**文本大模型**

*文本大模型，例如step-1系列，基于生成式人工智能技术，通过在大量数据上进行训练，以理解和生成自然语言。*

**文本大模型可以用于但不限于以下任务：**

* **内容生成**：生成文章、故事、对话等。
* **代码生成**：根据描述生成代码片段。
* **摘要**：将长篇文章或文档缩减为简短的摘要。
* **翻译**：将一种语言翻译成另一种语言。
* **问答**：回答各种问题，从简单的事实查询到复杂的推理问题。
* **创意写作**：创作诗歌、歌词或其他形式的创意文本。
* **语言理解**：进行情感分析、文本分类等。
* **对话系统**：与用户进行自然语言对话，模拟聊天机器人或虚拟助手。

使用这些模型时，通常需要提供一个prompt，即一段引导性的文本，它告诉模型需要执行的任务类型以及相关的上下文信息。设计一个有效的prompt是至关重要的，因为它直接影响到模型输出的质量和相关性。

**多模态大模型**

*多模态大模型，在文本大模型基础上，增加了对多种模态（如图像、音频、视频等）数据的处理能力，如step-1v系列，目前能够理解处理文本以及图像。 多模态大模型的核心优势在于更符合人类认知世界的习惯，因为人类通过多种感官接收多模态信息，这些信息通常是互补和协同的，使用多模态大模型可以更好地理解这些信息以及进行推理。*

**多模态大模型可以用于但不限于以下任务：**

* **图像到文本的翻译**：为图像生成描述性文本。
* **对话系统**：结合文本和图像信息回答复杂问题。
* **创作**：根据图像和文本创作文学作品。
* **医疗影像诊断**：结合医学影像（如X光片、MRI图像）和病人的病历文本，多模态模型可以帮助医生更准确地诊断疾病。
* **搜索引擎**：搜索引擎可以利用多模态技术来理解用户的查询意图，结合文本、图像，提供更相关的搜索结果。
* **自动驾驶**：自动驾驶汽车可以使用多模态模型来处理来自多个传感器（如摄像头、雷达、激光雷达）的数据，以更好地理解周围环境并做出驾驶决策。