Project 97: Reverse Prompt Engineering (Guess What Prompt Made This)

Description:

This tool takes a model-generated response and tries to guess what prompt could have produced it. It helps you understand how prompts guide output—perfect for learning prompt design through reverse engineering.

reverse_prompt_engineering.py

```python
import os
import openai
import gradio as gr

# Load OpenAI key
openai.api_key = os.getenv("OPENAI_API_KEY")

# Function to guess the original prompt
def reverse_engineer_prompt(model_output):
    try:
        # Prompt the LLM to guess what led to the output
        messages = [
            {
                "role": "system",
                "content": (
                    "You are a reverse prompt engineer. Given a piece of model
                    "Try to infer the user's intent and reconstruct a suitable
                )
            },
            {
                "role": "user",
                "content": f"Here is the AI-generated output:\n\n{model_output
            }
        ]

        # Call OpenAI
        response = openai.ChatCompletion.create(
```

```python
        model="gpt-3.5-turbo",
        messages=messages
    )

    return response["choices"][0]["message"]["content"].strip()

    except Exception as e:
        return f"Error: {str(e)}"

# Gradio UI
iface = gr.Interface(
    fn=reverse_engineer_prompt,
    inputs=gr.Textbox(label="Paste the AI-Generated Output"),
    outputs="text",
    title="🕵️ Reverse Prompt Engineering",
    description="Paste an AI response, and this app will guess what kind of pr
)

# Launch the reverse engineer
iface.launch()
```