Street Interview Video Creation Using AI Tools

Introduction

This document describes the process of creating a short street interview–style video set in Times Square, New York City, using AI tools. The goal was to generate a natural 20–25 second dialogue between a reporter and a tech job seeker, split into multiple segments due to tool limitations, and then merge them into one cohesive video.

To give context, the project started with the following script, which served as the base prompt:

Reporter: "Hi! We're talking to people here in Times Square about the tech job market. Are you job hunting right now?"

Interviewee: "Yeah, I just graduated and I'm applying for software engineering roles."

Reporter: "How's the market treating you?"

Interviewee: "It's tough—lots of competition—but networking helps more than online applications."

Reporter: "If your resume was on a Times Square billboard, what would the headline say?"

Interviewee: "Haha, probably: 'Code, Coffee, and Creativity!'"

Reporter: "Love it. Best of luck out there!"

How the video was created

1. **Prompt Setup**
    - The AI prompt was defined to generate a natural back-and-forth dialogue in a Times Square setting from Chatgpt.
    - The script above was used as the foundation.
2. **Video Generation in Google Veo3 (Flow Application)**
    - Since Google VO3 only supports **8-second clips**, the dialogue had to be broken down into smaller pieces.
    - Each segment (e.g., first reporter question + interviewee response) was generated as a separate video.
    - This step was repeated four times to cover the full script.

3. **Merging in CapCut**
    - The four separate clips were imported into **CapCut (Mac)**.
    - The clips were arranged in sequence and merged into a single continuous video.
    - The merged output simulated one street interview session.

Challenges Faced

1. **Prompt Interpretation Limitations**
    - Even though the prompt was clearly worded, the AI video tool did not always interpret the instructions accurately.
    - For better accuracy, scripts had to be downloaded separately from AI text generators (ChatGPT or Google AI) before being used in the Flow application.
2. **Role Consistency**
    - The intention was that the reporter would pass the microphone to the interviewee while asking questions.
    - However, the AI often reversed this action or displayed it inconsistently.
    - This lack of fixed process meant that some clips looked correct while others did not.
3. **Consistency Across Clips**
    - Because each segment was generated independently, maintaining a consistent look and behavior across all four clips required careful editing in CapCut.

Conclusion

Overall, the activity was both **fun and insightful**. While the AI tools are powerful for generating quick, creative visuals, there are still limitations in consistency, role accuracy, and prompt interpretation. Combining text-based AI for script creation and video-based AI for visuals, followed by manual editing in CapCut, provided the best results. The project demonstrated how AI can be used to simulate realistic, fast-paced street interviews and highlighted areas where current tools can improve.