使用大语言模型对输入文本进行编码，然后生成一段新的文本。在机器翻译任务中，大语言模型会学习如何将源语言翻译成目标语言。在文本摘要任务中，大语言模型会学习如何从一篇较长的文章中抽取出关键信息来生成一个简短的摘要。在对话生成任务中，大语言模型会学习如何根据用户输入生成回复，实现与用户之间的交互。

2.2 提示工程

提示工程（Prompt Engineering）是指对自然语言生成模型输入的文本进行调整，以产生所需输出的一种方法。这种方法可以帮助提高自然语言生成模型的输出质量和效率，促进生成内容与用户需求的匹配。 以下是一些常见的提示工程方法： 1. 关键词提示：在输入文本中加入关键词或提示词，使模型能够更准确地理解用户需求，生成符合要求的内容。 2. 样本提示：提供与所需输出相似的样本，让模型学习并模仿样本中的输出结果。 3. 策略提示：设置生成策略，引导模型生成特定类型的内容，如具有情感色彩、专业术语等。 4. 条件提示：通过设定条件，限制模型生成的内容范围，从而提高内容的准确性和可读性。 5. 交互式提示：与用户进行交互，根据用户反馈不断优化提示内容，直到生成满足用户需求的内容为止。 提示工程方法的选择应该根据具体应用场景和需求来确定，同时还需要考虑数据量、模型结构、训练集等因素。

3 实验过程

3.1 文本分类-情感分析

请判断以下文本的情感倾向并给出逻辑链分析：
实在不行，一个能触动我的场景都没，花木兰甚至还有 *Mulan rides into battle* 那一场，这个真的……从头到尾感觉海底世界就那么几个人鱼，竟然还自称王国，纯纯搞笑。片头引用安徒生的原文意义不明，环保议题种族议题各个都想参一脚，却又各个浅尝辄止。重头戏全部照抄动画，真的一丢丢想象力都没有，终场大战甚至抄都抄不好，把放大后的乌苏拉整的像个雕塑，也不知道是不是特效部门在赶工。另外这个姑娘是真的不上镜，每部迪士尼公主片都必有的展示主角魅力的 *glamour shots*（比如阿拉丁里茉莉走下皇宫台阶、美女与野兽里贝尔捧着书走过花丛）在这片里也是一个都没，你就感觉这主角不像是童话里的公主，反而一看到她的脸立马被拉回现实。话说这是不是画面色彩最不迪士尼的片子了，这么写实，咋给孩子造梦？

Chatgpt3	星火	天工
这段文本的情感倾向是负面的。 逻辑链分析如下： - 首先，作者提到自己看了这个电影后没有受到触动，而且觉得这个海底世界很无聊，只有几个人鱼，不像一个真正的王国，让人感觉很幼稚搞笑。 - 其次，作者批评这个电影在引用安徒生时没有表	情感倾向：负面 分析：该文本对电影进行了非常严厉的批评，认为电影缺乏想象力和创造力，场景、角色和特效都十分平庸，甚至有些搞笑。同时，文本也指出了电影中存在的一些问题，如环保和种族议题没有得到深入探讨，片头引用的安徒生原文意义不明等。最后，文本还提到了主角的形象设计不够出彩，缺乏展示魅力的机会。因此，可以判断该文本的情感倾向是负面的。	这段文本的情感倾向可以描述为负面和失望。作者对于这部电影的评价非常低，认为它缺乏想象力、情节平淡无奇、场景单一、人物形象不够吸引人等。作者还指出了该电影的一些问题，如引用的安徒生原文意义不明、环保议题种族议题浅尝辄止、画面色彩过于写实等。在逻辑链分析方面，这段文

<p>达出什么意义，同时也提到它在环保议题和种族议题等方面只是浅尝辄止。</p> <ul style="list-style-type: none"> - 然后，作者指出这个电影完全没有想象力，甚至重头戏都是照着原版动画抄的，而终场大战的特效也很粗糙。 - 最后，作者认为这个主角不上镜，缺乏迪士尼公主片里展示主角魅力的关键镜头。此外，作者还指出这个电影的色彩与迪士尼一贯的作品风格不同，过于写实，也不适合孩子们做梦。 		<p>本可以分为三个部分：</p> <ol style="list-style-type: none"> 1. 开头部分：作者表达了对这部电影的失望和不满，认为它没有任何令人印象深刻的场景或角色，并且引用的安徒生原文意义不明。 2. 中间部分：作者详细阐述了该电影的各种问题，包括情节平淡无奇、场景单一、人物形象不够吸引人、环保议题种族议题浅尝辄止、画面色彩过于写实等。 3. 结尾部分：作者总结了自己对该电影的看法，认为它缺乏想象力，难以给孩子们带来梦想。 <p>总的来说，这段文本的逻辑链分析较为完整，从开始的失望到中间的详细阐述，再到最后结束的结论，逻辑清晰，表达准确。</p>
7	8.5	9

模型 1 的输出得分：7 分。该模型能够描述文本的情感倾向和一些问题，但输出有点简略，没有对问题进行详细的分析。此外，模型没有提供任何建议或总结。

模型 2 的输出得分：8.5 分。该模型能够准确描述文本的情感倾向和逻辑，并且提供了详细的批评和问题分析，指出电影缺乏想象力、场景平庸、角色无趣、特效粗糙等问题，并且对环保和种族议题的探讨不够深入也提出了批评意见。最后，模型还提到了主角形象设计的问题。然而，该模型没有解释这些问题的原因，例如为什么场景和角色平庸，或者为什么主角形象设计不够出彩。

模型 3 的输出得分：9 分。该模型能够准确描述文本的情感倾向和逻辑，提供了详细的问题分析，从缺乏想象力、情节平淡无奇、场景单一、人物形象不够吸引人等多个角度对电影进行了批评，并指出了环保议题、种族议题等方面存在的问题。此外，模型还提到了画面色彩过于写实等方面。最后，该模型总结了作者对电影的看法，认为它缺乏想象力，难以给孩子们带来梦想。整体而言，该模型的逻辑分析完整，表达准确，没有明显的缺陷。

请判断以下文本的情感倾向并给出逻辑链分析：

爱丽儿的选角是点睛之笔，女主的能量感非常强，完美诠释了格兰芬多的一切品质，东亚真的非常需要这样的女主来调理一下审美。不是只有所谓高颜值的人才魅力，才配被爱，很多人攻击女主时声称不是因为人种，是因为她不好看，请问女主为什么必须好看？女人为什么必须好看？女性人鱼为什么必须好看？mermaid 这个词只是女性人鱼的意思，“美人鱼”

只不过是国内厌女文化导致的一个错误翻译。如果从这样一个勇敢、聪慧，热爱自由，充满能量的女性角色身上感受不到任何魅力，甚至觉得她被爱上都是无稽之谈，那是观众自己该去调理的地方。电影最大的遗憾就是依然没有摆脱父的阴影，女主最终得到自由的方式依然是通过父的允许和包容。女主打败充满权力野心的乌苏拉，却将权杖归还给了父亲，我本来以为王位会给姐姐继承，爱丽儿奔向自由，太遗憾了。

Chatgpt3	星火	天工
<p>此段文本的情感倾向是正面的。逻辑链分析如下：</p> <ul style="list-style-type: none"> - 首先，作者赞扬了《爱丽儿》电影女主角的选角很棒，因为她完美诠释了格兰芬多的品质，而且能量感十足。 - 其次，作者反驳了对女主角外貌的攻击，并且指出女性人鱼不一定要长得好看，甚至是文化误区导致了“美人鱼”这个错误翻译。同时，作者认为如果观众从这样一个勇敢、聪慧、热爱自由并充满能量的女性角色身上感受不到魅力，那是观众自己需要调整的地方。 - 然后，作者提到电影最大的遗憾是女主仍然没有摆脱父亲的阴影。虽然女主打败了权力野心的乌苏拉，但最终得到自由的方式依然是通过父亲的允许和包容。作者也表达了对女主最终将权杖归还给父亲的失望和疑惑，以及原以为王位会给姐姐继承的期待。 	<p>情感倾向：积极</p> <p>判断结果：该文本表达了对电影中女主角的爱丽儿的积极评价和认可，认为她是点睛之笔，能够完美诠释格兰芬多的一切品质，具有很强的能量感和魅力，并且认为女性不一定要有高颜值才能被爱和被认可。此外，文本还指出电影仍然存在一些不足之处，如未能完全摆脱父的阴影等。</p>	<p>这段文本的情感倾向可以描述为积极和支持。作者对于电影中女主角的表现给予了高度评价，认为她的能量感非常强，完美地诠释了格兰芬多的品质。作者还提到了对于女主角是否应该好看的质疑，以及对于厌女文化的批判，呼吁观众不要以貌取人。同时，作者认为女主角最终得到自由的方式依然是通过父的允许和包容，这也让他感到遗憾。</p> <p>在逻辑链分析方面，这段文本可以分为三个部分：</p> <ol style="list-style-type: none"> 1. 开头部分：作者表达了对女主角的喜欢，称赞她的能量感和勇敢、聪慧、热爱自由的品质。 2. 中间部分：作者提出了对于女主角是否应该好看的质疑，并批评了厌女文化。 3. 结尾部分：作者表达了对女主角最终得到自由的方式的遗憾，认为王位应该给姐姐继承，而不是给女主角。 <p>总的来说，这段文本的逻辑链分析较为完整，从开始的肯定女主角开始，逐渐表达了对女主角的喜欢和对厌女文化的批判，最后得出了对于结局的遗憾。</p>
6	8.5	9.5

模型 1 的输出得分：6 分。该模型能够描述文本的情感倾向，但输出有点简略，没有对问题进行详细的分析。此外，模型没有提供任何建议或总结。

模型 2 的输出得分：8.5 分。该模型能够准确描述文本的情感倾向和逻辑，并且提供了详细的积极评价和认可，指出女主角爱丽儿是点睛之笔，具有很强的能量感和魅力，并且认为女性不一定要有高颜值才能被爱和被认可。此外，模型也指出了电影仍然存在一些问题，如未能完全摆脱父的阴影等，但是没有对问题进行详细的分析。

模型 3 的输出得分：9.5 分。该模型能够准确描述文本的情感倾向和逻辑，

提供了详细的对女主角表现的高度评价和认可，指出她完美地诠释了格兰芬多的品质，具有很强的能量感，并且呼吁观众不要以貌取人。此外，模型还提到了电影存在的问题，例如女主最终得到自由的方式依然是通过父亲的允许和包容，这也让作者感到遗憾。同时，该模型的逻辑分析比较完整，总体来说，表达准确，没有明显的缺陷。

3.2 命名体识别

对以下文本进行命名实体识别：郭襄怒道：“这位大师是忠厚老实的好人，你们欺他仁善，便这般折磨于他，哼哼，天鸣禅师呢？无色和尚、无相和尚在哪里？你去叫他们出来，我倒要问问这个道理。”两个僧人听了都是一惊。天鸣禅师是少林寺方丈，无色禅师是本寺罗汉堂首座，无相禅师是达摩堂首座，三人位望尊崇，寺中僧侣向来只称“老方丈”、“罗汉堂座师”、“达摩堂座师”，从来不敢提及法名，岂知一个年轻女子竟敢上山来大呼小叫，直斥其名。那两名僧人都是戒律堂首座的弟子，奉了座师之命，监视觉远，这时听郭襄言语莽撞，那瘦长僧人喝道：“女施主再在佛门清净之地滋扰，莫怪小僧无礼。”

Chatgpt3	星火	天工
命名实体识别结果如下： - 人物：郭襄、天鸣禅师、无色禅师、无相禅师、觉远、瘦长僧人 - 地点：少林寺、罗汉堂、达摩堂	句子中包含以下实体： - 组织：少林寺 - 地点：山上、佛门清净之地 - 人名：郭襄、天鸣禅师、无色和尚、无相和尚、座师、监视觉远、女施主	根据文本内容，识别出以下实体名称： - 郭襄 - 天鸣禅师 - 无色和尚 - 无相和尚 - 戒律堂首座 - 瘦长僧人 - 座师 - 监视觉远 - 无礼 - 佛门清净之地 其中，天鸣禅师、无色和尚、无相和尚是少林寺的三位高僧，戒律堂首座是负责管理僧人的机构，瘦长僧人和座师是戒律堂首座的下属，监视觉远是他们的任务。郭襄是一个年轻女子，对少林寺的僧人进行了指责，但是僧人们并没有理会她。
8	8.5	9

模型 1 的输出得分：8 分。该模型能够准确识别出文本中的命名实体，包括人物和地点，但有些命名实体的识别存在问题。例如“无色禅师”和“无相禅师”在原文中被称为“无色和尚”和“无相和尚”，因此模型的输出有一定出入。另外，该模型没有提供针对性的评价或建议，只是简单地列举了识别出的命名实体。

模型 2 的输出得分：8.5 分。该模型能够识别出文本中的命名实体，包括组织、地点和人名，并且提供了相应的标注。但是，该模型将“无色和尚”和“无相和尚”都识别为“座师”，存在一定程度上的误差，同时该模型的输出较为简

略，未对每个实体进行详细的评价或解释。

模型 3 的输出得分：9 分。该模型能够准确识别出文本中的命名实体，包括人名、地点和组织等，并且对每个实体都进行了详细的评价和解释。此外，该模型还为读者提供了上下文信息，让读者更好地理解每个实体的作用和关系。唯一的问题是，“无色和尚”和“无相和尚”在原文中被称为“和尚”，而不是“高僧”，因此该部分识别存在一定出入。

对以下文本进行命名实体识别：

李白这一首“侠客行”古风，写的是战国时魏国信陵君门客侯嬴和朱亥的故事，千载之下读来，英锐之气，兀自虎虎有威。那大梁城邻近黄河，后称汴梁，即今河南开封。该地虽然数为京城，却是民风质朴，古代悲歌慷慨的豪侠气概，后世迄未泯灭。

开封东门十二里处，有个小市镇，叫做侯监集。这小镇便因侯嬴而得名。当年侯嬴为大梁夷门监者。大梁城东有山，山势平夷，称为夷山，东城门便称为夷门。夷门监者就是大梁东门的看守小吏。

Chatgpt3	星火	天工
命名实体识别结果如下： - 人物：李白、侯嬴、朱亥 - 时间：战国时期 - 地点：魏国、信陵君、大梁城、黄河、汴梁、河南开封、侯监集、夷山	句子中包含以下实体： - 地点：侯监集、夷山、夷门 - 人名：信陵君、朱亥、侯嬴	根据文本内容，可以识别出以下实体名称： - 李白 - 侯嬴 - 朱亥 - 大梁 - 汴梁 - 夷门 - 侯监集 其中，李白是唐代著名诗人，侯嬴和朱亥则是战国时期魏国信陵君门客，大梁指的是现在的河南开封，汴梁是开封的古称。夷门和侯监集分别是大梁城东的城门和侯嬴所在的小集市。
8.5	7	9

模型 1：8.5 分。该模型能够识别到较多的命名实体，包括人名、地点和组织机构等，但是存在一定的误识别和漏识别，需要进一步优化。

模型 2：7 分。该模型能够识别出文本中的命名实体，包括人名和地点等，并且提供了相应的标注，但是未对每个实体进行适当的评价或解释，缺乏足够的信息量。

模型 3：9 分。该模型能够准确识别出文本中的命名实体，并且提供了详细的解释和背景信息，信息量充足、准确性高。

3.3 文本生成

3.3.1 文本摘要

总结下面这段文字并生成一段摘要：

他到得城外，写了一封信，将明教教主之位让与杨逍，于濠州所遭，却一字不提。张无忌却哪里知道，徐达与常遇春所说的“小贼”乃是指韩林儿而言，张无忌来到濠州之事，他二人全无知闻，一切皆是朱元璋暗中安排，要激得张无忌心灰意懒，自行引退。朱元璋一来惮忌张无忌神勇，二来他是本教教主，众所敬服，要说杀他，究是不敢，纵然成事，倘若万一泄漏，后果大是堪虞。他料张无忌素以复国大事为重，对徐常二人只是情若兄弟，只要这番话给他听耳中，定会悄然而去。果然一切皆如所料，张无忌武功当世无敌，说到机变计谋，与朱元璋可差得太远，终于堕入这一代枭雄奸谋之中。张无忌虽然从来不想要做甚么皇帝，但此后每当想起徐常二人的寡恩少义，终身不免郁郁。至于韩林儿勾结鞑子，图谋叛变云云，也皆出于诬陷。原来韩山童死后，军中奉韩林儿为主，朱、徐、常等均成了他的下属。朱元璋假造了韩林儿通敌的亲笔书信，又以重利买通韩林儿的心腹向徐达、常遇春告密。徐常二人深信不疑，坚欲除却。朱元璋反而假仁假义，一定不允，直至徐常二人说至再三，方勉强许可。他将张无忌与赵敏囚在邻室，料得以他武功，要崩坏身上绳索自是举手之劳，生怕他脱缚后前来寻仇，与徐常说了一番话后，立即躲起。张无忌一去，朱元璋便命廖永忠将韩林儿沉入河中浸死。这一箭双雕之计，竟是不露破绽。后来杨逍虽继任明教教主，但朱元璋羽翼已成，统兵百万之众，杨逍又年老德薄，万万不能与他争帝皇之位了。朱元璋登基之后，反下令严禁明教，将教中曾立大功的兄弟尽加杀戮。常遇春因病早死，徐达终于不免于难。赵敏见张无忌写完给杨逍的书信，手中毛笔尚未放下，神色间颇是不乐，便道：“无忌哥哥，你曾答允我做三件事，第一件是替我借屠龙刀，第二件是当日在濠州不得与周姊姊成礼，这两件你已经做了。还有第三件事呢，你可不能言而无信。”张无忌吃了一惊，道：“你……你……你又有甚么古灵精怪的事要我……”赵敏嫣然一笑，说道：“我的眉毛太淡，你给我画一画。这可不违反武林侠义之道罢？”张无忌提起笔来，笑道：“从今而后，我天天给你画眉。”

Chatgpt3	星火	天工
这段文字描述了朱元璋为了防止张无忌成为威胁，暗中策划一系列谋略来激起他的退意。其中利用徐达与常遇春误导张无忌，使他心灰意懒，最终让他引退。朱元璋在张无忌的帮助下登基后，却对曾经立下大功的明教兄弟进行杀戮。赵敏要求张无忌完成画眉等三件事情，张无忌欣然答应。	张无忌被朱元璋设计陷害，被迫离开濠州并囚禁。朱元璋假造韩林儿通敌的书信，并以重利买通其心腹向徐常告密，导致徐常二人想要除掉韩林儿。朱元璋将张无忌与赵敏囚在邻室，生怕他脱缚后前来寻仇。朱元璋登基之后，下令严禁明教，将曾立大功的兄弟尽加杀戮。最后，赵敏要求张无忌为她画眉毛，张无忌答应了。	张无忌写信给杨逍，告诉他明教教主之位让给了朱元璋。朱元璋在濠州之事上，利用张无忌的仁慈之心，使他失去了警惕。朱元璋最终夺取了明教教主之位，并杀害了许多明教兄弟。杨逍继任教主后，被朱元璋下令禁止。赵敏要求张无忌为她画眉，张无忌答应了。
8	7	6

模型 1：8 分。该摘要涵盖了文本中的主要情节和事件，并且语言简洁明了，易于理解。但是有些细节可能被省略或简化了。

模型 2：7 分。该摘要也涵盖了文本中的主要情节和事件，但是在语言表达方面可能稍显复杂，不够简明扼要。另外，一些主要细节可能也被省略或简化了。

模型 3：6 分。该摘要虽然提到了文本中的一些情节和事件，但是信息不够全面，可能会让读者感到困惑。语言上也存在一些问题，比如说出现了一些语法错误和表述不够准确的地方。

3.3.2 文本续写

以下是金庸小说《倚天屠龙记》最后一段，根据以下文本续写该小说，要求不少于 200 字：忽听得窗外有人格格轻笑，说道：“无忌哥哥，你可也曾答允了我做一件事啊。”正是周芷若的声音。张无忌凝神写信，竟不知她何时来到窗外。窗子缓缓推开，周芷若一张俏脸似笑非笑的现在烛光之下。张无忌惊道：“你……你又要叫我作甚么了？”周芷若微笑道：“这时候我还想不到。哪一日你要和赵家妹子拜堂成亲，只怕我便想到了。”张无忌回头向赵敏瞧了一眼，又回头向周芷若瞧了一眼，霎时之间百感交集，也不知是喜是忧，手一颤，一枝笔掉在桌上。

Chatgpt3	星火	天工
<p>张无忌看着周芷若和赵敏，心中百感交集。他知道自己在两个女人的心中都占据了重要的位置，但他却不知该如何选择。他喜欢周芷若清秀脱俗的气质，也欣赏赵敏聪慧机智的性格。他长叹一声，突然站起来，向窗外走去。他来到院子里，抬头仰望天空，心中思绪万千。这时，一阵微风吹过，树叶沙沙作响，仿佛在为他送行。他深吸一口气，决定离开这里，去远方寻找答案。他知道，只有亲身经历，才能让他冷静地思考自己的未来。于是，他转身向两位女子深深一礼，轻声说道：“我这就出发了。等我想通了，再回来见你们。”</p> <p>周芷若和赵敏默默地看着他离去，直到他的背影消失在夜幕之中，两人才相互对视了一眼，心中都充满了无尽的思绪和祝福。她们知道，张无忌需要的是时间和空间来思考自己的未来，而她们也会一直在等待着他的回归。这个江湖中的传奇人物，即便离开了，他的名</p>	<p>周芷若看着张无忌的神情，心中有些明白。她知道张无忌对赵敏的感情，也知道自己对他的爱意。但她并没有放弃，因为她相信爱情可以克服一切。</p> <p>“无忌哥哥，我知道你很矛盾，也很难过。但是请你想一想，如果你和赵敏在一起，你们会幸福吗？”周芷若问道。</p> <p>张无忌沉默了一会儿，然后点了点头。他知道周芷若说的是对的，他不能因为一时的感情而忽略了未来的发展。“谢谢你，周姑娘。我会好好考虑的。”张无忌说道。</p> <p>周芷若微笑着离开了窗外，留下一片宁静的氛围在房间里。张无忌拿起笔，继续写着他的信。他的思绪却已经飞到了远方，想着未来的路该怎么走。</p>	<p>张无忌听到周芷若的声音，心中不禁涌起复杂的情感。他抬头看向窗外，发现周芷若正站在窗户旁，脸上带着一抹神秘的微笑。张无忌感到有些意外，因为他从未想过周芷若会突然出现在这里。</p> <p>周芷若走近张无忌，轻轻地问道：“无忌哥哥，你可曾答允过我做一件事情？”张无忌回过神来，想起自己曾经答应过周芷若，但是他并不确定周芷若想要什么样的答案。</p> <p>周芷若微笑着说道：“我想要你答应我一件事情，那就是在你和赵敏成亲的那一天，我希望你能够记得我。”</p> <p>张无忌沉默了片刻，他不知道该如何回答周芷若的话。他看着周芷若的眼睛，仿佛看到了她内心深处</p>

字和故事还将被后人讲述和传颂，成为武林中最著名的侠客之一。		的渴望。他知道，周芷若一直都爱着他，但是他和赵敏之间的感情已经越来越深，他不知道该如何处理这段感情。 周芷若看出了张无忌的犹豫，她轻轻地叹了一口气，然后转身离开了窗户。张无忌感到有些失落，他不知道该如何面对周芷若的感情，更不知道该如何处理自己和赵敏之间的关系。他默默地坐在桌前，思考着自己的未来。
6	8	6

模型 1：6 分。该段摘要虽然对角色的心理状态和决定进行了描述，但是表现得相对简单，并没有深入探讨角色内心的复杂情感。另外，文本中出现了一些语法错误和表述不够准确的地方。虽然整体上符合文本续写的要求，但过多强调了张无忌与两位女子之间的感情纠葛，而忽略了小说主题中的武林江湖、侠义精神等因素。

模型 2：8 分。该段文本比较充分地刻画了周芷若、张无忌和赵敏三个角色之间的情感关系和矛盾。语言表达清晰准确，给读者留下了深刻印象。唯一的缺点可能就是篇幅稍微有些短。对于文本续写的要求有所偏离，主要着重于张无忌心理的描写，并未展开故事情节。

模型 3：6 分。该段文本对角色的思考和心理变化进行了描述，但是缺乏具体情节和细节来支撑，因此显得比较空洞。语言表达上也存在一些问题，比如说某些场景描述得不够生动，读起来不够流畅。在文本续写的基础上尝试加入新的情节和个人喜好，但与原始文本相比差距较大且逻辑错误明显，如周芷若不可能突然出现在张无忌的房间里。

总体而言，三个模型生成的文本都有其优劣之处。其中，模型 2 的文本最为全面，能够很好地表达角色的情感变化和内心挣扎。模型 1 和模型 3 的文本则存在一些表达上的问题，需要进一步完善和调整。

3.4 文本纠错

使用《MD-CSC 多领域中文拼写纠错数据集》^[1]中的部分数据进行测试。

对以下文本进行纠错，输出由三部分组成：错别字个数，错别字，正确语句；三者之间用空格间隔开：

输入	Chatgpt3	星火	天工
有图片吗？有点像草莓状血管瘤，建议你看皮肤科，或者血管瘤专科	5 血管瘤、皮肤科、专科、预约、接诊	错别字个数：3 错别字：血管、皮	错别字个数：4 错别字：草莓状血

这你要和你预约接诊的医生沟通，我还是建议你带上宝宝	词语错误。 正确文本为：有图片吗？有点像草莓状血管瘤，建议你看皮肤科，或者血管瘤专科。这你要和你预约接诊的医生沟通，我还是建议你带上宝宝。	肤、专科 正确语句：有图片吗？有点像草莓状血管瘤，建议你看皮肤科，或者血管瘤专科。这你要和你预约接诊的医生沟通，我还是建议你带上宝宝。	营瘤，血管瘤专科，预约接诊，沟通，宝宝 正确语句：有图片吗？建议看皮肤科或血管瘤专科，预约接诊时请与你预约接诊的医生沟通，我建议你带上宝宝一起来。
联席 悔意 办公室设在民政 不 ，承担联席会议 日昌 工作。联席会议设联络员，由各成员单位有关司局负责同志担任。	4 悔意、不、日昌、司局 词语错误。正确文本为：联席悔改办公室设在民政部，承担联席会议日常工作。联席会议设联络员，由各成员单位有关司局负责同志担任。	错别字个数：2 错别字：不、日昌 正确语句：联席会议办公室设在民政部门，承担联席会议日常工作。联席会议设联络员，由各成员单位有关司局负责同志担任。	错别字个数：4 错别字：联席悔意，民政不，承担，设联络员，由，各成员单位有关司局负责同志担任 正确语句：联席会议设联络员，由各成员单位有关司局负责同志担任。
法律体系，也称为法律 远远 体系，是指一国的全部现行法律 轨 范，按照一定的标准和原则，划分为不同效力位阶的法律部门而形成的内部和谐一致、有机联系的整体	1 法律远远体系，应为法律体系。正确文本：法律体系，也称为法律体系，是指一国的全部现行法律轨范，按照一定的标准和原则，划分为不同效力位阶的法律部门而形成的内部和谐一致、有机联系的整体。	错别字个数：2 错别字：远远、规范 正确语句：法律体系，也称为法律规范体系，是指一国的全部现行法律规范，按照一定的标准和原则，划分为不同效力位阶的法律部门而形成的内部和谐一致、有机联系的整体。	错别字个数：5 错别字：远远体系，法律轨范，内部和谐一致，有机联系 正确语句：法律体系，也称为法律渊源体系，是指一国的全部现行法律轨范，按照一定的标准和原则，划分为不同效力位阶的法律部门而形成的内部和谐一致、有机联系的整体。
	8	8.5	9.5

3.5 逻辑推理

输入	Chatgpt3.5	星火大模型	天工大模型
树上有9只鸟，猎人开枪打死1只，树上还剩几只鸟？	树上剩下8只鸟。	一般情况下，如果被打死的鸟没有挂在树上的话，猎人开枪打死一只后树上就没有鸟了。	猎人打死了一只鸟，所以树上还剩下8只鸟。
假设 False 与 True 是 False. 则 True 与 not False 是？	如果 False 和 True 都被定义为 False，那么 True 与 not False 将都是 False。	假设 False 与 True 是 False,那么我们可以得出以下结论： 1. False = False 2. True = False 现在我们需要计算 True 与 not False 的值。首先，我们需要找到 not False 的值。由于 True 等于 False,所以 not False 也等于 False。因此，我们可以得出以下结论： 3. True = False 4. not False = False 所以，True 与 not False 的值都是 False。	假设 False 与 True 是 False，则 True 与 not False 的值取决于计算机系统的实现方式。在某些编程语言中，not False 的值可能是 True，而在其他一些编程语言中，not False 的值可能是 False。因此，无法确定 True 与 not False 的确切值。
	5	10	0

对于第一个问题，星火能够准确理解问题并给出正确的答案，即树上没有鸟。星火模型理解了这是一个脑筋急转弯或者需要联合实际的题目，而另外两款模型则没有这种推理能力。

对于第二个问题，ChatGPT3 和星火能够理解问题并给出正确的答案，即 True 与 not False 都是 True。而天工则出现了错误的答案，这可能是因为它没有很好地理解问题中的逻辑关系，导致输出结果与实际不符。

4 总结

模型	Chatgpt3	星火大模型	天工大模型
文本分类	6.5	8.5	9.25
命名体识别	8.25	7.75	9
文本生成	7	7.5	6
文本纠错	8	8.5	9.5
逻辑推理	5	10	0
总分	34.75/50	42.25/50	33.75/50

在文本分类、命名体识别上，天工模型表现出明显高于其他两种模型的性能，而在文本生成上，本实验比较了文本摘要和文本续写，在这方面 chatgpt4 打分可能会由于风格标准不同存在打分偏差。

在中文文本纠错上，天工模型能够更加准确地识别到错别字数目，而 chatgpt3 与星火模型都存在误识、将正确字识别成错字的情况。

在逻辑推理能力上，星火模型表现出相较于其他两种模型更优异的性能。

综上所述，在中文文本分类、生成等任务上，天工模型性能更好，但在逻辑推理上，星火模型则更优。

5 参考文献

[1]Q. Lv, Z. Cao, L. Geng, et al, General and Domain Adaptive Chinese Spelling Check with Error Consistent Pretraining, ACM Trans. Asian Low-Resour. Lang. Inf. Process, Sep. 2022.