

# Matryoshka Pilot

Learning to Drive Black-Box LLMs with LLMs

Changhao Li<sup>1\*</sup>, Yuchen Zhuang<sup>1\*</sup>, Rushi Qiang<sup>1</sup>, Haotian Sun<sup>1</sup>  
, Hanjun Dai<sup>2</sup>, Chao Zhang<sup>1</sup>, Bo Dai<sup>1</sup>

\* Equal Contribution, <sup>1</sup>Georgia Institute of Technology, <sup>2</sup>Precur AI

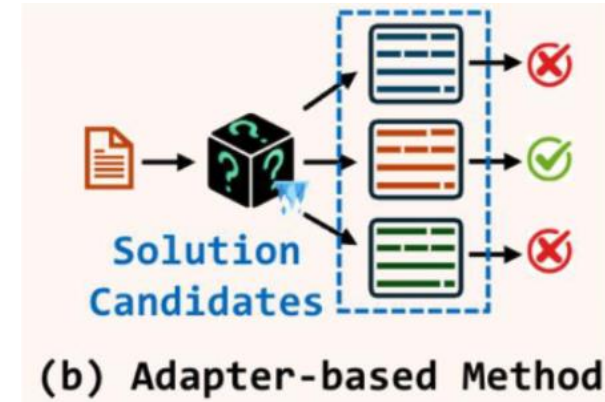


# How to improve black-box LLM performance?



- Chain-of-Thought [Wei et al., 2023]
- Least-to-Most [Zhou et al., 2022]
- AdaPlanner [Sun et al., 2023]

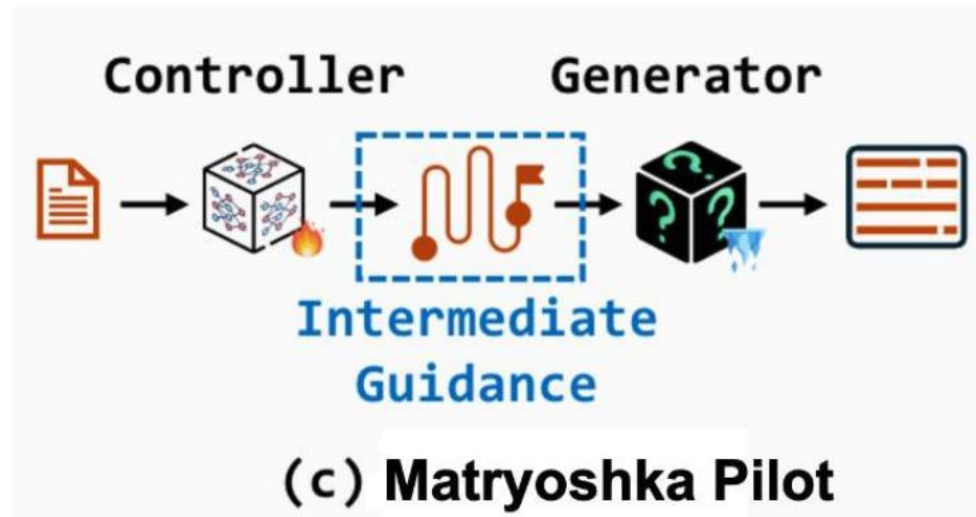
...



- Tree-of-Thoughts [Yao et al., 2023]
- RAP [Hao et al., 2023]
- Bbox-adapter [Sun et al., 2024]

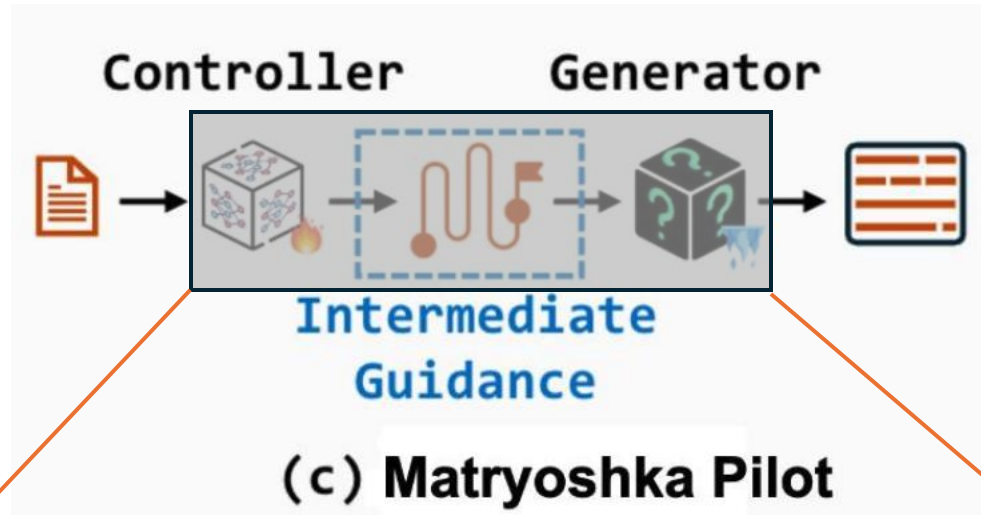
...

# How to improve black-box LLM performance?

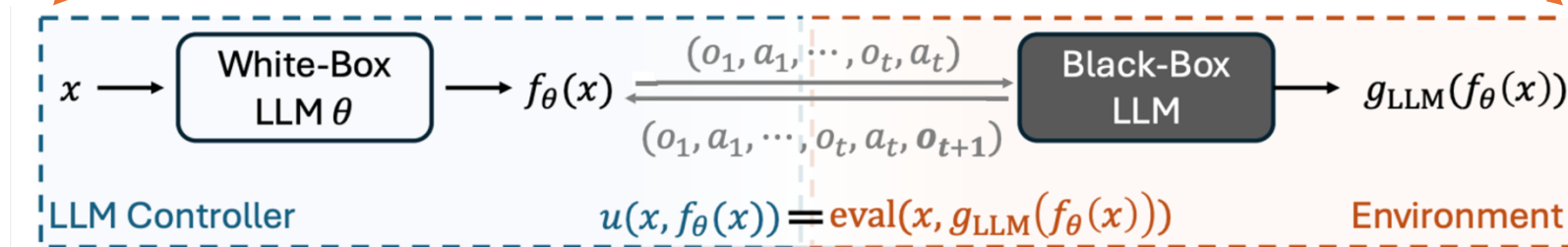


**Controller-Generator Framework**

# How to improve black-box LLM performance?



## Controller-Generator Framework



Motivation

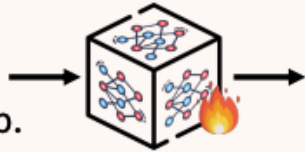
Matryoshka Pilot

Experiment

## Planning (AlfWorld): Text-based Virtual Household Environment

**Your task is to:**

clean some soapbar  
and put it in countertop.



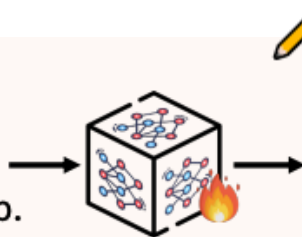
**Task Description**

**Controller**

## Planning (AlfWorld): Text-based Virtual Household Environment

**Your task is to:**  
clean some soapbar  
and put it in countertop.

**Task Description**



**Controller**

**General plan:** I need to get a list of receptacles to find the soapbar, take the soapbar to a sinkbasin, clean it, and put it in a countertop.

**[Step 1]** get a list of receptacles where the soapbar is likely to appear.

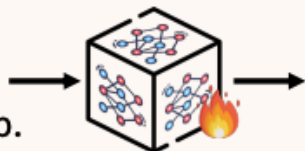
**[Step 2]** .....

**Task Decomposition**

**Planning (AlfWorld): Text-based Virtual Household Environment**

**Your task is to:**  
clean some soapbar  
and put it in countertop.

**Task Description**



**Controller**

**General plan:** I need to get a list of receptacles to find the soapbar, take the soapbar to a sinkbasin, clean it, and put it in a countertop.

**[Step 1]** get a list of receptacles where the soapbar is likely to appear.

**[Step 2]** .....

**Task Decomposition**

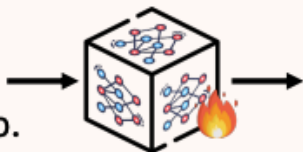


**Generator**

### Planning (AlfWorld): Text-based Virtual Household Environment

**Your task is to:**  
clean some soapbar  
and put it in countertop.

**Task Description**      **Controller**



**General plan:** I need to get a list of receptacles to find the soapbar, take the soapbar to a sinkbasin, clean it, and put it in a countertop.

**[Step 1]** get a list of receptacles where the soapbar is likely to appear.

**[Step 2]** .....

**Task Decomposition**



**Generator**

```
def solution(agent, start_from=1):  
    # General plan: .....
```

**Generation**

```
    if start_from <= 1:  
        # [Step 1] .....
```

**Execution** ✓

```
        answer = ask('.....')  
        recep_to_check = literal_eval(answer)  
    .....
```

# How to improve intermediate guidance?

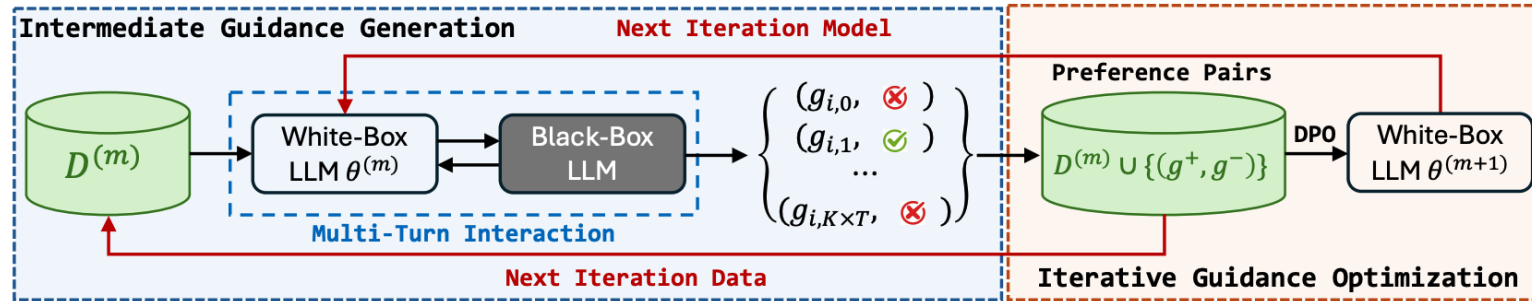


Figure 4: Overview of iterative guidance optimization. By iteratively updating both the model and the reference policy, M-Pilot progressively refines its intermediate guidance.

## Iterative Guidance Optimization

*Conceptualizing black-box LLM as environment*

Motivation  
Matryoshka Pilot  
Experiment

- Iterative Guidance Optimization + Controller-Generator Framework Enhances Black-box LLM performance across diverse tasks

Dataset (→)	LaMP-1		LaMP-2N		LaMP-2M		LaMP-3		LaMP-4		
Method (↓)	Acc. ↑	F-1 ↑	Acc. ↑	F-1 ↑	Acc. ↑	F-1 ↑	MAE ↓	RMSE ↓	R-1 ↑	R-L ↑	BLEU ↑
gpt-4o-mini	0.514	0.513	0.655	0.473	0.413	0.325	0.371	0.673	0.132	0.116	0.992
RAG (k=1) [34]	0.626	0.624	0.733	0.539	0.444	0.378	0.311	0.631	0.141	0.126	1.296
RAG (k=4) [34]	<u>0.632</u>	<u>0.632</u>	0.792	<u>0.611</u>	<u>0.502</u>	0.430	<b>0.272</b>	<b>0.579</b>	<u>0.161</u>	<u>0.146</u>	<u>2.953</u>
PAG [32]	0.624	0.624	0.775	0.559	0.496	<u>0.443</u>	0.316	0.645	0.143	0.130	1.968
M-Pilot	<b>0.640</b>	<b>0.639</b>	<b>0.823</b>	<b>0.607</b>	<b>0.527</b>	<b>0.465</b>	<u>0.277</u>	<u>0.581</u>	<b>0.174</b>	<b>0.160</b>	<b>4.298</b>
w/o IDPO	0.611	0.611	<u>0.807</u>	0.575	0.496	0.432	0.311	0.636	0.131	0.120	1.341

Methods (↓)	Tasks (→)	Pick	Clean	Heat	Cool	Exam	Pick2	All
BUTLER [39]		46.00	39.00	74.00	<b>100.00</b>	22.00	24.00	37.00
ReAct [55]		37.50	64.52	69.57	42.86	38.89	17.65	47.76
Reflexion [38]		50.00	41.94	65.22	52.38	66.67	47.06	52.99
AdaPlanner [41]		<b>100.00</b>	<b>93.55</b>	78.26	95.24	66.67	<b>88.24</b>	88.06
M-Pilot		<b>100.00</b>	<b>93.55</b>	<b>100.00</b>	95.24	<b>100.00</b>	<b>88.24</b>	<b>96.27</b>
w/o 2 <sup>nd</sup> -round IDPO		100.00	93.55	100.00	100.00	83.33	88.24	94.78
w/o 1 <sup>st</sup> , 2 <sup>nd</sup> -round IDPO		100.00	93.55	86.96	95.24	55.56	88.24	88.06
w/o Guidance Optimization		100.00	93.55	91.30	85.71	11.11	88.24	81.34

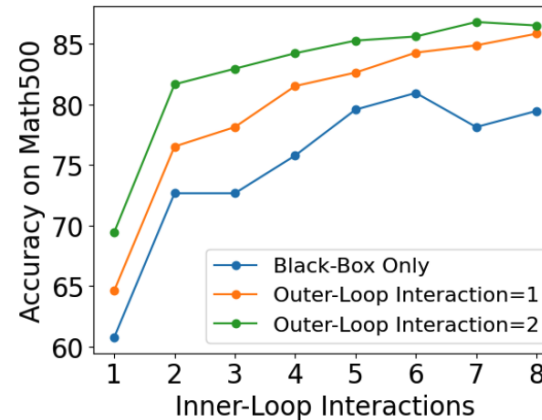
Dataset (→)	GSM8K		GSM-HARD	
Method (↓)	gpt-3.5	4o-mini	gpt-3.5	4o-mini
CoT	0.809	0.932	0.406	0.500
Least-to-Most	0.811	0.908	0.425	0.498
PAL	0.802	0.920	0.638	0.748
PAL-Self-Debug	0.864	0.943	0.701	0.774
M-Pilot	<b>0.931</b>	<b>0.964</b>	<b>0.761</b>	<b>0.801</b>
w/o IDPO	<u>0.896</u>	<u>0.954</u>	<u>0.729</u>	<u>0.780</u>

LaMP with *Gpt-4o-mini* as Black-Box Generator.

AlfWorld with *Gpt-3.5-turbo* as Black-Box Generator.

GSM with *Gpt-3.5-turbo* / *GPT-4o-mini* as Black-Box Generator.

Dataset (→)	LaMP-1		LaMP-2N		LaMP-2M		LaMP-3		LaMP-4		
Method (↓)	Acc. ↑	F-1 ↑	Acc. ↑	F-1 ↑	Acc. ↑	F-1 ↑	MAE ↓	RMSE ↓	R-1 ↑	R-L ↑	BLEU ↑
M-Pilot (4o-mini)	<b>0.640</b>	<b>0.639</b>	<b>0.823</b>	<b>0.607</b>	<b>0.527</b>	<b>0.465</b>	<u>0.277</u>	<u>0.581</u>	<b>0.174</b>	<b>0.160</b>	<b>4.298</b>
gpt-3.5-turbo	0.590	0.589	0.790	0.594	0.399	0.325	0.357	0.693	0.166	0.150	3.433
Plug-and-play (gpt-3.5)	<b>0.594</b>	<b>0.593</b>	<b>0.798</b>	<b>0.609</b>	0.469	0.412	<b>0.286</b>	<b>0.599</b>	<b>0.176</b>	<b>0.161</b>	<b>4.222</b>
w/o IDPO (gpt-3.5)	0.585	0.585	0.790	0.608	<b>0.472</b>	<b>0.425</b>	0.334	0.670	0.160	0.147	3.015
gemi-1.5-flash	0.518	0.510	0.700	0.498	0.368	0.279	0.546	0.825	0.135	0.113	1.494
Plug-and-play (gemi)	<b>0.573</b>	<b>0.565</b>	<b>0.825</b>	<b>0.615</b>	0.504	<b>0.418</b>	<b>0.298</b>	<b>0.614</b>	<b>0.183</b>	<b>0.170</b>	<b>5.002</b>
w/o IDPO (gemi)	0.568	0.561	0.811	0.602	<b>0.505</b>	0.411	0.365	0.715	0.164	0.150	3.439



MATH500 with *Gpt-3.5-turbo* as Black-Box Generator.

Plug-and-play results of *GPT-3.5-turbo* and *Gemini-1.5-flash* as Black-Box Generators on LaMP

Motivation  
Matryoshka Pilot  
Experiment

- Matryoshka Pilot demonstrates strong efficiency in both cost and sample usage.

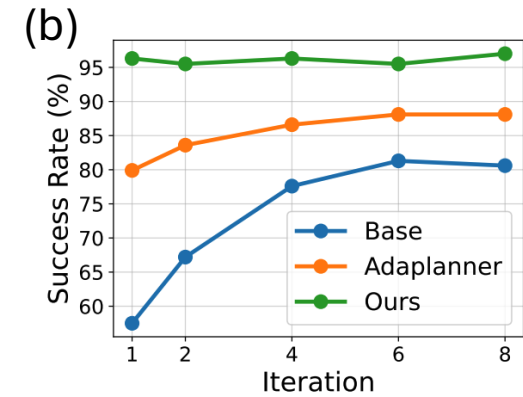
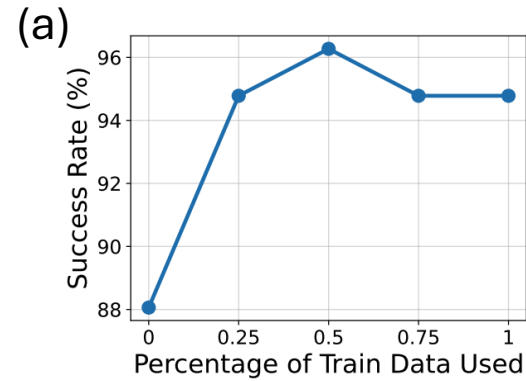
(a)

Method	API cost (\$)	Task Performance
M-Pilot	0.818	97.8
Adaplanner [41] (w/o controller)	1.151	88.8

(b)

Method	API cost (\$)	Task Performance
M-Pilot	1.740	93.1
Self-Consistency [49] (w/o controller)	4.946	81.3

Comparison of cost and task performance on (a) AlfWorld and (b) GSM8K



Success rate (%) of Matryoshka Pilot versus (a) training data size and (b) number of iteration loops

*Thanks !*

## Matryoshka Pilot: Learning to Drive Black-Box LLMs with LLMs

Changhao Li<sup>1\*</sup>, Yuchen Zhuang<sup>1\*</sup>, Rushi Qiang<sup>1</sup>, Haotian Sun<sup>1</sup>  
, Hanjun Dai<sup>2</sup>, Chao Zhang<sup>1</sup>, Bo Dai<sup>1</sup>

\* Equal Contribution, <sup>1</sup>Georgia Institute of Technology, <sup>2</sup>Precur AI

