

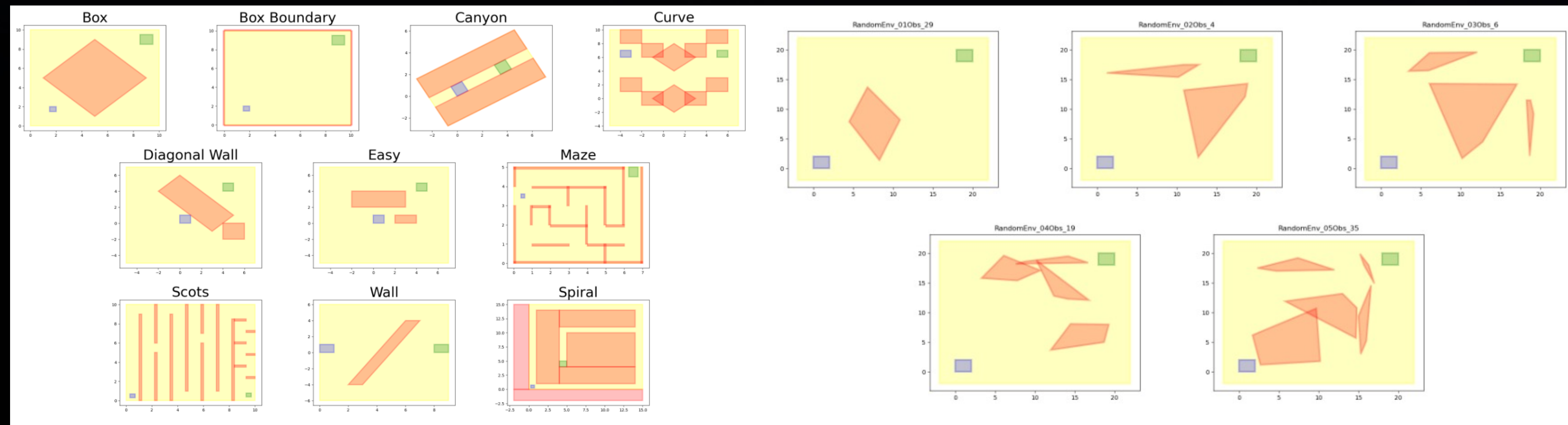
Can LLMs plan paths with correct hints?

Sayan Mitra, University of Illinois, Urbana-Champaign

<https://mitras.ece.illinois.edu/>

Erik Bowen Wu, Karlsruhe Institute of Technology

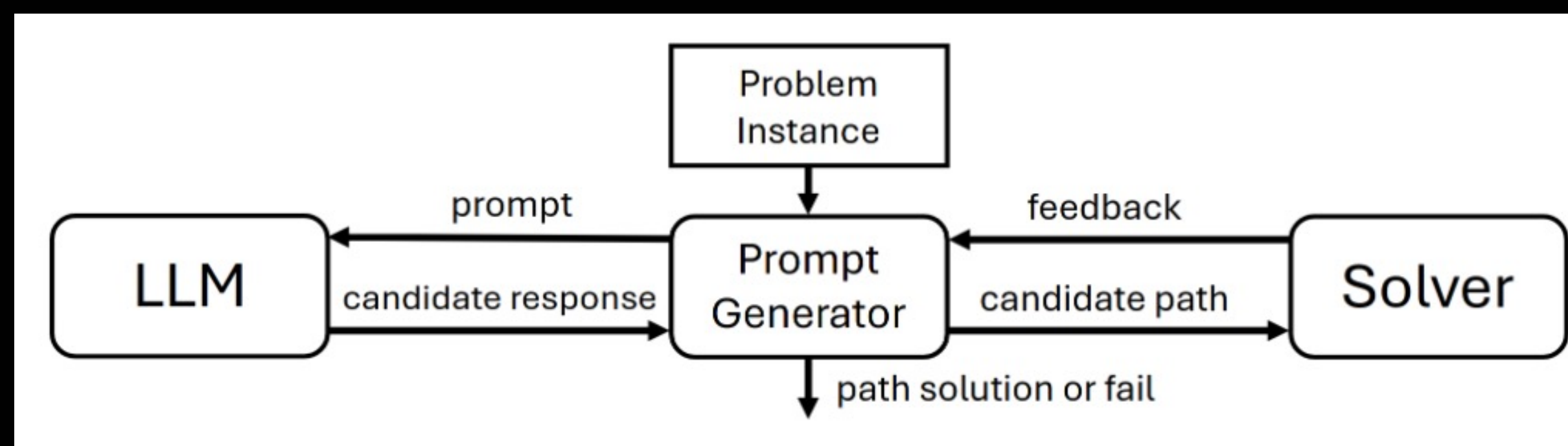
Summary. Good plans need global reasoning, and bad plans can be easily identified. Can Large Language Models, as approximate knowledge repositories, with the aid of solvers that can point out mistakes, be used to effectively solve hard planning problems? Our experiments with 3 LLMS (GPT4-0, Gemini Pro, Claude Sonnet), finetuning, and 3 types of hinting, including visual hints, suggest otherwise.



Path planning problem:

Find a sequence of waypoints from the initial set to the goal set that does not collide with any of the obstacles.

We used 12 problems from literature and dozens of randomly generated problems and tried to solve them with 3 LLMS and iterative hints



Hinting strategies:

Initial prompt describes the generic planning problem and the representations. Iteratively for each of LLM's candidate solutions:

- Collision hints
- Longest correct subpath
- Free space hints
- Visual / picture hints
- Fine-tuning

Collision hints

Problem	Gemini Pro 1.5			GPT-4o			Claude 3.5		
	S%	N	PL	S%	N	PL	S%	N	PL
Box	90	10.9	4.4	50	8.6	5.2	0	-	-1
Box B.	100	1.3	3.7	100	1.0	3.0	100	1.0	3.6
Canyon	0	-	-	100	2.2	4.7	70	7.1	5.7
Curve	40	8.0	6.5	70	9.1	5.4	10	3.0	9.0
D. Wall	70	8.9	5.0	0	-	0	-	-	-
Easy	100	3.3	4.5	100	2.6	5.2	0	-	-
Maze 2D	0	-	-	0	-	0	-	-	-
Scots	20	13.5	12.5	0	-	0	-	-	-
Spiral	60	14.2	6.2	0	-	0	-	-	-
Wall	100	1.6	4.2	100	1.6	4.5	10	11.0	5.0
1 Obs	95	1.7	4.1	90	1.7	4.9	85	1.9	4.5
2 Obs	95	2.5	4.9	50	2.3	5.7	60	1.8	4.3
3 Obs	75	2.9	4.9	15	4.0	6.3	25	3.4	4.6
4 Obs	45	3.7	4.9	5	1.0	5.0	0	-	-
5 Obs	30	3.7	4.8	10	3.0	7.0	15	3.0	5.0

Problem	Gemini Pro 1.5			GPT-4o			Claude Sonnet 3.5		
	S%	N	PL	S%	N	PL	S%	N	PL
Box	90	8.8	6.3	50	17.3	8.0	30	17.3	5.0
Box B.	100	1.3	3.7	100	1.0	3.9	100	1.2	4.1
Canyon	30	15.9	5.3	100	2.6	4.6	80	5.6	5.8
Curve	30	17.9	11.0	100	4.3	5.8	20	16.5	8.5
D. Wall	90	8.9	7.7	0	20.0	-	0	-	-
Easy	100	3.9	6.6	100	6.4	5.6	20	12.5	6.0
Maze	10	20.1	13.0	0	20.0	-	0	-	-
Scots	0	-	-	10	18.7	10.0	0	-	-
Spiral	30	17.1	9.0	0	20.0	-	0	-	-
Wall	100	8.0	6.5	100	5.4	6.9	10	20.0	7.0
1 Obs	100	1.9	4.8	70	1.8	5.1	55	2.1	4.4
2 Obs	60	2.8	6.3	50	2.6	5.5	35	1.4	4.1
3 Obs	65	4.1	6.7	20	2.2	6.2	20	3.2	5.7
4 Obs	20	5.0	7.5	10	5.5	9.0	10	2.0	4.0
5 Obs	35	3.8	6.8	15	3.0	7.0	15	5.3	6.0

Bubeck, Chandrasekaran, et al. "Sparks of AGI: Early experiments with gpt-4," 2023.

Song, "Llm-planner: Few-shot grounded planning for embodied agents with LLMs," ICCV'23

Kambhampati, et al. "Position: Llms can't plan, but can help planning in llm-modulo frameworks," ICML'24.

Wei, et al. "Chain-of-thought prompting elicits reasoning in large language models," Neurips'22.

Stechly, Valmeekam, and Kambhampati, "Chain of thoughtlessness? an analysis of cot in planning," '24.

- (1) With hints LLMs can solve the moderate problems, but not the harder problems
- (2) More hints better, but just collision is a lot better than no hints; images did not help.
- (3) Fine-tuning helped marginally and improves the quality of the solutions.

Collision + Free Space + Longest subpath



Computational Cybersecurity in Compromised Environments

2024 Fall Workshop | September 17-19 | SRI International in Menlo Park, CA