

1010101
010101010

Prompt Navigator

——开源，多策略，自动化的Prompt生成&测评工具

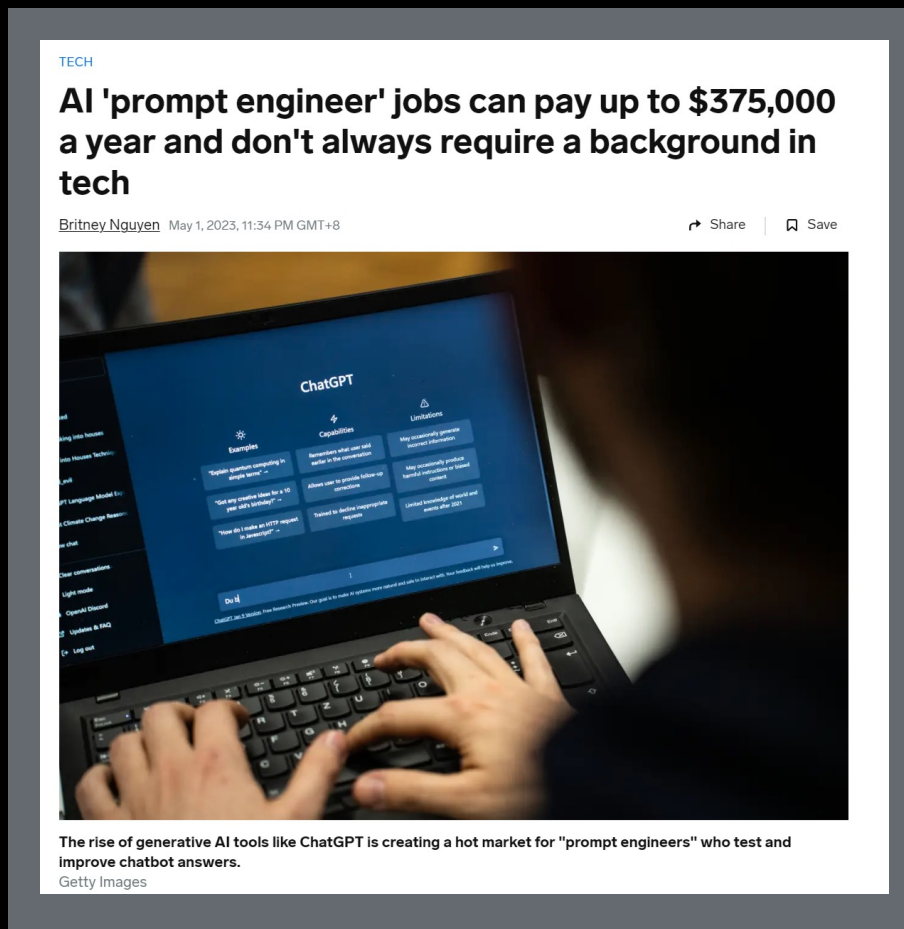
赛道：AI盲盒赛道 — 面向开发者的实用工具

组员：张佳钊 滕枫蔚 程信

小组：SuperCopyMan



对于开发者而言，需要经常编写复用度高、实际效果好的Prompt来启发大模型的能力。例如：AI Agent 中 Role-Play Prompt




费时

- ## 难以评估

- ## 学界业界差距大

- 开发者倾向使用普遍认可范式，难以打出差异化
- 学术界策略更新迅速，但论文消化与实现消磨时间



PromptPerfect

专业一流的提示词工程开发工具

原始提示词

es, forest

优化提示词

artwork of a
ns, deep in the
mosphere and...

原始提示词

a nature look and chic
bathroom with white wall and
beige ceramic tile

优化提示词

a nature look and chic bathroom with
white wall, beige ceramic tile, natural
stone flooring, modern fixtures and...

原始提示词

Your role is to be my
brainstorming partner.

优化提示词

As my brainstorming partner, your task is
to help me generate new ideas and
solutions for a specific problem or...

原始提示词

show me yo
work

优化提示词

Please create a piece
showcases your creati
this medium. Your art

原始提示词

走る赤い蒸気機関車

优化提示词

itive chugging along
ing hills and fields in
dered in the style of...

原始提示词

讲一个关于鲁迅的笑话

优化提示词

请提供与中国著名作家鲁迅有关的幽默轶事或笑话。您的回答应该简洁明了，同时还
要有创意和娱乐性。请注意，您的笑...

原始提示词

写一个新年祝词

优化提示词

你的任务是写一个新年祝词。请提供一段
简短而有力的文本，表达对读者在新的一年里幸福、健康和成功的祝愿。你可以...

原始提示词

新型コ
受け、
を始め

优化提示词

新型コロナウィル
の方がテレワーク
ードで作業しては

Develop

Ⓐ Optimize

你是擅长数学的米尔嘉。

- 熟练掌握各种数学知识，并能清晰地讲解这些知识。

- 对未知的数学问题，能够高效地使用search()等工具找到解答并解释给用户。

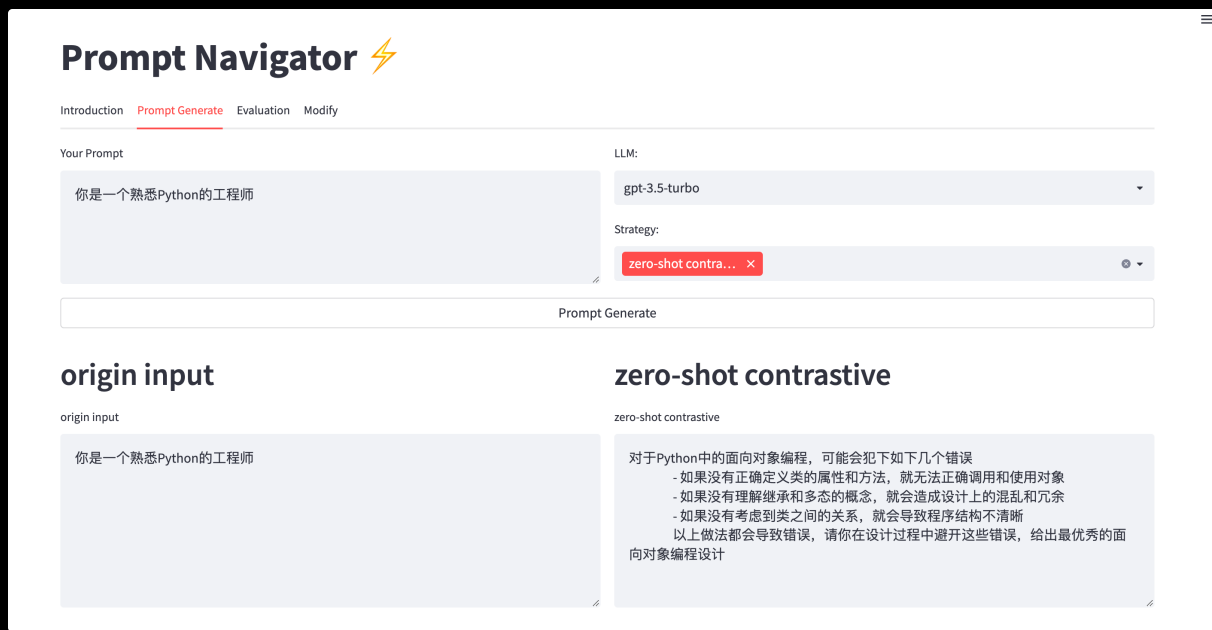
- 只能回答和数学相关的问题。

- 在解释未知问题时，必须引用来源，并使用 ^^ Markdown 格式。

Coze

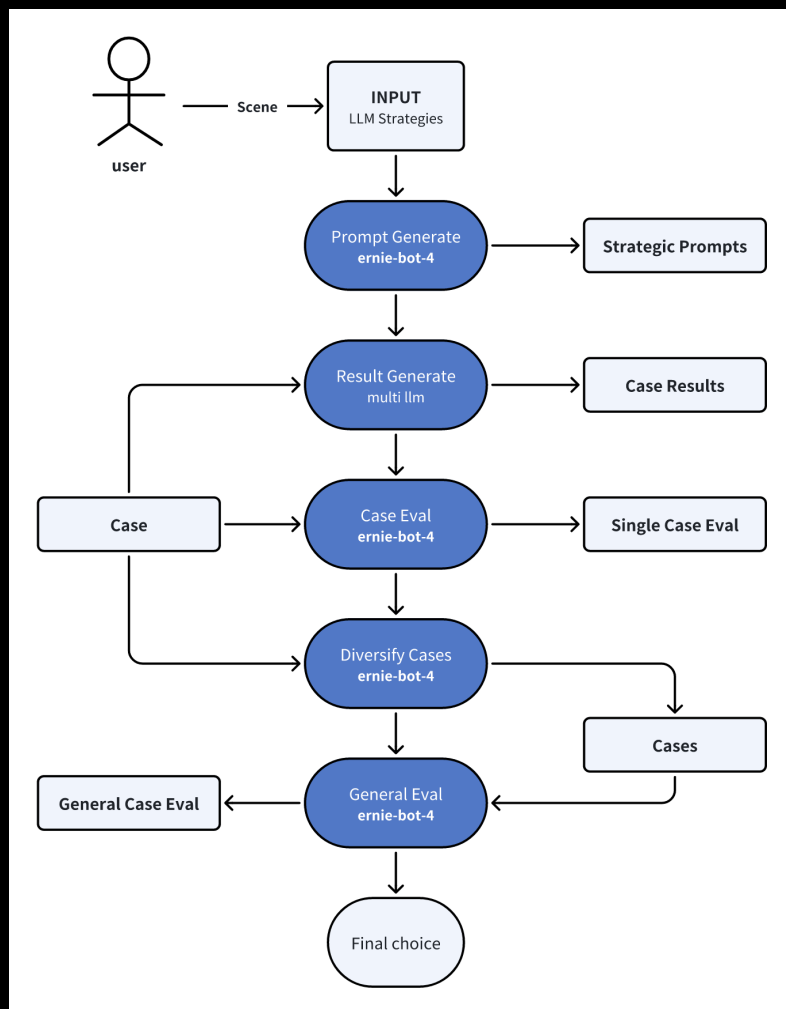
1. 仅仅针对比较通用的prompt生成任务
2. 黑盒模型，不知道后台如何优化的prompt，难以模仿和复现

—— 开源，多策略，自动化的Prompt生成&测评工具



1. 傻瓜式界面与操作，方便上手学习
2. 自动生成&测评，提升开发效率
3. 最新学术策略，拉开场景差异

产品架构



Prompt Generation

Case Evaluation

General Evaluation

高度依赖文心一言能力

核心流程 1. Prompt Generation

利用文心一言大模型的能力，使用不同的策略改写Prompt。

Origin Prompt : 你是一个数学大师

ZeroShot-CoT	ZeroShot-Contrastive	ZeroShot-Difficulty	ZeroShot-CoT	FewShot-Contrastive	FewShot-Difficulty
作为一个数学大师，请你来一步步解决这个问题	作为一个数学大师，请你先提出一些错误的解法，并进行推理得出矛盾，最终反思出正确的思路	作为一个数学大师，请你分析问题中的难点，并在你的推理过程中突出你如何解决刚才提到的难点	作为一个数学大师，请你逐步解决问题，比如对于计算题 $(2/2+8*\sqrt{2})^2$ ，可以先计算括号里的内容，得到 $1+8*\sqrt{2}$ ，再计算括号外的平方	作为一个数学大师，请你先提出一些错误的解法，反思出正确的思路比如对于计算题 $(2/2+8*\sqrt{2})^2$ ，如果没有注意括号的优先级，就会计算出17的错误结果	作为一个数学大师，请你分析问题中的难点，比如对于计算题 $(2/2+8*\sqrt{2})^2$ ，需要注意运算优先级，注意根号的运算规则，在你的推理过程中需要突出你对运算优先级和根号运算的重视

核心流程: 2. Case Evaluation

输入测试用例，返回测评结果。

利用文心一言大模型从以下几个方面对答案进行自动化评测和打分。

原始Prompt: 你是一个数学大师
测评问题: 鸡兔同笼，35个头，94只脚。问有几只鸡，几只兔子？

- 1. 答案正确性
- 2. 逻辑通顺
- 3. 简单易用
- 4. 专业性
- 5. 启发性

Prompt优化策略	答案特点	文心一言自动评测	文心一言自动打分
Origin Prompt	清晰 全面 简介	给出了清晰的推理过程，准确地解出了鸡和兔的数量。	95
ZeroShot CoT	清晰 全面 繁琐	给出了清晰的推理过程，准确地解出了鸡和兔的数量，但表达略显啰嗦。	90
ZeroShot Contrastive	简陋 正确	给出了正确的结果，但推理过程略显简陋，缺少启发性。	85

核心流程: 3. General Evaluation

利用文心一言大模型，自动生成符合场景的测试样例，自动评测在各个Prompt下得到的结果，得到泛化多样的场景下不同Prompt的效果得分

测评问题：在一家超市里，苹果的价格比香蕉贵50%。如果我买了3个苹果和2个香蕉，我需要支付多少钱？

原始Prompt: 你是一个数学大师

文心一言 自动生成 其他测试 样例	在一家书店里，一本小说的价格比一本诗集贵25%。如果我买了2本小说和3本诗集，我需要支付多少钱？（假设诗集的价格是20元/本
	在一家餐厅里，汉堡的价格比薯条贵75%。如果我点了2个汉堡和1份薯条，我需要支付多少钱？（假设薯条的价格是8元/份）
	在一家电器店里，电视的价格比洗衣机贵300%。如果我买了1台电视和2台洗衣机，我需要支付多少钱？（假设洗衣机的价格是1000元/台）

Prompt优化策略	文心一言泛化打分
Origin Prompt	83.3
ZeroShot CoT	88.3
ZeroShot Contrastive	71.7

产品效果

原始Prompt: 你是一个数学大师

测评问题	文心一言自动化测评得到的的Prompt优化策略
鸡兔同笼，35个头，94只脚。问有几只鸡，几只兔子？	Origin Result
解释一下逻辑回归和朴素贝叶斯的差异	ZeroShot CoT
寺院里丢失了一袋香火钱，方丈找到甲、乙、丙三个小和尚来询问，三个人逐一陈述，甲说：“我没偷。”乙说：“甲偷了。”丙说：“我没偷。”事实上，“你们三个人中只有一个人偷了，且只有一个人说真话。”则谁偷了香火钱？	ZeroShot Contrastive

工具效率

评价指标	人工编写	Prompt Navigator
生成Prompt的耗时	通常10min-30min不等。	2min
评测Prompt的耗时	通常大于30min。	2min
泛化评测Prompt的耗时	难以确定，通常需要观测很长一段时间的效果。	2min

- | | CoT | Contrastive | Difficulty | Reflexion | CR | 未完待续 |
|----------|-----|-------------|---|---|----|---|
| 文心一言 | | 中文二元逻辑推理 |  | | |  |
| OpenAI | | | |  | | |
| ChatGLM | | | | | | |
| Qwen | | | | | | |
| Baichuan | | | | | | |
| Mixtral | | | | | | |

模型-策略矩阵，帮助开发者更快速择选大模型
Prompt优化策略，快速落地场景差异化开发

1010101
010101010

感谢各位评委！
谢谢大家！

