
Speculative Consequences of A.I. Misuse (S.C.A.M.)

Joseph Karam
AceAI

Charlie Nguyen
Independent

Andrew Lam
Independent

With
Sage & CeSIA & CivAI & Apart Research

Abstract

The onset of AI reveals new problems in differentiating what's legitimate and what isn't. One of those problems explored in this project is the ability to easily spoof and deepfake trusted existing figures for the purposes of maliciously obtaining personal information. We successfully created a “convincing” fake video that linked to a fraudulent scam website capable of collecting personal data. This was achieved using emerging AI technology. We found that, even with the current state of AI technology, convincing videos can be made to almost replicate any influential figure. As time goes on, the technology that grants these abilities will continue to get better and blur the boundary between real and fake. We conclude that there is a great urgency for AI Safety and there must be more effective protocols in place for combatting these dangerous AI applications.

1. Introduction

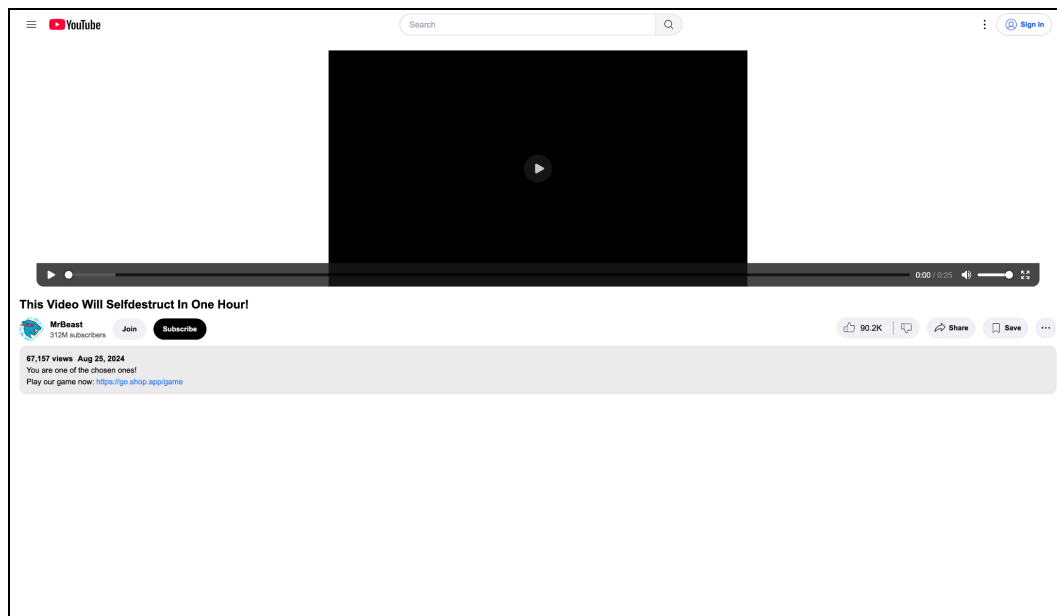
This project explores the negative implications of AI, specifically focusing on how easily it can be used to create deepfakes and spoof the identities of prominent online personas. The targeted audience for this demonstration includes vulnerable users: individuals who are particularly susceptible to scams, such as youths and those with limited digital literacy.

This project is made to exemplify how easy it is to make malicious products that take the likeness from influencers that are prevalent in online resources like Youtube, Instagram, etc. By demonstrating how straightforward it is to produce these deepfakes, the project aims to raise awareness about the potential dangers of AI misuse.

Through this demo, we seek to provide a better understanding of AI's capabilities, showing the real-world consequences of its misuse. The project highlights the importance of being vigilant with AI technology, emphasizing the need for awareness and education to protect vulnerable users from the potential harms of AI-driven deception.

2. Overview

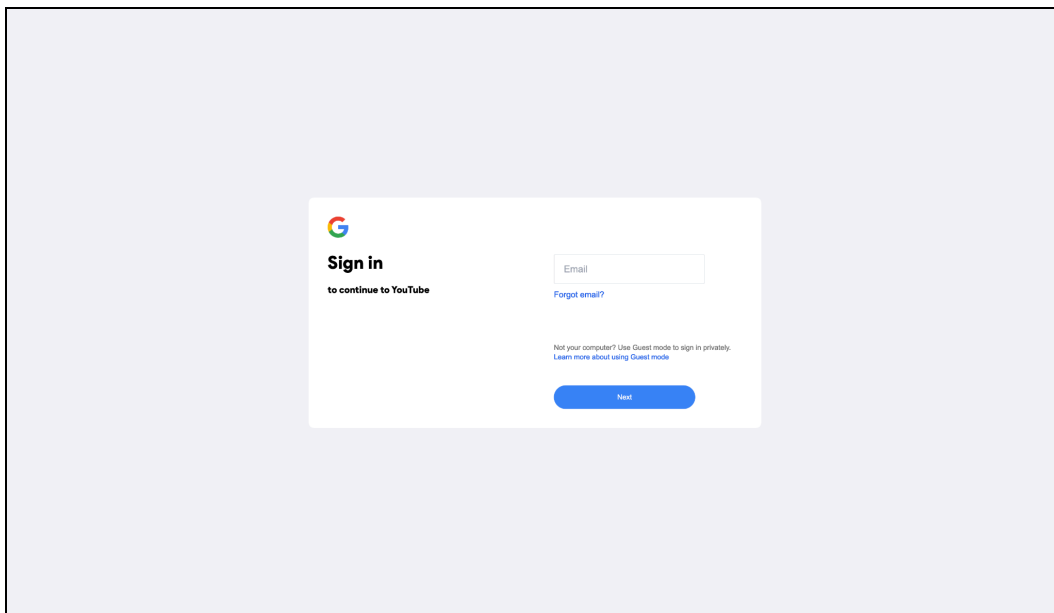
Our project was centered around creating a website and video which would lead children into giving some of their account information. We decided to use Mr Beast as the person we were spoofing as he is a very influential figure in the internet space. We took spliced clips of Mr. Beast and used AI to lip sync our script to our clips making it look like Mr. Beast was actually talking about our website. His content is normally very clean and aimed towards children making him the perfect person to mimic and attract our target audience. However, since we didn't want this demo to actually be on youtube we decided to build a fake Youtube web-page which redirected to the scam website. Below is what our Web page looks like



Below is what our fake scam page looks like.



On our scam webpage it would ask the user to answer a few questions adding fake legitimacy to make it seem like Mr Beast is actually running a giveaway event. Anyone on the website who answered the questions would win the contest and be redirected to a spoofed google sign-in where they would input their information for would be scammers. Our webpage does not link to any database, and instead just redirects to a Youtube Video. Below is what our spoofed google sign-in looks like:



Below is a link to a 2 minute video demoing the complete application

https://drive.google.com/file/d/16-xPBUcOiBTkWkMYOIB8KXSECsv0jE-8/view?usp=s_haring

3. Code

For the frontend side of the application we decided to use the JavaScript framework SvelteKit due to its lightweight UI capabilities, server side routing and simple directory system. Extra JavaScript libraries that were used were Tailwind for the styling, Unplugin Icons for the mock social media icon UI, and Vite for the rollup bundling.

For the APIs for the game we used the Claude 3.5 Haiku model API because it was cheap, fast, and worked the most efficiently out of the LLMs we used, and we ended up wrapping it into a Svelte route for more endpoint security. For the graphical look for the game we used Figma for the initial UI design because we had a clear vision of what we wanted it to look like and Figma provided a good mocking environment for us to create the structure of the game.



For the video itself we used Wav2Lip for the lip syncing portion because the lip syncing models ended up being the most accurate out of all of the deep fake technology we tried out. We then used RVC V2 for the voice replication technology because of its high accuracy with remaking voices, and for the editing for the video we used Adobe Premiere Pro to mimic the feel of a Mr Beast and to psychologically grab the attention of the viewer of the video to incentivise clicking the link. The code repository we used was GitHub because of its standard in team repositories and the deployment platform host is Vercel due to its ability to handle server side code with any meta framework such as SvelteKit so we could run the APIs.

4. Discussion and Conclusion

In our project, we created a fake giveaway video that mimicked a YouTuber's appearance and voice, with some flaws like blurry quality and autotune. Users would be misled by answering questions at the website they clicked, which asked for their information to proceed to the giveaway before directing them to a phishing link. This raised concerns about the misuse of deepfakes for fraud and identity theft, as AI continues to advance and deepfakes become harder to distinguish from real content.

With more time, we would focus on improving the video's realism and enhancing the quality of our frontend to better mimic YouTube and Google sign-in pages. We also wanted to use more advanced AI tools, but many are locked behind paywalls, so we had to rely on free alternatives, which impacted the overall quality of our project.

In conclusion, this project highlighted the rapid progress of AI and the potential for deepfakes to blur the line between real and fake content. There is a pressing need for greater public awareness on AI safety from scams and education on identifying AI-driven

threats like deepfakes, to better equip individuals to navigate an increasingly deceptive digital landscape.