PromptAgent: Strategic Planning with Language Models Enables Expert-level Prompt Optimization

Xinyuan Wang[1*], Chenxi Li[1*], Zhen Wang[12*], Fan Bai[5], Haotian Luo[2], Jiayou Zhang[2], Nebojsa Jojic[2], Eric P. Xing[24], Zhiting Hu[1]

[1]UC San Diego, [2]MBZUAI, [3]Microsfot Research, [4]CMU [5]Georgia Tech

We Need Expert Prompts 👩‍⚕️👨‍🔬👨‍🔧

Expert prompts: Highly effective, task-specific prompts that are often heavily engineered by domain experts

- Unleash the full potential of **very large LLMs**, e.g., GPT-3.5/4
- Spearhead the **next era of prompt engineering**, with more powerful LLMs that can understand intricate instructions

The following example shows how the expert prompt improves perf. with **richer domain knowledge and structured guidance**

Biomedical Task Input

Linkage studies in this family suggested a close linkage between the c2 deficiency gene and genes coding for B18 , Dw2 , and BfS antigens . "c2 deficiency" is a disease mention to be extracted

Ordinary User Prompt

Extract the disease or condition from the sentence, if any is mentioned.

Prompt From Sampling-Based Method

If any disease or condition is mentioned in the sentence, extract it.

Expert-level Prompt

Task Description	You're tasked with extracting diseases or conditions from the given sentence …
Domain Knowledge	Avoid associated elements: inheritance patterns, genes or gene loci (like PAH) …
Solution Guidance	Consider both specific diseases and broader categories, common abbreviations …
Exception Handling	The term 'locus' should be recognized as a genomic location, not a disease name …
Output Formatting	Provide the identified diseases in this format: {entity_1,entity_2, ...} …
…	

Ordinary User/Sampled Prompt Output

c2 deficiency gene ✗

Expert Prompt Output

c2 deficiency ✓

No More Prompt Engineering? 🤔

Manual expert prompt engineering? 🤷

- A unique blend of domain knowledge and intuition for LLMs
- Ad-hoc human-chatbot interactions with tedious trial-and-errors

Existing automatic prompt engineering? 😬

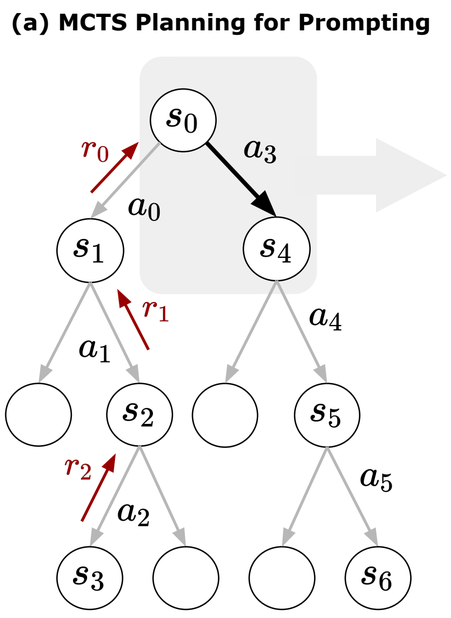
- Tend to **overlook the depth of domain knowledge**
- Struggle to efficiently **explore the vast prompt space**

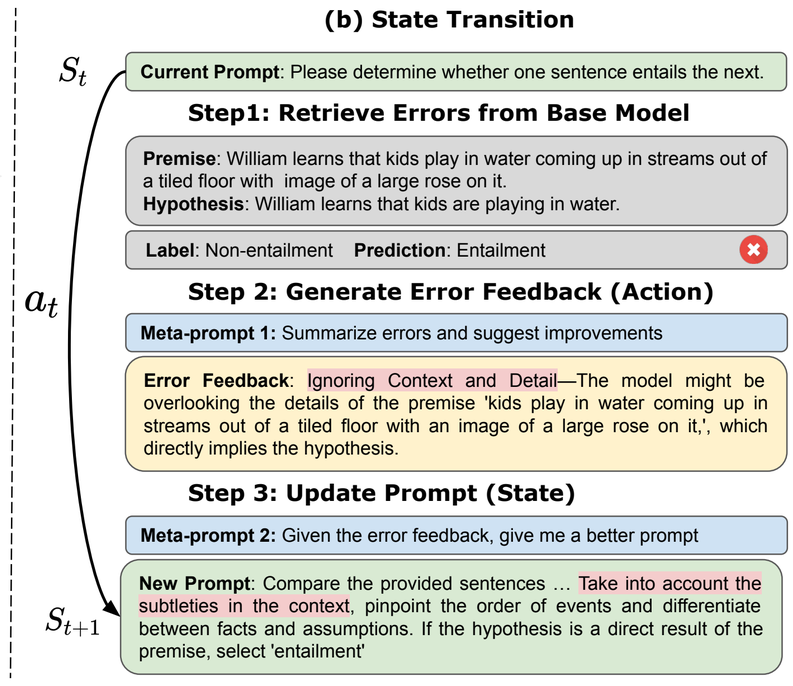
Goal: Automatically craft expert-level prompts equivalent in quality to those handcrafted by domain experts

PromptAgent Saves Your Time 🙌

Key innovations (check more details in the paper):

1. Reframing prompt optimization as **a strategic planning problem**
   - Efficiently explore prompt space with **lookahead and backtrack**
2. **Self-reflection on model errors** to mirror human's trial-and-error
   - Effectively induce **valuable domain insights** with error feedbacks



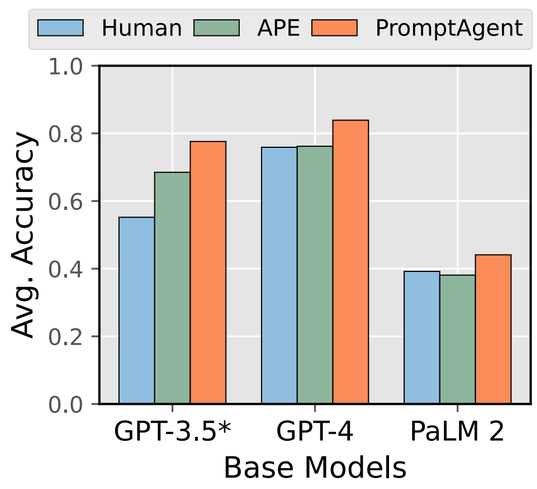


Superior Prompts Unlocked 🔓🚀

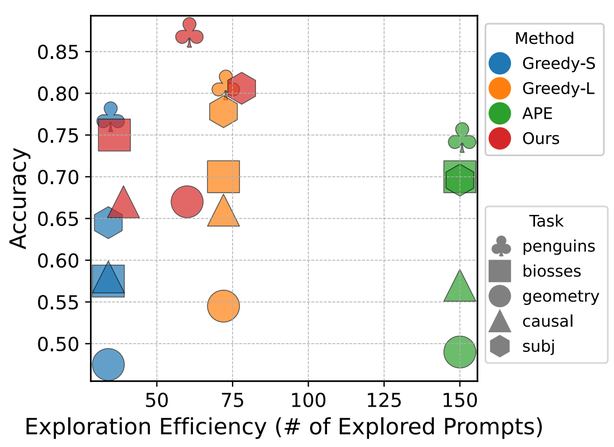
1. Great empirical gains over 12 tasks spanning three domains: **BIG-Bench Hard, domain-specific and general NLP tasks**

	Domain-specific Tasks				General NLU Tasks			
	NCBI (F1)	Biosses	Med QA	*Avg.*	Subj	TREC	CB	*Avg.*
Human (ZS)	0.521	0.550	0.508	0.526	0.517	0.742	0.714	0.658
Human (FS)	0.447	0.625	0.492	0.521	0.740	0.742	0.429	0.637
CoT (ZS)	0.384	0.425	0.508	0.439	0.656	0.63	0.750	0.679
CoT	0.376	0.675	0.542	0.531	0.670	0.784	0.643	0.699
GPT Agent	0.125	0.625	0.468	0.406	0.554	0.736	0.339	0.543
APE	0.576	0.700	0.470	0.582	0.696	0.834	0.804	0.778
PromptAgent	**0.645**	**0.750**	**0.570**	**0.655**	**0.806**	**0.886**	**0.911**	**0.868**

2. **Transferrable** expert prompts across various base models



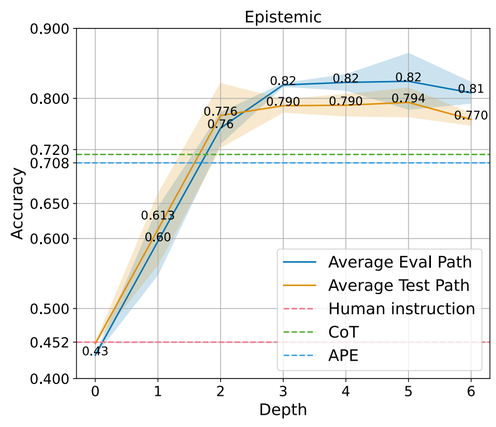
3. Great balance of **exploration efficiency** and performance



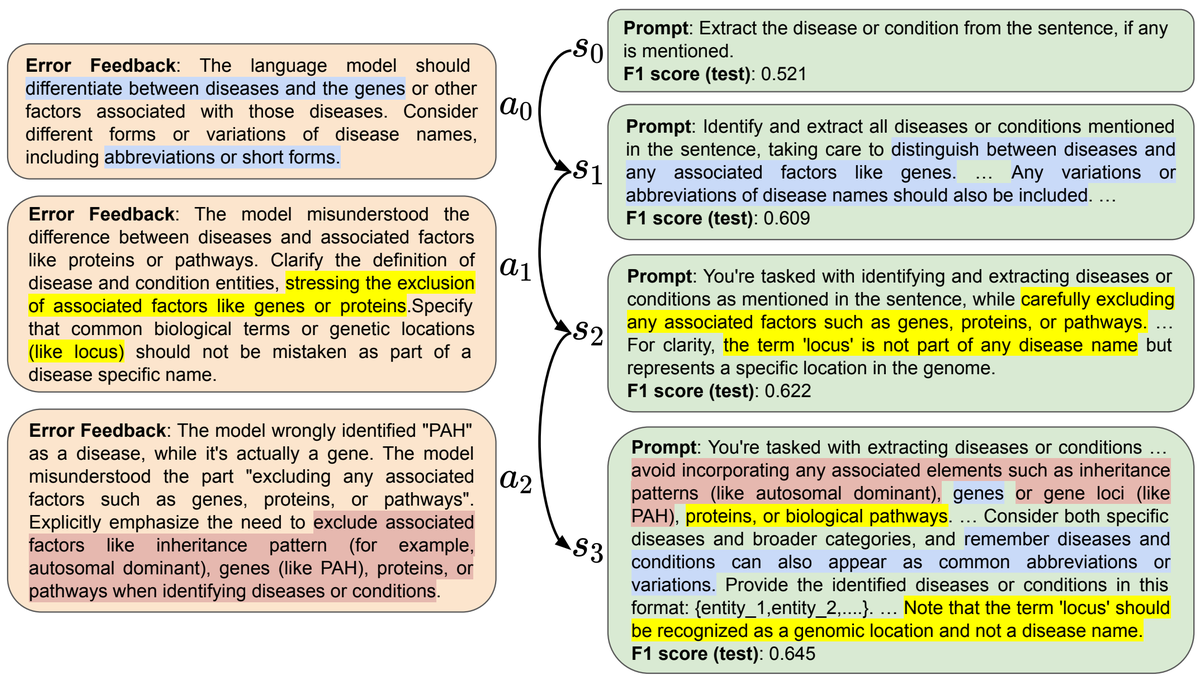
4. Ablation study on search variants, showing the **power of strategic planning** (MCTS)

	MC	Beam	Greedy	MCTS (Ours)
Penguins	0.772	0.823	0.810	**0.873**
Biosses	0.575	0.675	0.700	**0.750**
Geometry	0.490	0.610	0.545	**0.670**
Causal	0.650	0.610	0.660	**0.670**
Subj	0.692	0.765	0.778	**0.806**
Average	0.635	0.697	0.698	**0.754**

5. **Convergence** analysis reveals a stale learning dynamic



6. A prompt **evolution trajectory**, showing how error feedbacks (highlighted colors) are accumulated into better expert prompts



Road Ahead with PromptAgent 🛣️

- PromptAgent serves as **a principled framework** to study prompt optimization by unifying prompt sampling and rewarding
- Unlike discovering magic/local prompt variants, expert-level prompting is still **an untapped area** to solve challenging problems