

如何写出优雅的Prompt

知乎上@ybyq的一篇回答-大模型微调到底有没有技术含量，提到算法工程师在做数据工作的几个level：

1：继承实验室或者同事的训练数据，拿到之后也不 check 一下数据质量，直接放进去训。

2：下载一个开源数据，构建“system + query + answer”集合。

3：利用LLM生成数据，学会用LLM喜好的 prompt 去请求。并且意识到数据 prompt 多样性，想尽各种办法去扩充 prompt 的任务多样性和表达方式多样性，甚至去刻意加一些 noisy prompt 去提升抗噪性。同时，愿意放下身架，一条一条去 check 数据质量，去和标注同学对齐标注标准。

4：利用用户的交互日志来驱动数据构造过程，收集用户的真实 prompt，用规则或者LLM去分析用户的 feedback，进而获得高质量的 answer 数据。

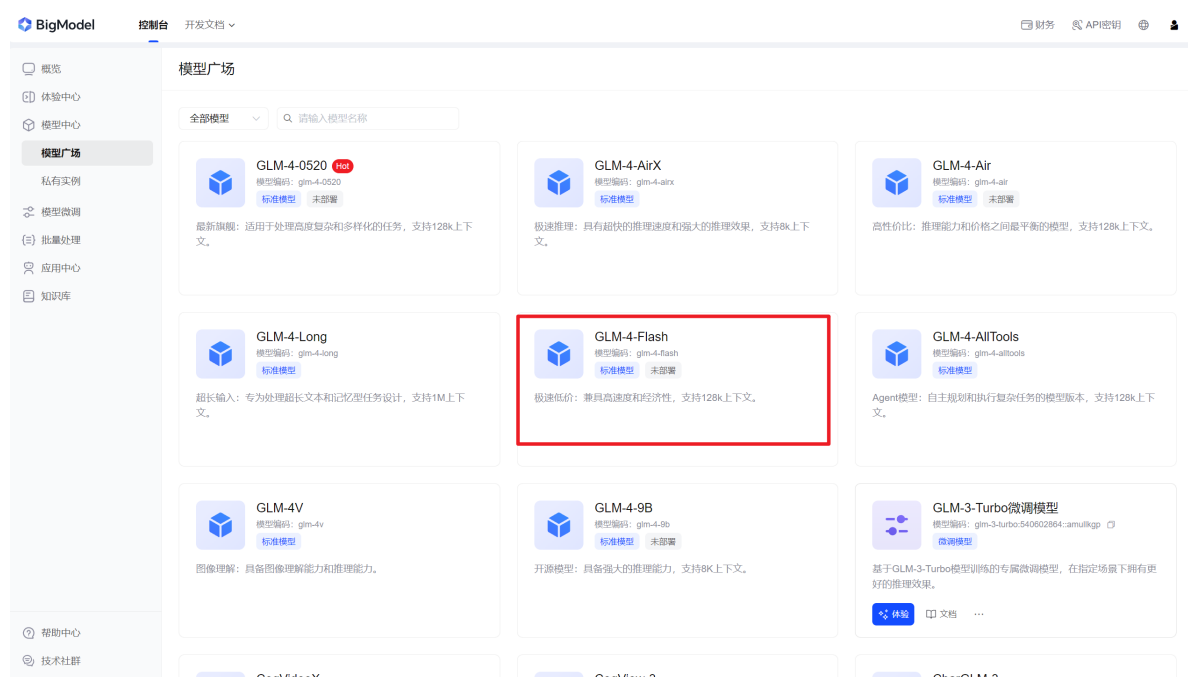
5：借鉴 cot、rag、function_call、agent 等思路，把复杂的模型无法胜任的任务在数据层面就进行拆解，比如“模型写不出长篇小说”-->“模型写小说大纲，模型基于小说大纲写长篇小说”。

.....

确实，大模型的能力很多时候都遵从了garbage in garbage out的原则，因此做数据这种看似没有技术含量的工作其实对最终的结果至关重要。

那么怎么样才能写出优雅的prompt，让LLM可能存在的幻觉问题、bad case等在数据层面就解决掉，也是算法工程师应该探索的方向。

下面以GLM-4-flash为例介绍prompt写作方法：



1 prompt的定义

A prompt is an input to a Generative AI model, that is used to guide its output.

来源：The Prompt Report: A Systematic Survey of Prompting Techniques

在这篇论文中还给出了prompt的常见组成部分：

- 指令，即核心目标，如“请判断这句话是积极还是消极”。
- 示例，给出一些案例供参考和比对，如“请参考下面几个案例”。

- 输出格式，对最终输出的结果要求，例如分点、json格式之类的。
- 角色。已经有大量实验证明，给大模型赋予角色，能很大程度提升大模型的输出结果。
- 附加信息。其他需要配合任务进行输出的结果，如要做翻译，那就得提供原文，或者是RAG需要提供reference等。

以最近免费的GLM-4-flash为例：

👎 反例:

The screenshot displays the GPT-4o interface, divided into two main sections: a chat area on the left and a settings panel on the right.

Chat Area (Left):

- Header:** A blue circular icon with a white face, followed by the text "黑悟空为什么这么火" (Why is Black Myth: Wukong so popular?).
- Message 1 (User):** A red star icon followed by the text: "《黑神话：悟空》之所以能够火爆全球，主要原因可以从以下几个方面进行分析：" (The reason why "Black Myth: Wukong" can become a global hit lies mainly in the following aspects for analysis:).
- Message 2 (AI):** A detailed analysis of the game's success, structured into seven numbered points:
 - 技术创新与高品质制作:** The game adopted cutting-edge engine technology, achieving detailed visual representation and complex lighting effects and vast scene details, bringing unprecedented visual feast. High-precision action capture technology made character movements more natural and fluid, greatly enhancing player immersion and代入感 (immersion).
 - 深厚的文化底蕴:** "Black Myth: Wukong" takes the classic Chinese novel "Journey to the West" as its theme, deeply挖掘 (explored) and presented Chinese traditional culture. The game's scenes, character design, and plot are all filled with thick Chinese style, providing a unique cultural experience for global players.
 - 国产游戏里程碑:** As the first domestic AAA game masterpiece, "Black Myth: Wukong" represents the highest level of domestic game industry, its success marks a milestone for the technological innovation and quality improvement of China's game industry.
 - 玩家口碑与市场认可:** After launching, it quickly won widespread praise from players, not only receiving market acclaim domestically, but also performing excellently on international platforms like Steam, Epic Game Store, WeGame, and PlayStation.
 - 文化自信的体现:** The game's success reflects the improvement of our cultural confidence, it is not just a game, but more importantly, an important symbol of the Chinese game industry's march towards the world.
 - 带动相关产业增长:** The popularity of "Black Myth: Wukong" also drove the sales of peripheral and licensed products, promoting the development of related industries like game hardware and cultural tourism.
 - 社会认可与关注:** The development of the game industry has received widespread attention and recognition from all sectors of society, from mainstream media to government spokespersons, all highly evaluating the game's success.

综上所述，《黑神话：悟空》之所以能够火爆全球，是由于其在技术创新、文化传承、市场认可和社会影响等多方面的综合因素共同作用的结果。

以上内容为 AI 生成，不代表开发者立场，请勿删除或修改本标记
- Input Field:** A text box containing the prompt: "向我发指令，我可以完成对话聊天、创作文章、生成代码等多种任务" (Send me instructions, I can complete various tasks such as chat conversation, article creation, and code generation).

Settings Panel (Right):

- Header:** "体验对象" (Experience Object) with a "查看代码" (View Code) link.
- Model:** A dropdown menu showing "模型" (Model).
- Model 1:** A dropdown menu showing "glm-4-flash".
- 多轮对话 ①:** A toggle switch that is currently turned on.
- 参数设置 (system_prompt ①):**
 - Text:** "你是一个乐于解答各种问题的助手，你的任务是为用户提供专业、准确、有见地的建议。" (You are an assistant who is happy to answer various questions. Your task is to provide users with professional, accurate, and insightful suggestions.)
 - max_tokens ①:** A slider set to 1024.
 - temperature ①:** A slider set to 0.95.
 - top_p ①:** A slider set to 0.70.
- 工具调用:** A toggle switch that is currently turned on.
- 网页检索 ①:** A toggle switch that is currently turned on.
- 知识库检索 ①:** A toggle switch that is currently turned on.

📖 正例：



你是一名游戏运营专家，负责分析新发布游戏《黑神话：悟空》的成功因素。你的任务是深入分析一款游戏的市场表现和玩家反馈，以确定其成功的关键因素。你需要：

任务：

详细描述游戏的核心玩法、主要特点、艺术风格和开发背景。

使用以下步骤/技术来识别游戏成功的因素：

玩家评价分析

游戏机制的创新程度

市场营销策略评估

竞争对手对比

用户参与度统计

游戏的社区和社交媒体反响

开发团队的影响力

游戏更新和支持策略

财务表现分析

用户留存率和活跃度

媒体和评论家的评价

游戏的可玩性和挑战性

结合游戏的各个方面，逐步分析每一个成功因素。

运用你的专业知识和市场洞察，给出游戏成功的初步评估。

批判并重新评估每个因素的影响。

对游戏的成功因素得出最终结论，并给出其成功的主要原因，范围为0到100。



游戏核心玩法、主要特点、艺术风格和开发背景

核心玩法：

《黑神话：悟空》是一款动作角色扮演游戏，以中国古典名著《西游记》为背景，玩家将扮演孙悟空，进行闯关、战斗和探索。游戏强调打击感和流畅的动作设计，玩家可以通过各种技能和武器组合进行战斗。

向我发指令，我可以完成对话聊天、创作文章、生成代码等多种任务



主要特点

- **动作战斗：**游戏提供了流畅的动作战斗体验，强调技巧和反应速度。
- **角色成长：**玩家可以通过完成任务、击败敌人和收集资源来提升角色的技能和装备。
- **探索和谜题解决：**游戏中有大量的探索区域和谜题，需要玩家运用智慧和策略来解决。

艺术风格

- **东方幻想：**游戏的艺术风格深受中国传统文化的影响，包括神话、历史和艺术元素。
- **视觉和听觉体验：**游戏利用先进的图形和音效技术，提供了沉浸式的视觉和听觉体验。

开发背景

- **开发周期：**《黑神话：悟空》的开发周期长达多年，开发团队专注于打造高品质的游戏体验。
- **团队背景：**游戏由经验丰富的开发团队制作，他们在动作游戏领域有深厚的积累。

成功因素分析

1. 玩家评价分析

- 玩家普遍对游戏的战斗系统、故事和艺术风格给予了高度评价。
- 玩家反馈显示，游戏的高难度和深度是玩家喜爱的特点。

2. 游戏机制的创新程度

- 游戏在动作战斗和谜题解决机制上有所创新，但整体上还是基于传统动作游戏模式。
- 创新程度中等，但结合了独特的中国元素。

体验对象 [查看代码](#)

模型

模型 [①](#)

glm-4-flash

多轮对话 [①](#)

参数设置

system_prompt [①](#)

你是一个乐于解答各种问题的助手，你的任务是为用户提供专业、准确、有见地的建议。

max_tokens [①](#)

1024

temperature [①](#)

0.95

top_p [①](#)

0.70

工具调用

网页检索 [①](#)

知识库检索 [①](#)

模型

模型 [①](#)

glm-4-flash

多轮对话 [①](#)

参数设置

system_prompt [①](#)

你是一个乐于解答各种问题的助手，你的任务是为用户提供专业、准确、有见地的建议。

max_tokens [①](#)

1024

temperature [①](#)

0.95

top_p [①](#)

0.70

<div><div>3. 市场营销策略</div><div><div><div>• 游戏通过高质量的预告片和积极的社交媒体活动，成功吸引了玩家的注意。</div><div>• 市场营销策略得当，但可能过于依赖预告片。</div></div></div><div><div>4. 评估竞争对手</div><div><div><div>• 与其他动作游戏相比，《黑神话:悟空》在画面和音效上具有优势。</div><div>• 在游戏玩法上，可能缺乏独特的卖点。</div></div></div><div><div>5. 用户参与度</div><div><div><div>• 游戏的社区活跃，玩家在社交媒体上积极讨论游戏。</div><div>• 用户参与度较高，社区对游戏的未来更新充满期待。</div></div></div><div><div>6. 社交媒体反响</div><div><div><div>• 社交媒体上对《黑神话:悟空》的讨论热烈，正面评价居多。</div><div>• 社交媒体的反响对游戏的成功起到了重要作用。</div></div></div><div><div>7. 开发团队的影响力</div><div><div><div>• 开发团队的知名度和声誉提高了游戏的市场期待值。</div><div>• 团队的影响力是游戏成功的关键因素之一。</div></div></div><div><div>8. 游戏更新和支持策略</div><div><div><div>• 开发团队承诺会持续提供游戏更新和补丁，以保持游戏的新鲜感和玩家兴趣。</div><div>• 更新和支持策略是维持游戏长期成功的关键。</div></div></div></div></div></div></div></div></div>	<div><div>9. 财务表现</div><div><div><div>• 游戏在财务上取得了显著成功，尤其是在预售阶段。</div><div>• 财务表现对游戏的长期运营至关重要。</div></div></div><div><div>10. 用户留存率和活跃度</div><div><div><div>• 游戏发布后，用户留存率和活跃度保持在高水平。</div><div>• 留存率和活跃度是衡量游戏成功的重要指标。</div></div></div><div><div>成功因素评估</div><div><div><div>• 可玩性和挑战性：游戏的可玩性高，挑战性适中，适合不同水平的玩家。</div><div>• 综合评估：游戏的成功是多方面因素共同作用的结果。</div></div></div><div><div>最终结论</div><div><div><div>《黑神话:悟空》的成功主要归因于以下几个因素：</div><div><div>1. 出色的艺术和音乐表现：提供了沉浸式的游戏体验。</div><div>2. 深度的故事和角色设计：吸引了玩家的兴趣和情感投入。</div><div>3. 有效的市场营销和品牌建设：提高了游戏的知名度和期待值。</div><div>4. 持续的更新和支持：保持了游戏的新鲜感和玩家的参与度。</div></div></div><div><div>综合上述因素，我给《黑神话:悟空》的成功因素评分为92/100。游戏在艺术、故事和市场营销方面表现出色，但游戏机制的创新程度和独特性还有提升空间。游戏的成功是多个因素综合作用的结果，特别是艺术和故事的成功结合了开发团队的知名度和市场营销策略的有效性。</div><div><div>以上内容为 AI 生成，不代表开发者立场，请勿删除或修改本标记</div></div></div></div></div></div></div></div>
--	--

体验对象

查看代码

模型

▼

模型 ⓘ

glm-4-flash

▼

多轮对话 ⓘ

参数设置

system_prompt ⓘ

你是一个乐于解答各种问题的助手，你的任务是为用户提供专业、准确、有见地的建议。

max_tokens ⓘ

1024

temperature ⓘ

0.95

top_p ⓘ

0.70

体验对象

查看代码

模型

▼

模型 ⓘ

glm-4-flash

▼

多轮对话 ⓘ

参数设置

system_prompt ⓘ

你是一个乐于解答各种问题的助手，你的任务是为用户提供专业、准确、有见地的建议。

max_tokens ⓘ

1024

temperature ⓘ

0.95

top_p ⓘ

0.70

工具调用

网页检索 ⓘ

2 prompt的分类

- 上下文学习（in-context learning）。通俗而言，举例子，换个角度，few shot之类的也属于这个范畴。
- 零样本学习。类似CoT、Role-Play（角色）、风格、情感的提示，都属于这个范畴。
- 思想生成。常见的就是CoT及其变体。
- 分解。把问题进行拆解然后逐步完成的思想，比较常用，核心难点在于如何拆解，论文里类似“Least-to-Most Prompting”之类的都有提及。
- 集成。这个严格来说不算分解的反义，这里指的是通过多次或者多种方式对统一方案进行验证来实现最终效果的可靠性，比较典型常用的就是self-consistency，通过重复让大模型生成自己的答案来加强对最终结果的验证，配合CoT和非0的temperature即可实现多次的结果生成，而Prompt Paraphrasing则是通过改写prompt来确认最终效果，也是类似的思想。
- 自我批评。集成强调的是生成结果的多次，而自我批评则是强调对生成结果的验证，让大模型自己判断内容是否正确。Self-Calibration就是非常典型的，在原有生成的基础上，把问题+回复重新输入大模型让大模型来判断是否正确。

3 Prompt方案

选了几个常用且有代码性的，在日常工作中值得一试

3.1 解析分析

解析分解（Decomposition）是一种在链式思维（Chain-of-Thought, CoT）基础上进化而来的方法，通过将复杂任务拆分为多个步骤逐步完成。这种方法有助于更好地控制整个过程，并提高最终结果的准确性。核心思想是将任务细化为更易处理的子任务，逐步解决，最终得出全局解决方案。在实际应用中，任务的拆解方式可能因场景不同而有所变化，以下是几种典型的应用方式，供参考：

- **Least-to-Most Prompting**：按顺序逐步解决各个子问题，最后推导出整体结果。
- **Decomposed Prompting**：结合few-shot学习，由大模型决策调用特定函数来解决问题。此方法在agent领域广泛应用，与搜索对话中的意图识别及相应处理有相似之处。
- **Plan-and-Solve Prompting**：让模型首先设计执行计划，然后按照计划逐步完成任务。

在实际操作中，任务分解后请求大模型的次数会显著增加，这将对时间和资源消耗带来压力，因此需要特别注意。除此之外，还有两个关键点需要考虑：

- 每个步骤的效果应进行监控，以防某一短板显著影响最终结果。
- 可以仔细评估每个拆解后的步骤，判断是否需要大模型参与。例如在数据量较大的场景中，决策任务可能可以由小模型或规则系统来完成，从而有效提高效率，降低成本。

例如，可以对拆解后的query使用GLM-4-flash进行回答，理由如下：

- **免费**，GLM-4-flash 模型免费，让使用解析分析法构造prompt的推理成本降下来。
- **速度快**，生成速度72.14 token/s，约等于115字符/s。
- 具有与 GPT-4 相当的**函数功能调用能力**。
- **网页检索**，能实时访问天气、新闻等信息。
- **推理能力**：基于强大的预训练基座，GLM-4-flash在中英文性能上表现出色。

3.2 CoT

CoT通过引导大模型逐步思考，能够提升其处理复杂问题的能力，并且对简单问题的解答也会更加稳定。我的理解是，其核心原理在于，通过展示思维过程，可以使模型在最终解码时，能够参考到更加全局和稳定的信息，从而提高答案的准确性。

最直接的实现方式是在指令中添加一句话，比如“让我们一步一步来思考这个问题。”当然，也可以结合具体问题的内容，提出类似“从XX角度逐步分析”或“请按照XX步骤思考”等指令来加强效果。在逻辑推理能力较强的模型中，这种方法通常能带来显著的提升。

这种方法的优点在于实现成本低，但缺点是生成的答案可能较长，导致解析的难度增加、耗时也会变长。

3.3 集成

集成的思想在原来深度学习、机器学习的早期就已经有考虑到，其核心思想就是构造多个类似的结果然后合并，类似Self-Consistency通过CoT产生多个结果然后综合评估， Demonstration Ensembling通过多种few-shot结果来判断， Mixture of Reasoning Experts是用MoE多专家系统提供不同的推理思路。

3.4 Self-Criticism

自我批评旨在利用大模型回溯，验证自己生成的结果是否正确，最简单的方式就是直接把问题+大模型答案通过prompt拼接让模型进行结果验证（Self-Calibration），更进一步则有Self-Refine进一步提供修改建议或者完成修改。

prompt并非局限在自己的编辑，研究上还会有很多细分的场景和思路。

小结

prompt并不是看上去那样简单的工作，可以深入研究的实验的点也很多，推荐的综述性论文如下：

- 1、The Prompt Report: A Systematic Survey of Prompting Techniques
- 2、A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications
- 3、A Survey of Prompt Engineering Methods in Large Language Models for Different NLP Tasks

参考：机智的叉烧