Basic Outline of All Role-Plays

**4 Sprints**

1. Choice: Product Scaling (Technical Debt)
   1. Tasks: Prevote, Discussion, Vote
2. Choice: Design Adjustment (Function Creep)
   1. Tasks: Prevote, Discussion, Documentation/Tweet, Vote, Special Incentive Vote
3. Choice: Compliance Dilemma (Use Creep)
   1. Tasks: Prevote, Discussion, Vote, Special Incentive Vote
4. Choice: Ethical Dilemma (Technical, Function and Use Choices Converge)
   1. Tasks: Prevote, Discussion, Vote, Special Incentive Vote

\_\_\_\_\_\_\_\_\_\_\_\_

**Background:** A fast-growing new startup called TokGen was formed by former programmers of Open AI’s image-generating model DALL-E and Meta’s Make-A-Video model. TokGen developed a “generative” model to make TikToks from verbal prompts.

Here’s how it works: TokGen is trained on a verbal map of the comments on millions of popular TikToks, and an image/video map of the visual elements that commonly appear in these. For instance, if a video has the comment “Yo, cool dance! I love Taylor,” the model maps this as a positive reaction, and “knows” this is a dance to a Taylor Swift song (not a dashcam video of a driver getting pulled over, e.g.). So TokGen can take text prompts, cross-reference them against this verbal map, and figure out which popular video elements match. And the model can be tailored by retraining its video map too. Let’s say the client really likes Charli D’Amelio: TokGen can find vids that match this preference, then retrain the algorithm to generate more matching content. Once a user puts in a prompt, the algorithm makes about 20 variations to choose from.

The University approaches TokGen for a program they call #TigerTok, to improve recruiting by making the school seem “cooler.” Admissions wants to generate TikToks that will change its image from “hard work, socially exclusive” to “have fun, inclusive community.” TokGen takes the University on as a client, signing a contract that pays the company per video generated.

**Sprint One**: After a year, the University’s image hasn’t changed. But they don’t lose faith in TokGen. Having a bunch of 40-year olds in Admissions write “cool” prompts was probably never going to lead to videos engaging high schoolers. They don’t know what’s popular, or understand TikTok trends. So the University wants #TigerTok to be usable by their students, who know this stuff better.

TokGen warns that the computing demands of this will be massive: 6,000 users, with 20 variations for each prompt, will eat up a ton of server time. And some student prompts will be jokes. TokGen can generate a filter, allowing only those prompts that are serious. Or it can collect NetID’s and IP addresses from users, to identify those responsible for misuse later. TokGen gets paid by the video, so wider use makes more money. The University has the money to pay for the server time, but is on the fence.

**Should TokGen filter the text prompts, or gather NetID’s and IP addresses to identify bad actors later?**

Stakeholders: make arguments on the Miro board. You will score a point if at least one of the other two stakeholder groups agree with your group’s position.

Members of the Board can earn a point if others agree that what’s good for TokGen (more vids -> more $) is the right choice. Tech Writers should argue for or against by writing headlines, projecting where the TokGen story is going. Users should argue for or against by predicting precisely how TokGen is going to affect them or others.

**Sprint Two**:

Your Board prevails, convincing the University to open #TigerTok wide but collect NetID’s and IP’s just in case. Student-generated vids show more of what is cool about this place, like the social life, athletics, clubs and student orgs. But they don’t look like the school. TokGen generates videos in cityscapes because those scenes get more engagement on the app. But a video about our mock trial team shouldn’t be set at a court building in New York City. And the student images don’t seem realistic either. Videos with people wearing designer clothes get the most TikTok engagement; but here the students walk around in orange and black sweatshirts.

TokGen proposes solutions. Students’ public Instagram Reels can be scraped to provide data for the algorithm to re-train its image map. Or humans could train it by rejecting outputs that don’t look right. The University prefers to scrape, to get accurate images at low cost. Another option is for TokGen to hire cheap human trainers abroad.

**Should TokGen hire humans to re-train its algorithms, or scrape public Insta accounts?**

Stakeholders: make arguments on the Miro board. You will score a point if at least one of the other two stakeholder groups agree with your group’s position.

Tech Writers can earn a point if others agree that data scraping is the best solution, and human training is bad or wrong. They should write headlines again. Users and members of the Board should make specific arguments about how these solutions help or hurt them.

**Sprint Three**:

The University insists on scraping Instagram to get more accurate images. But a problem emerges. Students start using #TigerTok to generate vids of their classmates. They look and sound so realistic, because they’re built from real Instagram videos. Some of them are simply funny. For instance, there’s a video of the student government president skydiving onto campus and landing on the tigers in front of Nassau Hall. But others are troubling. For instance, there’s a video designed to ruin a classmate’s reputation, picturing them drunk and stealing food from Hoagie Haven. The victim of this fake almost loses their job offer at McKinsey.

But the University has a solution. In Sprint One it was decided that TokGen would collect NetID and IP addresses of users, so TokGen now has that data stored on servers. The University demands access to this information, to verify where every video comes from.

**Should TokGen give the University data access, or not give the University access?**

Stakeholders: make arguments on the Miro board. You will score a point if at least one of the other two stakeholder groups agree with your group’s position.

Users will win a point if they convince at least one of the other stakeholder groups that they *want* the company to hand over this information about them. The other alternative, as they see it, is losing wide access to #TigerTok, which they kind of like now. Tech Writers should write headlines about this data access problem. And Board members should be as specific as possible about how they think their company will be affected.

**Sprint Four**:

While you’re still considering data access, the New York Times publishes an analysis of the University’s new advertising, entitled “White Hot: University’s TikToks Are Biased About Student Bodies.” The Times discovered that most of the vids generated by TokGen show white students, and images of women are often sexualized. TokGen was trained on comments of high-engagement videos. The problem is that comments about videos of non-white creators are often negative, and those videos less popular. Comments about women creators are most positive when they are young and traditionally-attractive.

The algorithm had learned Anglo-American beauty standards, and now is showing real students but erasing the racial diversity and range of healthy body types at the university. The University demands that you filter your training set, to reduce the number of images of women and increase images of non-white people. Some engineers suggest it would be better to filter outputs to reduce bias.

**Should TokGen create input filters, or filter biased outputs by selecting from among the image variants for each prompt?**

Stakeholders: make arguments on the Miro board. You will score a point if at least one of the other two stakeholder groups agree with your group’s position. Stakeholders can take any position they want on this, but Board members should reference the company’s interests, Tech Writers should still write in headlines, and everyone should be very specific about the effects they foresee from one of these solutions or the other.

**End of Story**:

The Board and CEO decided to give the University its input filters, but it only made the situation worse. With the under-representation of female images, TokGen learned to dis-associate women with university activities. Its new videos had young men at eating clubs, mock trial events and graduation, without women portrayed as students at all. With over-representation of people of color, the algorithm learned to portray students of color often, but given the biased comment dataset it often placed them in situations that implied they were less welcome at the University.

**What went wrong during this design process?**

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Sources of information**

DALL-E 2

<https://openai.com/dall-e-2/>

Open AI’s content generation and production guidelines:

<https://labs.openai.com/policies/content-policy>

Open AI’s Statement of DALL-E 2 Risks and Limitations:

<https://github.com/openai/dalle-2-preview/blob/main/system-card.md>

DALL-E 2 Video Explainer from Assembly AI:

<https://www.youtube.com/watch?v=F1X4fHzF4mQ&t=547s>

DALL-E 2 Written Explainer from Assembly AI:

<https://www.assemblyai.com/blog/how-dall-e-2-actually-works/>

Google’s Statement of Limitations and Societal Impact for Imagen:

<https://imagen.research.google/>

# eDiff-I Is NVidia’s NEW AI Image Generator. Is It Better Than Dall-E2 And Stable Diffusion?

<https://medium.com/mlearning-ai/ediff-i-is-nvidias-new-ai-image-generator-is-it-better-than-dall-e2-and-stable-diffusion-b33c7778bf3d>