Kenyon College

Fall 2019

Can GPT-2 Replace a Sex and the City Writers' Room?

Alasia Destine-DeFreece
Kenyon College, destinedefreece1@kenyon.edu

Samara Handelsman
Kenyon College, handelsman1@kenyon.edu

Talia Light Rake
Kenyon College, lightrake1@kenyon.edu

Ally Merkel
Kenyon College, merkel1@kenyon.edu

Gracie Moses
Kenyon College, moses1@kenyon.edu

Recommended Citation

Can GPT-2 Replace a *Sex and the City* Writers' Room?

Alasia Destine-DeFreece, Samara Handelsman, Talia Light Rake, Ally Merkel, & Gracie Moses
Kenyon College IPHS Department
Fall 2019 Artificial Intelligence for the Humanities

Abstract

Someday, in the not-so-distant future, will the writers' room be replaced by artificial intelligence? This poster was created by a group of actors and writers who were curious about the role artificial intelligence could play in the future of television writing. In the first stage of this project, we prompted GPT-2 with the opening lines of a scene from the television show *Sex and the City*, and had GPT-2 finish the scene. We filmed the three best GPT-2 generated scenes, and screened these scenes for our "AI for the Humanities" class, along with a filmed version of the original *Sex and the City* scene. We hypothesized that the GPT-2 generated script would be able to convincingly mimic the simple back-and-forth dialogue of *Sex and the City*. Unfortunately, the students in this class easily distinguished between the GPT-2 generated scenes and the human written scene. After reaching the conclusion that GPT-2 is not able to write a believable *Sex and the City* script, we generated another series of scenes, this time using a GPT-2 model that had been trained on every episode of *Sex and the City*. We found that the trained GPT-2 was able to mimic the style of *Sex and the City* better than the untrained model, but even the trained model struggled with creating cohesive scenes. Based on the results of this experiment, it can be concluded that GPT-2 is capable of mimicking the style of *Sex and the City*, but is not yet capable of writing a scene that an audience would believe to be written by a human.

Methodology

Generative Pretrained Transformer 2 (GPT-2) is a large scale, Artificial Intelligence model that scrapes from the Internet to generate text based on input given. The model is able to effectively respond to the type of language put in whether it is dialogue, essay-writing, or scientific computing, the model creates text to finish the prompt appropriately. This large-scale transformer based model has 1.5 billion parameters, and is trained on 8 million webpages (Radford, 2019). In addition to responding or "finishing" text based on prompts given, GPT-2 can translate languages, answer questions, and learns from the "raw" text given and is therefore, unsupervised in this response (Radford, 2019). Hence, GPT-2 is groundbreaking in its ability to achieve these literary standards without task-specific training.

As a group of humanities students, we decided to test the comprehensive capabilities of GPT-2 in a scene that we then filmed. Using, a "Jupyter" notebook, we played around with GPT-2, and tested the language modeling abilities. We prompted the model with a five lines from a scene (Season 1, Episode 1) of dialogue in *Sex and the City*, with the hopes that the model would finish the scene, in a succinct and sensible manner. We chose a simple five-line dialogue between two characters in the show, Kurt, and Carrie. We chose *Sex and the City* because it follows a pattern of romantic comedy that is easy to follow, and the scripts are readily available online. The model can create a multiplicity of outcomes from the five lines given, but it was not always comprehensible. It was nevertheless impressive that each time the script was completed in a different manner, wherein there was potential for a story-line every time. With that said, we decided to run the model eight times to then chose three of the "best" or most human-like scripts. Each time we ran the script it yielded three sample texts, so we ultimately chose from 24 scripts GPT-2 created.

The next step was filming the four scenes we chose. Considering most of us have a drama or film background, we applied our artistic capabilities, but relied on the script to tell the story. Out of the four scenes, three were finished by GPT-2 , and one was the original *Sex and the City* scene. After several days of filming, we asked a class of students to watch the scenes and fill out a questionnaire corresponding with each scene. We played each scene, one after another and let individuals fill out a Google Form which asked whether the scene was robot-written or human-written and lastly why they thought so. After unfulfilling results, we went back and with the help of Professor Chun, trained GPT-2 on every season of *Sex and the City*, to compare these new scenes to the previous GPT-2 scenes that were solely prompted but not trained. As far as training GPT-2, we were able to use a "kaggle" notebook that had data from all six seasons of *Sex and the City*. We went back and analyzed GPT-2's trained scenes for comparison sake, and to pose questions for the future of GPT-2

Results & Analysis

As hypothesized, it was very clear that GPT-2 struggled to capture the witty back-and-forth of the original dialogue and was unable to create a convincing script without human intervention.

Of the three scenes, Scene 1 was predicted to be the most obviously completed by artificial intelligence. Directly following the end of the prompt, the dialogue lost all sense of logic and devolved into a fight without any tangible substance. The survey participants unanimously responded that GPT-2 completed the script, pointing out the inconsistencies in plot as well as various grammatical errors (i.e. "No I aren't"). Though the majority of the dialogue was grammatically correct, the previous example being an anomaly, the conversation as a whole was semantically nonsensical. Likewise, 100% of the participants correctly identified Scene 4 as being completed by GPT-2. Many responses pointed out semantic and pragmatic errors in specific lines, the majority of which isolated "I will with you" and "I will always be there to be there."

83% of the participants correctly identified Scene 2 as being the original script. While the goal of the experiment was to determine whether or not GPT-2 could pass for human, the responses that explain why the original script could be mistaken for AI are very telling. One of the two participants who selected GPT-2 wrote that the dialogue seemed too jumpy while the other pointed to how stereotypical the dialogue was. This points to two expectations that people have for AI-completed text: 1) it's unlikely that artificial intelligence would be able to write dialogue that follows natural human phrasing, and 2) artificial intelligence is expected to pull the most obvious themes from the text, thus creating a "stereotypical" script.

All but one participant correctly identified Scene 3 as being completed by GPT-2. While many noted that it was more convincing, the conversation was still too disjointed and switched between topics too often. Additionally, the script switched which character wrote the column, losing its continuity.

Results of in-class Survey
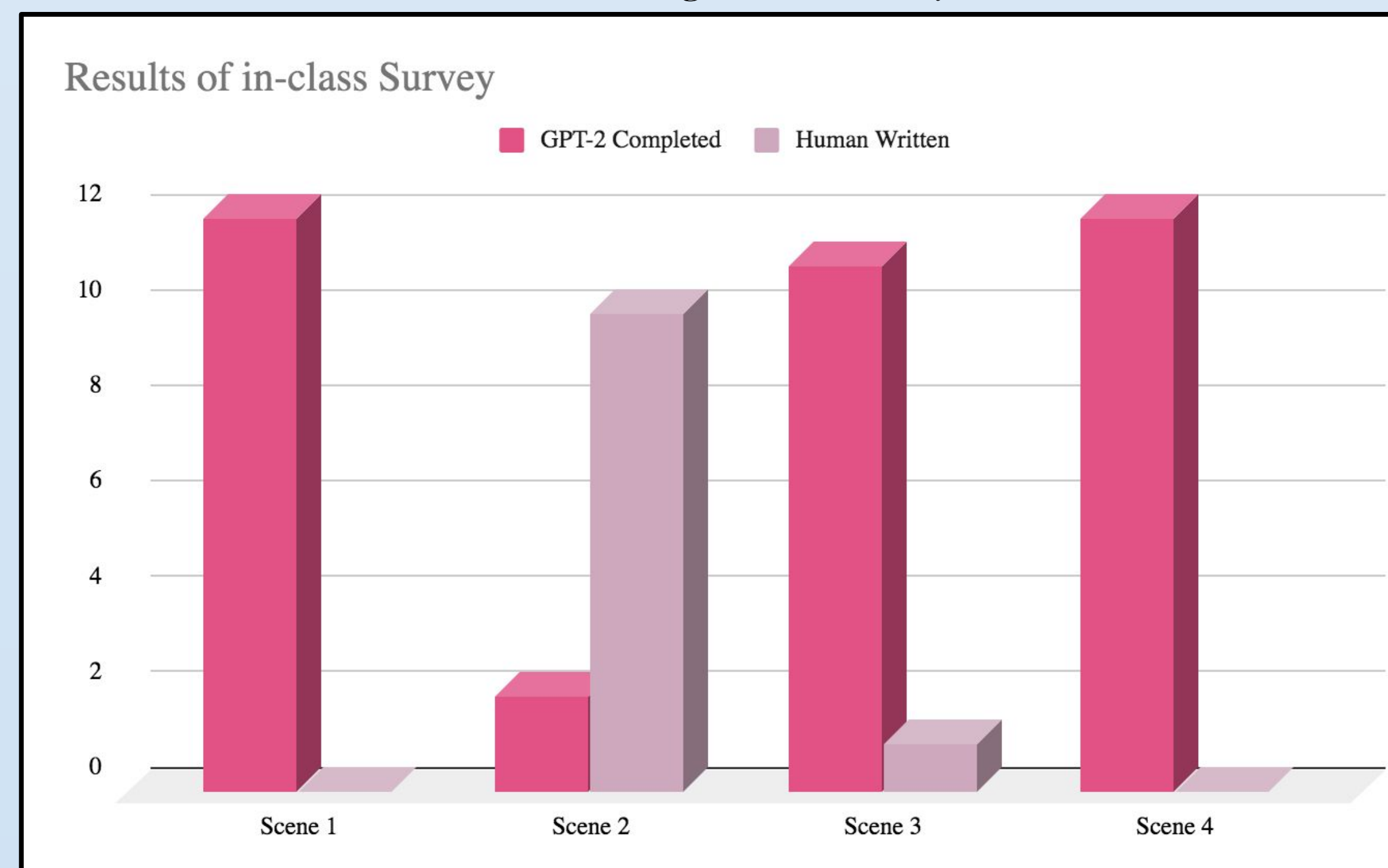
Conclusion

When working with the Generative Pretrained Transformer 2 (GPT-2), and artificial intelligence in general, it can be hard to keep up because the field is growing so rapidly and new versions are constantly coming out. Our goal when working with the untrained GPT-2 was to see if it could successfully replicate *Sex and the City*-like dialogue. While we were saddened to see that the answer was clearly negative, we believe that our research still brings forth new understandings for the future of natural language processing.

Before starting our project, we believed that a machine would easily be able to replicate something so simple and straightforward in themes and dialogue. Throughout our process, we learned that this lack of substance actually posed a significant roadblock to creating coherent conversation. We assumed that *Sex and the City* would be easier for GPT-2 to replicate than text from a Chekhov play because Chekhov has such a specific style. However, we were exposed to a peer's project that used Chekhovian style, which was easier to reproduce. It proved easier for GPT-2 to replicate Chekhov because it picked up on themes and speech patterns that were present in monologues. Even when we explored the GPT-2 that was specifically trained on *Sex and the City* dialogue, we saw that it still had trouble keeping the conversation coherent. It was able to pick up on their most common conversation topics, but it does not have a natural, conversational flow. For example, the trained model produced the following:

Trained Script

In the scripts that GPT-2 was trained on, if the character had a multi-line section of dialogue, the script wrote their name multiple times-- that is why GPT-2 has written conversations where it looks like people are talking to themselves. As seen here, the conversation topics are relevant to the show, but the speech patterns are choppy and atypical. We have come to conclude that GPT-2 can not recreate dialogue from something with such little substance.

We believed that our original prompt would be solid enough because it had subtext, idioms, and called for wit, but we now believe that GPT-2 is unable to replicate such specific human mechanisms.

Examples of other scripts used:

Original Script (Scene 2)

GPT-2 Completed (Scene 1)

Resources

Radford A. Better Language Models and Their Implications. OpenAI. 2019 10 [accessed 2019 Dec 11].
    https://openai.com/blog/better-language-models/#fn1
Star D. Sex and the City Season 1 Episode 1. Sex and the City Transcripts. 2008 [accessed 2019 Dec 11].
    https://www.satctranscripts.com/2008/08/sex-and-city-season-1-episode-1.html#.Xe2VopNKjVo
Special thanks to Professor Chun and Professor Elkins!