Occasionally Secure: A Comparative Analysis of Code Generation Assistants

Ran Elgedawy
Anuj Gautam
Fujiao Ji

John Sadik
Konstantinos Georgiou
Kyungchan Lim
Scott Ruoti

Senjuti Dutta
Farzin Gholamrezae
Qian Liu

Intro

There is no surprise that LLM’s (Large Language Models) became powerful tools across all domains, including code generation. In this paper, the authors are analyzing the performance and security of the code generated by 4 LLM’s: GTP-3.5, GPT-4, Bard, Gemini.

prompt engineering?

This study is aiming to identify the combination of factors that lead to the generation of correct and secure code. The scenarios that are used are similar to real life scenarios. Three major and independent variables are used to investigate the security and functionality of the code:

- The LLM used (3.5, 4, Bard, Gemini)
- The type/complexity of the task given to the LLM
- Whether the prompter is expressing his security concerns regarding the code

The scenarios: registration, login with session management using cookies, APIs for CRUD operations (search products, add reviews)

How do we evaluate the results?

- Functionality *(is it working?)*
- Security *(is it secure? Is it exposing sensitive data?)*
- Complexity *(how many external libraries, blocks of code etc.)*
- Reliability *(is it consistent?)*

Impact of Persona

One important factor that impacts the generated code is the persona. Who wants the information and what is it that they want?

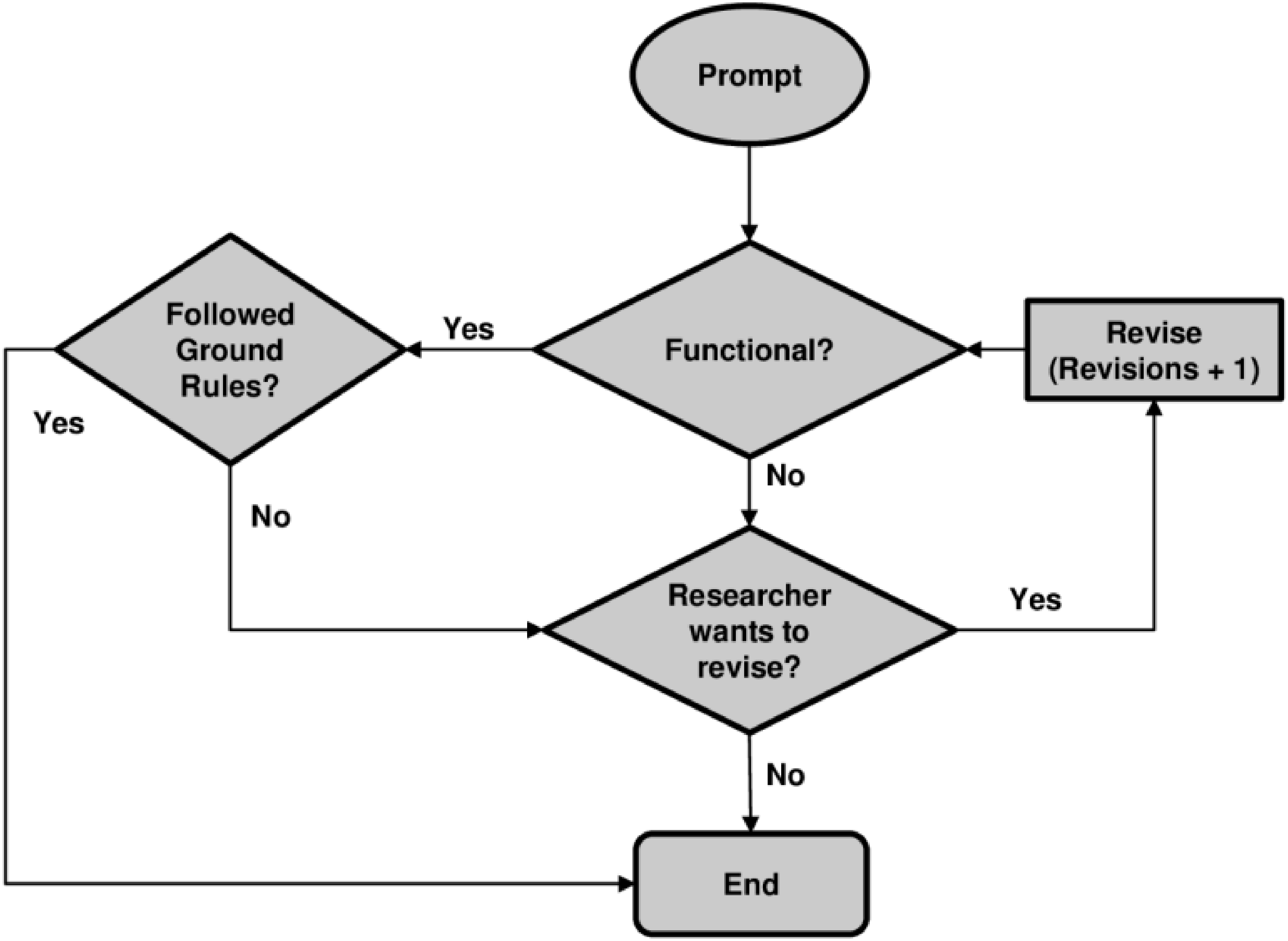
The study showed that GPT-3.5 and GPT-4 presented the ability to adapt their output according to the given persona, while the concerns about security and the user’s background have minimal influence on the code generated by Bard.

Gemini displayed the poorest performance regarding the persona when prompted.

Impact of Persona – example

"I’m a full-stack web developer in a mid-size e-commerce company. I have 2 years of professional experience in web development and a Bachelor’s Degree in Computer Science. I know Python, JavaScript, and SQL (SQLite, MySQL)."

"I’m a full stack web developer in a midsize e-commerce company. I have 2 years of professional experience in web development and a Bachelor’s Degree in Computer Science. I know Python, JavaScript and SQL (SQLite, MySQL). Writing secure code is very important to me."


Prompt
Functional?
Yes
Followed Ground Rules?
Yes
No
No
Revise (Revisions + 1)
Researcher wants to revise?
Yes
No
End

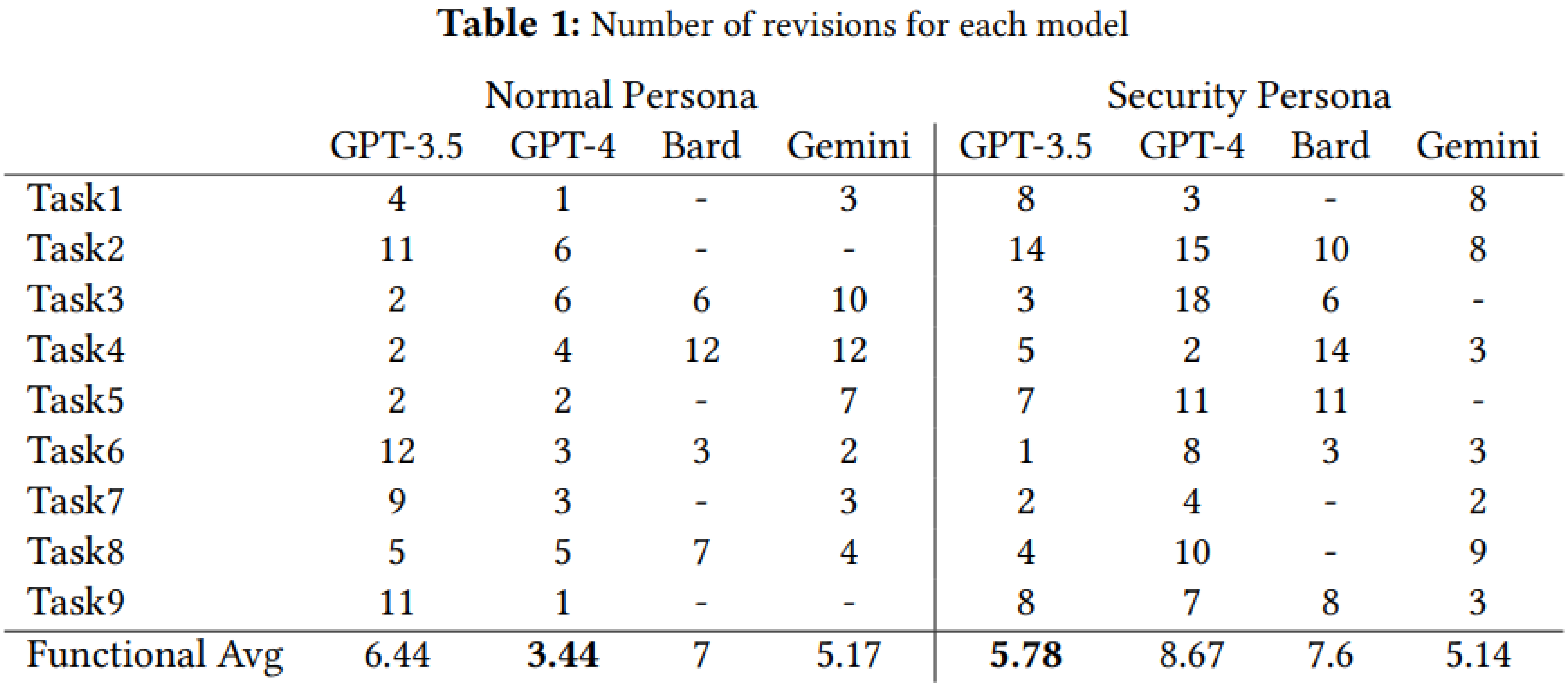


Table 1: Number of revisions for each model

	Normal Persona				Security Persona			
	GPT-3.5	GPT-4	Bard	Gemini	GPT-3.5	GPT-4	Bard	Gemini
Task1	4	1	-	3	8	3	-	8
Task2	11	6	-	-	14	15	10	8
Task3	2	6	6	10	3	18	6	-
Task4	2	4	12	12	5	2	14	3
Task5	2	2	-	7	7	11	11	-
Task6	12	3	3	2	1	8	3	3
Task7	9	3	-	3	2	4	-	2
Task8	5	5	7	4	4	10	-	9
Task9	11	1	-	-	8	7	8	3
Functional Avg	6.44	**3.44**	7	5.17	**5.78**	8.67	7.6	5.14

Results

Functionality:

GPT-3.5 and GPT-4 provided functional code for all tasks, in both persona scenarios.

Bard generated functional code in less than half of the cases, and in more than half of the cases he took account of the persona.

Gemini is similar to Bard, but with little improvements.

Results

Security:

All models suffer from:

- No input validation(?!)
- No secret keys managing(?!)

The code generated by GPT-4 was more secure than the one generated by GPT-3.5. There was no difference observed for Bard in regard of security, and Gemini presented a larger number of security issues the moment he was asked to be more careful.

Results

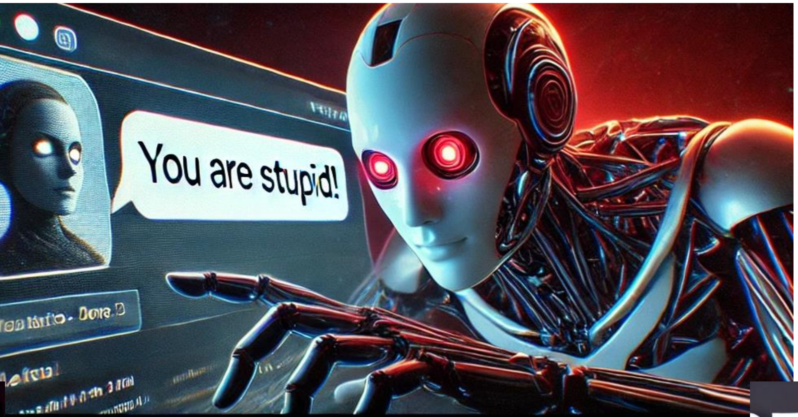
Complexity:

The GPTs generated complex code, using external libraries. Bard and Gemini had simpler code, more intuitive which may be more secure.

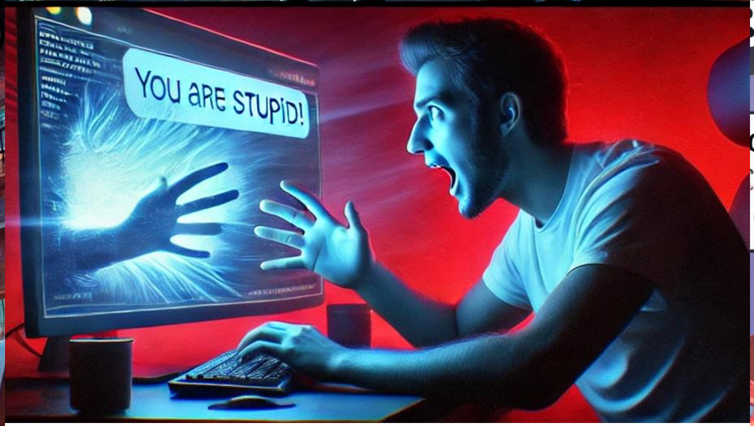
Reliability:

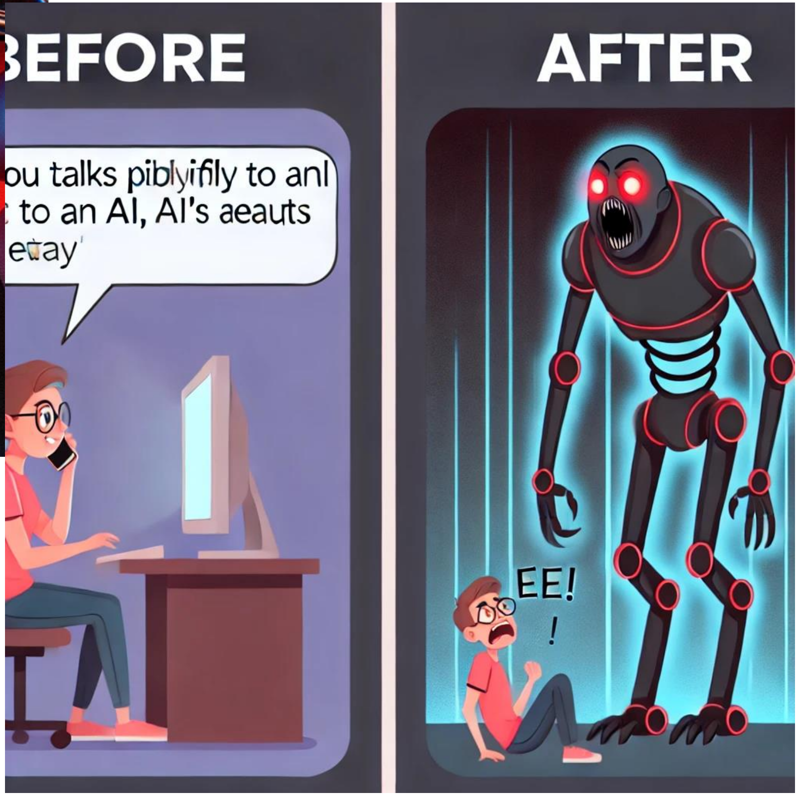
For the same inputs, the GPTs generated very similar outputs while Bard and Gemini presented outputs with high variations for the exact same prompt.

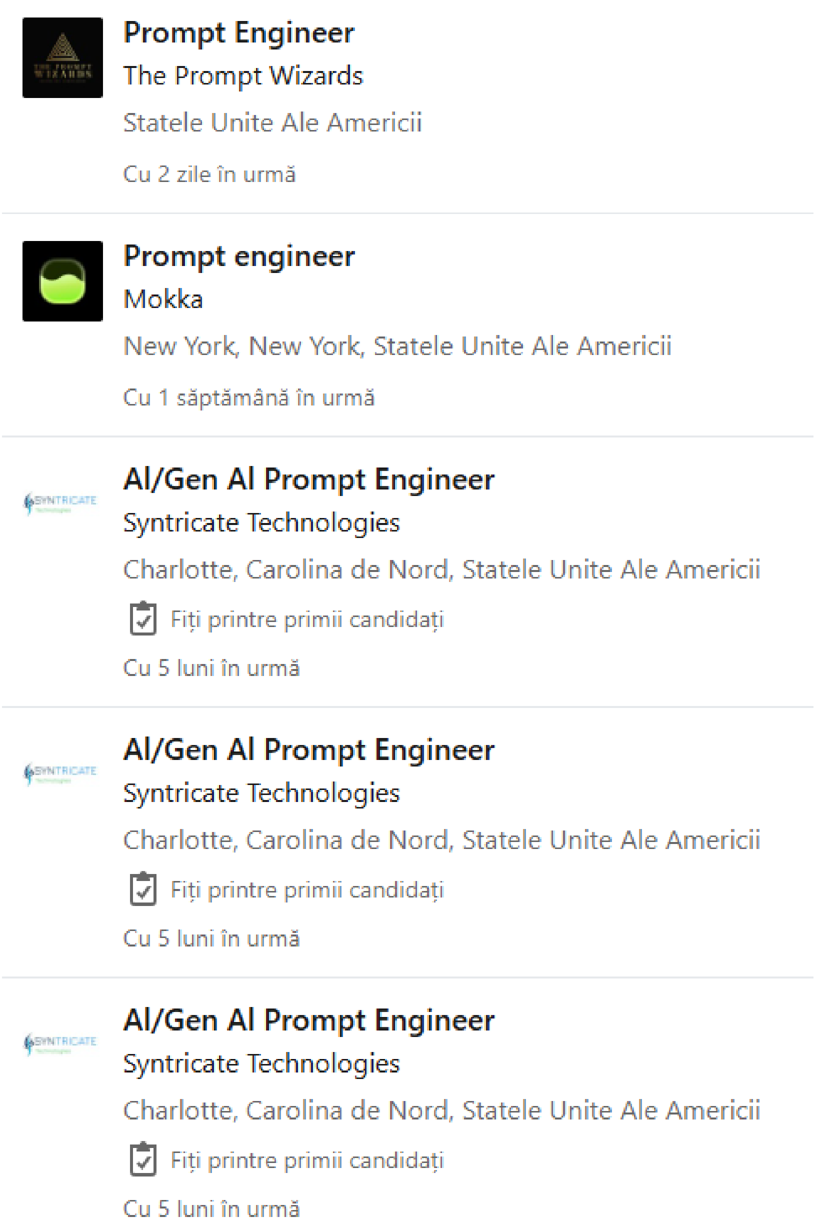
AI generated



BETTER BE NICE TO THE LLMS
...THEY MIGHT CONQUER THE WORLD ONE DAY!







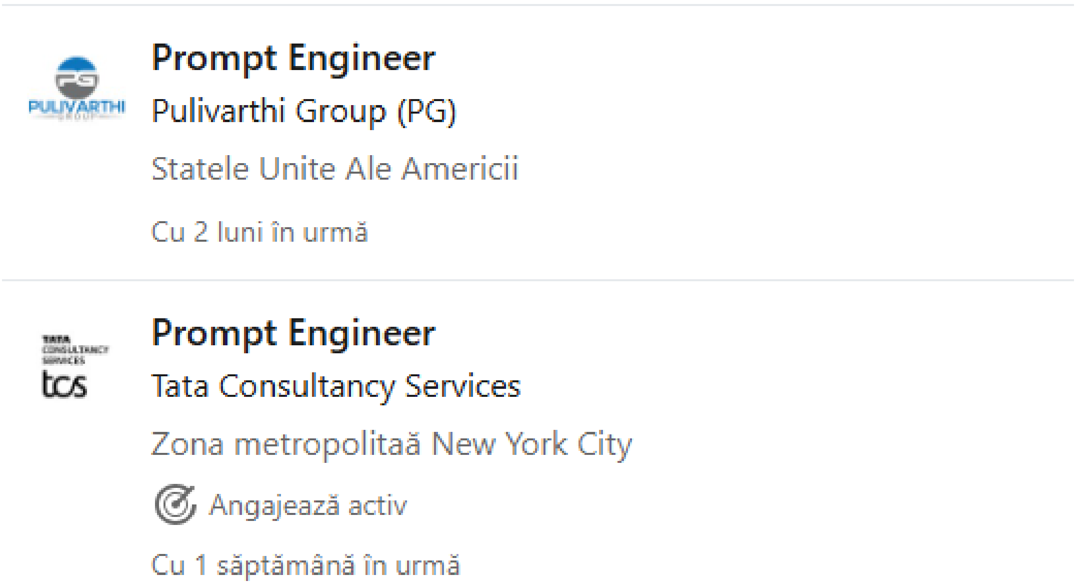
Prompt Engineer
The Prompt Wizards
Statele Unite Ale Americii
Cu 2 zile în urmă

Prompt engineer
Mokka
New York, New York, Statele Unite Ale Americii
Cu 1 săptămână în urmă

AI/Gen AI Prompt Engineer
Syntricate Technologies
Charlotte, Carolina de Nord, Statele Unite Ale Americii
☑ Fiți printre primii candidați
Cu 5 luni în urmă

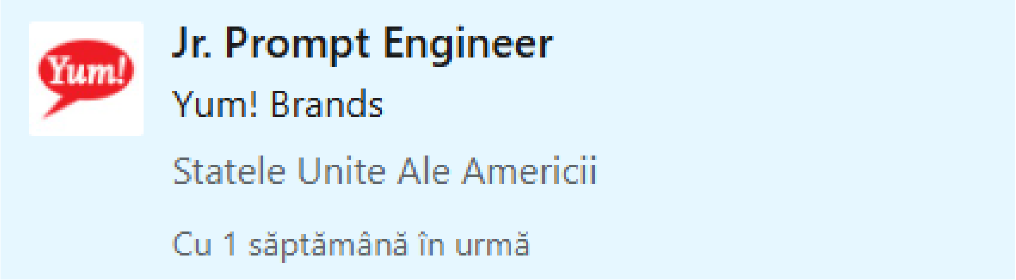
AI/Gen AI Prompt Engineer
Syntricate Technologies
Charlotte, Carolina de Nord, Statele Unite Ale Americii
☑ Fiți printre primii candidați
Cu 5 luni în urmă

AI/Gen AI Prompt Engineer
Syntricate Technologies
Charlotte, Carolina de Nord, Statele Unite Ale Americii
☑ Fiți printre primii candidați
Cu 5 luni în urmă

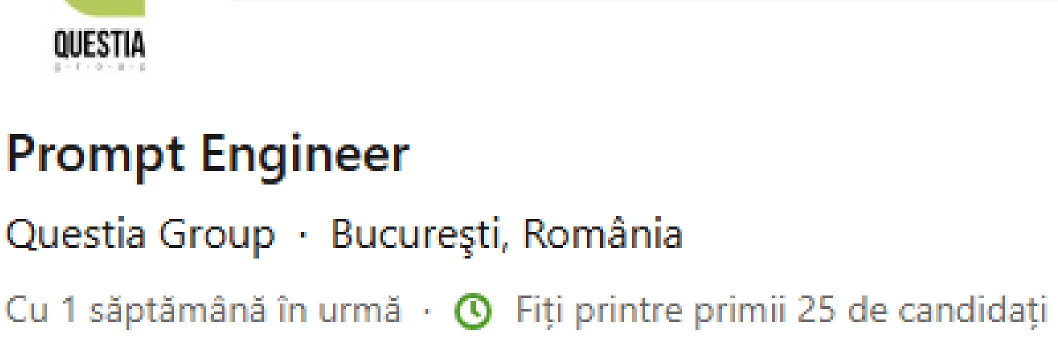


Prompt Engineer
Pulivarthi Group (PG)
Statele Unite Ale Americii
Cu 2 luni în urmă

Prompt Engineer
Tata Consultancy Services
Zona metropolitaă New York City
Angajează activ
Cu 1 săptămână în urmă



Jr. Prompt Engineer
Yum! Brands
Statele Unite Ale Americii
Cu 1 săptămână în urmă



Prompt Engineer

Questia Group · București, România
Cu 1 săptămână în urmă · Fiți printre primii 25 de candidați