

8 错误分析和故障排除

8.1 识别常见错误

人工智能语言模型，如 GPT-4 和 ChatGPT，并不完美，可能会出错。识别输出中的常见错误模式是提高我们创建的提示质量的第一步。一些常见的错误模式包括：

- **缺乏常识：**模型可能会产生缺乏常识或与已知事实相矛盾的输出，因为它们依赖训练数据中的统计模式，而不是对概念的真正理解。
- **重复和冗长：**人工智能模型有时会陷入循环，重复短语或句子。这可能导致输出中出现冗余信息或失去连贯性。
- **敏感性：**模型可能对输入短语的微小变化非常敏感，导致输出结果的不一致。
- **离题回答：**人工智能模型可能会提供与给定提示无关的答案，这是由于无法理解上下文或问题造成的。
- **不准确的信息：**有时，模型会生成看似合理但实际错误或荒谬的答案，这是由于它们在大型语言模型中的人工智能幻觉。
- **偏见和不适当的内容：**人工智能模型可能会无意中包含训练数据中存在的偏见，从而导致有偏见或不适当的输出。
- **回避问题：**模型可能在回答复杂或具有挑战性的问题时回避问题，给出不明确或模糊的答案。

通过识别这些常见的错误模式，我们可以更好地了解提示可能需要改进的地方或 AI 模型可能存在问题的地方。

8.2 误报和漏报

误报和漏报是使用 AI 语言模型时可能发生的两种错误。了解这些错误将帮助我们设计更好的提示并提高模型的整体性能。

误报（False Positive）：误报是指模型错误地将负例识别为正例的情况。在 AI 语言模型的背景下，这可能意味着模型为某个问题生成了一个错误的答案，尽管它看起来可能是合理的。为了减少误报，可以尝试：

- **精细调整提示：**使问题更具体、明确，以便更好地引导模型。

- 结合其他模型或算法：使用多模型集成，利用不同模型的长处，降低误报风险。
- 增加验证步骤：在模型生成答案后，引入额外的验证步骤，例如将答案与知识库进行比较。

漏报（False Negative）：漏报是指模型错误地将正例识别为负例的情况。在 AI 语言模型的背景下，这可能意味着模型未能生成正确的答案，尽管正确答案在其训练数据中是可用的。为了减少漏报，可以尝试：

- 改进输入提示：提供更多上下文信息或提问方式，以便模型更好地理解问题。
- 使用多个模型或不同的参数设置：多个模型或参数设置可能会提供不同的答案，从而减少漏报。
- 用增量学习：根据用户反馈和新数据源，持续改进和调整模型，以便它能更好地捕捉到所需信息。

分析 AI 模型输出中的误报和漏报将帮助我们识别提示中的错误类型并为改进提供方向。

8.3 解决错误的策略

现在我们已经介绍了常见的错误模式以及误报和漏报的概念，让我们探讨一些解决提示设计中这些错误的具体方法。

- (1) **优化指令：**调整输入提示的措辞，使您的指令更加明确、清晰和简洁，以引导 AI 模型获得所需的输出。您还可以尝试不同的措辞或方法，看看哪种方法能产生最佳效果。

例子：将问题从“最受欢迎的游戏”修改为“2022 年全球最畅销的电子游戏”。

- (2) **提供上下文和示例：**在提示中包含相关上下文和示例可以帮助 AI 模型更好地理解任务并产生更准确的输出。确保示例多样化并能代表预期的结果。

例子：要求模型为一家律师事务所编写营销文案时，提供相关背景信息和示例：“为一家专门处理知识产权案件的律师事务所编写一段吸引客户的营销文案。例如：我们在知识产权领域拥有丰富经验，致力于保护您的创新和独特思想。”

- (3) **分阶段提示：**尝试将问题分解为多个小问题，让模型逐步回答。这有助于提高问题的可理解性，从而减少误报和漏报。

例子：将任务分为两步：“列举中国五个最大的城市。”接着询问：“这些城市中哪些城市拥有著名的科技产业园区？”

- (4) 试验温度和最大令牌：调整温度和最大令牌设置会对模型的输出产生重大影响。较低的温度会导致更加集中和确定的输出，而较高的温度会导致更具创造性和多样化的回应。试验这些设置，为您的用例找到合适的平衡点。

例子：在生成创意故事时，可以提高温度以产生更多元的输出。在生成简洁的摘要时，可以降低温度以产生更集中的输出。

- (5) 重复提问：以不同方式提问相同的问题，以获取更多答案。这有助于提高答案的多样性和质量，从而降低误报和漏报的风险。

例子：尝试以不同方式提问相同的问题：“哪个城市拥有最大的农村人口？”和“农村人口最多的城市是哪个？”

- (6) 增量学习和反馈循环：根据用户反馈和新数据源，持续改进和调整模型。这将有助于模型学习到更多的知识，从而减少误报和漏报。

例子：收集用户对于智能聊天机器人回答的反馈，调整模型以纠正常见的误报和漏报，从而改善用户体验。

- (7) 与他人协作：与其他提示工程师或领域专家分享您的发现、见解和提示设计。通过利用他人的知识和经验，协作可以带来更有效的提示设计。

例子：在一个产品推荐系统项目中，提示工程师与领域专家合作，以确保生成的产品推荐符合行业标准和客户需求。

- (8) 引入验证和过滤层：在模型生成答案后，引入额外的验证和过滤步骤。例如，与知识库进行比较，以确保答案的准确性和一致性。

例子：AI 生成了一份关于某种疾病的治疗方案，验证层可以将答案与权威医学数据库进行比较，确保提供的信息准确无误。

- (9) 自定义微调：如果 AI 模型始终难以处理特定主题或领域，请考虑使用与您的用例相关的自定义数据集来微调模型。这可以通过为模型提供更有针对性的训练数据来提高模型的性能。

例子：为了改进医学问答系统，可以使用大量医学相关的训练数据对模型进行微调，从而提高模型在该领域的准确性和可靠性。

通过应用这些策略，将能够更有效地解决提示设计中的常见错误模式、误报和漏报，从而充分利用 AI 语言模型的潜力。

在本章中，我们介绍了错误分析和故障排除在提示工程中的重要性。通过识别常见的错误模式、理解误报和漏报并实施各种错误解决策略，您将顺利地**为 AI 语言模型创建更有效的提示**。