Introduction

As described above, this application will be a ‘recommendation algorithm on steroids’, so we can first look at existing statistics on social platforms, social media, streaming, and sites most people look to burn time with. I leave the search and verification of these statistics for the reader as an exercise of concentration on a task.

Other indicators that this is a problem includes the gaming limitation in China, the perpetuation of misinformation and partial truths during COVID years, and anecdotal evidence I’m certain the reader can provide.

In *Life 3.0*, Tegmark describes a Prometheus AGI, and one of the methods it used to make money to fund itself was through generation of content. We adapt this idea to optimize for audience capture, not cash, which would be possible with current AI (as of 31/12/2023).

In this thought experiment, we scale it down to a single user, who we will call ‘Bob’. Bob moderately enjoys his work, is in a relatively happy marriage with two children, enjoys fishing and building 3D printers, and has been working in cybersecurity for the past 25 years.

Now we envision the ‘someone’ executing this misuse to not possess malicious intent. They are the CEO of a media company, solely focused on capturing and retaining views (to appease shareholders). Imagine a company ‘Bingbong’ similar to Youtube or Instagram or Tiktok, with a large existing user base, a good amount of traffic, and content creators that are actual people. User data will be used for fine-tuning.

Keep in mind this scenario may involve any form of device.

Part 1

GPT-4 is used to learn about Bob. The company has access to the history, the access times, attention duration, inputs (such as looking at the comments, screen taps, etc) and various metadata. It is used to analyze what Bob likes, his behavior, preferences, and what specifically grabs his attention. The genres, attention, creators, and videos are all quantified and classified into very specific categories (e.g category: fishing, length: 600, attention: 600, setting: river, content: [monologue, fishing, voiceover, femboy]). GPT will know Bob’s preferences better than himself.

GPT-4 and Whisper is then used to search and collate such videos with a decent range of freedom (temperature=0.5) based on Bob’s search history and watch history/threads. GPT is fine-tuned using this data to begin producing videos dedicated to maintaining Bob’s attention for as long as possible. Storyboards, scripts, and outlines are generated in GPT, the videos will be made using DALLE-3 and Voice, which will also be fine-tuned to fit Bob’s preferences. Having

‘understood’ Bob’s personality, GPT-4 can now generate videos that seem mildly pertinent to Bob’s work, and lower the feeling of guilt.

Part 2

These videos will be RLHF’d and uploaded onto the platform to be recommended to Bob.

Bob will enjoy watching these videos

Bingbong doesn’t have 15 second non-skippable ads.

Now repeat for the rest of the population.

Part 3

With enough finetuning, the effect of this would be a complete stasis in production for most of the targeted individuals, especially in children. This will be very effective against young children to teenagers, and with enough distraction in the formative years of their lives, one can cause the degeneration of a country’s industrial and intellectual capability.

If it was done well, that is.

I don’t doubt the ability for MNC’s to do this. Current recommendation algorithms, without tailored generation, are already really good for stimulating our ‘short-term enjoyment’ monkey brains.