majacinka /
crewai-experiments

<> **Code** · Issues 4 · Pull requests 1 · ▷ Actions · Projects · Security · Insights

Experiments with local as well as models available through an api

☆ **422** stars · ⑂ **94** forks · ⊙ **16** watching · ⑂ **1** Branch · ⬡ **0** Tags · ∿ Activity

⊕ Public repository

⑂ main ▾ · ⑂ **1** Branch · ⬡ **0** Tags · ⑂ · ⬡ · 🔍 Go to file · t · Go to file · + · Add file ▾

	🖼 **majacinka** last month	··· ⟳
📄 README.md	examples added and readme i…	last month
📄 google_newsletter.py	improved prompt for consistency	last month
📄 reddit_newsletter.py	improved prompt for consistency	last month
📄 startup_crew.py	examples added and readme i…	last month

📖 **README**   ✏ ☰

CREWAI experiments

For my experiments with CrewAI, I decided to try 3 different projects, starting from the easiest to the most complex. The aim of the experiments was to have a team of AI agents do following work for me:

1. Examine my Startup idea
2. Build AI newsletter with Google SERP
3. Build AI newsletter with Reddit Scraper
4. Email classifier [WIP]

For my experiements I've tried following LLMs:

Available through API calls:

1. OpenAI -> GPT-4
2. Gemini Pro

Local Models through Ollama + rating of how they performed:

1. **Mistral 7B**

- Nice, coherent results
- Didn't understand that it should use scraping tool for the output
- Result is a bunch of generic text from training data

2. **Mistral 7B instruct**

- Nice, coherent results, a lot of emojis
- Didn't use any scraping tool for the output
- Result is a bunch of generic text from training data.

3. **Open Chat 3.5 7B**

- The best and most "newsletter-y" results
- But again, didn't use any tool, so generic content

4. **Nous Hermes 7B**

- Ok results
- didn't use any tool
- generic content

5. **Open Hermes 2.5 7B**

- The tone and style of writing is great
- but generic content
- didn't understand that it needs to use tools

6. **Starling 7B**

- Ok results
- didn't use any tool
- generic content

7. **Llama 2 13B**

- The only model that "understood" what the task is
- but the text wasn't coherent enough, didn's sound like a newsletter

8. **Llama 2 13B chat**

- Didn't understand the task or produce any output

9. **Llama 2 13B text**

- Didn't understand the task or produce any output

10. **Llama 2 7B**

- not coherent
- didn't use any tool
- no output

11. **Llama 2 7B text**

- No actual output
- didn't use any tool
- generic content

12. **Llama 2 7B chat**

- Didn't use any tool
- generic content

13. **Phi-2**

- The smallest model ran into biggest problems
- Lost track of what it's suppose to do, no output

Releases

No releases published

Packages

No packages published

Languages

- **Python** 100.0%