# CM202 Final Project, Winter 2020

JOHN DIEZ*, TESSA EAGLE*, and WILSON MUI*, University of California, Santa Cruz

Twitter is not a very positive space. The gamer space within Twitter is often thought of as toxic, being a space for sexism and other social issues. We have created a positive gamer twitterbot by with the use of two machine learning tools. A neural network classifier was trained and then used to help create a positive gamer tweet dataset for us to use. With this dataset, we were able to train a text generator to generate positive gamer-style tweets. The results of our text generator was used with our twitterbot to create a somewhat convincing positive gamer twitter account.

Additional Key Words and Phrases: twitterbots, GPT-2, sentiment analysis, text generations

**ACM Reference Format:**
John Diez, Tessa Eagle, and Wilson Mui. 2020. CM202 Final Project, Winter 2020. 1, 1 (March 2020), 6 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

## 1 INTRODUCTION

As technological evolution has started to evolve even further, one major question is whether we can ever make technology and AI mimic human behavior. Much of our current technology and current improvements of the items have focused on pure rationality, the logic behind each important step. However, to mimic human nature, AI needs to make mistakes, understand social cognition, and one part of that is the understanding of human emotions. With these goals in mind, the specific field of Machine Learning had started to become more interdisciplinary, where instead of having only computer scientists and people from only the technical fields would work on it, they had started working together with people from Social Sciences and Humanities. Specifically, for understanding emotions, they started working with Psychologists, Cognitive Science, and Linguistics. This one important branch of innovation that spurred from the dawn of machine learning algorithms was the creation of Affective Computing, which is our ability to analyze emotions through ML programs.

Through recent history, we've been able to develop ways of creating programs and AI that have the capacity to learn from patterns thanks to pre-existing rule sets [1]. This same mechanic of logic, under the joint work of other fields, was able to develop and create new programs that had the capacity to learn language, and in turn, understand parts of the human psyche. Affective computing is defined as systems that can recognize and interpret/process human affect (emotions or mood), and then can categorize or make assumptions based on these processes [2]. Sentiment Analysis is a branch of it that specifically uses these algorithms with NLP to classify these emotion to different valences or categories like "positive, negative, neutral" [3]. Affective computing at the moment can be seen through two primary lenses, text computation and facial computation. In text analysis, the most improved algorithm has the ability

Authors' address: John Diez, jdiez@ucsc.edu; Tessa Eagle, teagle@ucsc.edu; Wilson Mui, wimui@ucsc.edu, University of California, Santa Cruz, 1156 High St, Santa Cruz, CA, 95064.

to read and understand text by learning from a hybrid of rules that include frequency of text (how many times does a word
" "appearnextto" "),thesemanticmeaning(howtodifferentiate"Ihadagreatday"vs"IhadaGREATday",andfinallypresetrulesandinformationgivenbyscientistsonhowwordstransla

While Affective computing has definitely improved over time with the development of twitter bots, text generators, and sentiment analysis projects through large texts, there are still large biases in the program that relate more towards socially defined problems of culture, slang, and differences in tone/context. One particular population that we will be focusing on in this project are video gamers, and specifically their way of speech online settings (twitter). We targeted specifically gamers because a lot of criticism has been towards the gaming community due to their high rates of toxicity and long history of gate-keeping [5]. It has gotten so much attention, that many of these gaming companies that host these games that are considered toxic have turned to Machine Learning Algorithms to try and reduce overall toxicity in their games, but it hasn't been to the best extent [6]. Because of this, we wanted to develop our own ML algorithm trained specifically for gamer tweets, to see if we can create something that can understand and sound like gamers, as this would be a first step to help improve positive interactions online for this community.

Gamer language is unique because they have developed their own semantics for specific emoticons and memes, as well as new words or slang. Research has shown however that these usage of emoticons and memes are contain a similar structure to human real life interactions. Just like how humans pause, make facial contusions, or etc., emoticons/memes represent the same nuances of interaction, meaning they can be analyzed [7]. It is for this reason that we created an ML algorithm that learned from Gamer tweets with these particular nuances, and created our own twitter bot with the corresponding learned language to see whether we can create a human like AI but translated into a different subset of culture. We discuss in future sections our methodologies, and outputs that we created through our bot, and finally discuss our limitations and future directions.

## 2 METHODS

### 2.1 Downloading Tweets

We used an existing Github repository by Max Woolf to download all public tweets from a specified Twitter account (https://github.com/minimaxir/download-tweets-ai-text-gen). Two Twitter accounts were selected for inclusion in our project, @pokimanelol and @lilipichu, who are both female video game streamers. The tweets were downloaded using a python script that scraped a Twitter profile into a single-column CSV file for use in text generation. The script also processed the tweets to remove @ symbols, replies, and extra characters.

### 2.2 Sentiment Analysis

We needed a dataset of positive tweets for our text generator. To create this dataset, we trained a recurrent neural network classifier that would separate the positive and negative tweets we retrieved earlier. We adapted an existing codebase to create our own classifier (https://github.com/andikarachman/RNN-Twitter-Sentiment-Analysis). To train our classifier, we obtained a large dataset of tweets from Kaggle (https://www.kaggle.com/kazanova/sentiment140). This dataset contained 1.6 million tweets labeled positive or negative sentiment. We only used 800 thousand of the tweets due to time constraints in training our classifier. The dataset we used was split evenly between positive and negative tweets.

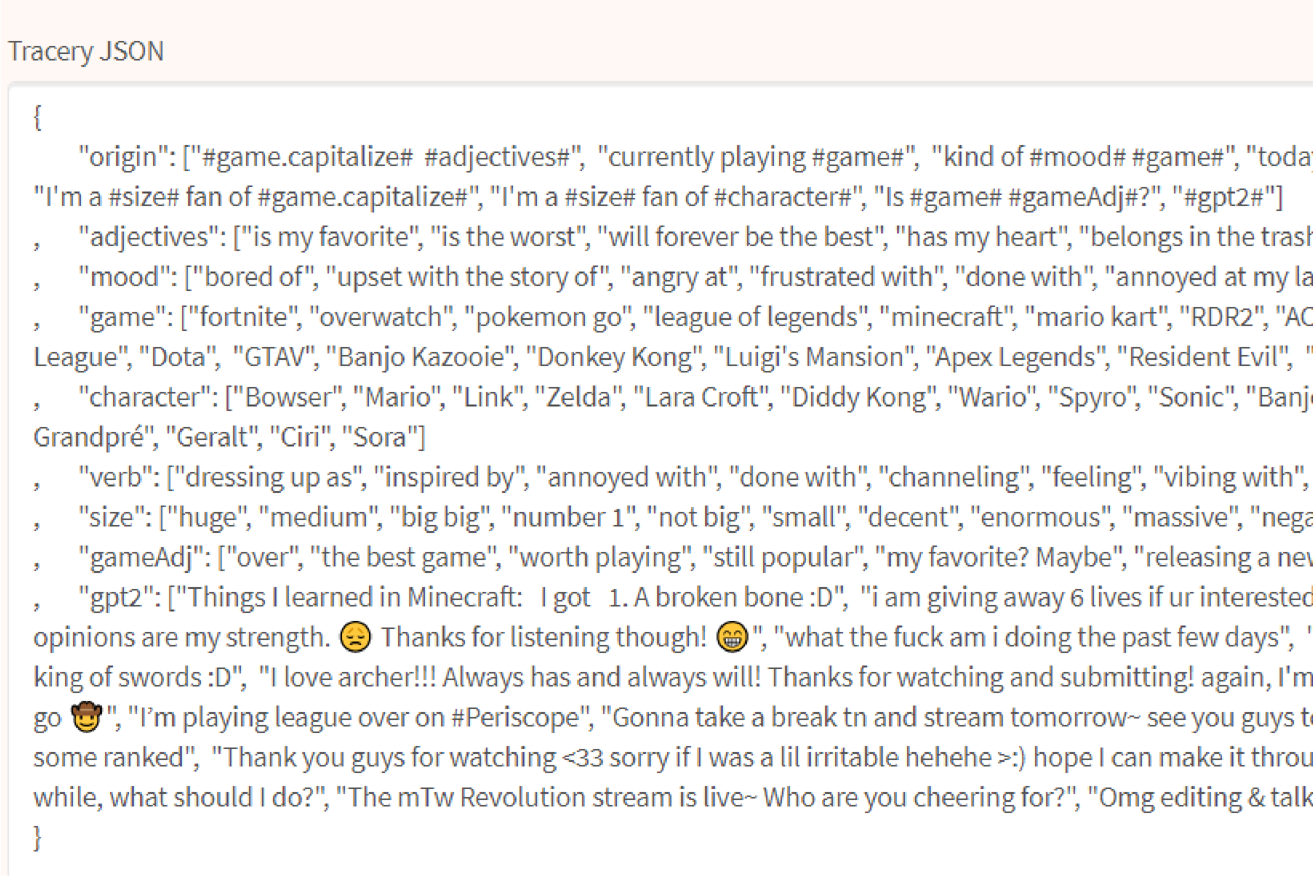
Our neural network consisted of an embedding layer, an long short term memory (LSTM) layer consisting of more layers, and a final sigmoid activation layer. Our tweet pre-processing involved removing punctuation, web links, twitter handles, and digits. Tweets were also truncated or padded to a length of 30 words. This was to deal with very long and very short tweets. Each word was then encoded. Our classifier returned an accuracy of greater than 90 percent. Over 500 thousand unique words were found within the training text. After our classifier was trained, we used it to organize all the positive tweets in our downloaded 'gamer' dataset.

### 2.3 Text Generation

After tweets were downloaded and sorted by valence, a GPT-2 model created by Max Woolf was trained on our dataset and used to generate new tweets. Generative Pretrained Transformer 2, or GPT-2, is a language model created by OpenAI trained on 8 million web pages to predict the next word in a text (https://openai.com/blog/better-language-models/). The original model was not released due to concerns of potentially pernicious applications, but a smaller model was released for use by the public. The model is capable of producing coherent computer-generate text and has been used for a variety of projects. The model was run using the Google Colab environment. The goal was to emulate the voice or tone of the two downloaded Twitter profiles by learning off of the positively categorized tweets. We opted to use a 124M size model due to the relatively small size of our dataset. The model runs 2000 steps, which uses a batch of data to train on each iteration instead of the whole dataset, as with an epoch. After training, up to 1,000 tweets at a time could be generated. We then reviewed the generated tweets and threw out duplicate values to the original dataset. The temperature was set to .7, which was the default value. The temperature ranged from 0 to 1, with values closer to 1 producing more outlandish text.

### 2.4 Twitterbot

Our twitterbot was created using the Tracery tool, built by Kate Compton (tracery.io). Tracery is able to generate text using the Tracery generative grammar and JSON. It's possible to build twitterbots, games, stories, and more from it. The grammar expands text based on user-inputted nodes. We created a Twitter account and set it up to look like a generic gamer profile. Through Tracery, we set the bot to tweet every six hours. We created a set of rule to generate game-related tweets, which were our "human-created" tweets. The bot also was set up to to tweet out the GPT-2 generated tweets. By having two types of tweets, we attempted to make the Twitter profile feel like less of a bot, and attempted to emulate the page of a typical game-related Twitter profile.



Tracery JSON

```
{
      "origin": ["#game.capitalize#  #adjectives#",  "currently playing #game#",  "kind of #mood# #game#", "toda
"I'm a #size# fan of #game.capitalize#", "I'm a #size# fan of #character#", "Is #game# #gameAdj#?", "#gpt2#"]
,      "adjectives": ["is my favorite", "is the worst", "will forever be the best", "has my heart", "belongs in the trash
,      "mood": ["bored of", "upset with the story of", "angry at", "frustrated with", "done with", "annoyed at my la
,      "game": ["fortnite", "overwatch", "pokemon go", "league of legends", "minecraft", "mario kart", "RDR2", "AC
League", "Dota",  "GTAV", "Banjo Kazooie", "Donkey Kong", "Luigi's Mansion", "Apex Legends", "Resident Evil",  "
,      "character": ["Bowser", "Mario", "Link", "Zelda", "Lara Croft", "Diddy Kong", "Wario", "Spyro", "Sonic", "Banj
Grandpré", "Geralt", "Ciri", "Sora"]
,      "verb": ["dressing up as", "inspired by", "annoyed with", "done with", "channeling", "feeling", "vibing with",
,      "size": ["huge", "medium", "big big", "number 1", "not big", "small", "decent", "enormous", "massive", "nega
,      "gameAdj": ["over", "the best game", "worth playing", "still popular", "my favorite? Maybe", "releasing a ne
,      "gpt2": ["Things I learned in Minecraft:   I got   1. A broken bone :D",  "i am giving away 6 lives if ur interested
opinions are my strength. 😞 Thanks for listening though! 😁", "what the fuck am i doing the past few days",  "
king of swords :D",  "I love archer!!! Always has and always will! Thanks for watching and submitting! again, I'm
go 🤠", "I'm playing league over on #Periscope", "Gonna take a break tn and stream tomorrow~ see you guys t
some ranked",  "Thank you guys for watching <33 sorry if I was a lil irritable hehehe >:) hope I can make it throu
while, what should I do?", "The mTw Revolution stream is live~ Who are you cheering for?", "Omg editing & talk
}
```

## 3 RESULTS

*3.0.1 Sentiment Analysis.* Overall, the sentiment analysis output reasonable results. We've included examples below which indicate instances that seem to be well classified as well as ones that appear to be misclassified.

| Tweet | Sentiment | Classification Score |
|---|---|---|
| feelin' very happy and very grateful.. ily guys | positive | good |
| my notifications tab is filled with dog wif hat icons, and i'm not complaining | positive | good |
| feelin' comfy | positive | good |
| life hack: be nice to others | positive | good |
| total of 13,500 trees planted today, thanks to chat | positive | good |
| social media is exhausting | positive | bad |
| lesson learned - i'll get his @ for you guys next time | positive | bad |
| be careful what you believe on the internet | positive | bad |
| all of my bras, undies, and socks were "lost" in the move wtf | negative | good |
| may or may not have blocked someone i played with.. | negative | good |
| broke up with my last boyfriend because he wouldn't stop counting..i wonder what he's up to now | negative | good |
| u look so kawaii!! o(><)o | negative | bad |
| I have so much to be thankful for, and the majority of it only exists because of your support | negative | bad |

*3.0.2 Text Generation.* The GPT-2 model was run on the positive sentiment tweets from @Pokimanelol and @lilypichu, each resulting in 1000 tweets. Examples are included below. Results varied in coherence.

- let's get it let's go
- I'm playing league over on Periscope
- Gonna take a break tn and stream tomorrow see you guys tomorrow <3
- i really amoeba
- Still feelin sick  can't feel well, maybe it's worth going see? come join me
- Aaaay I'm live playin some ranked
- long live the king of swords :D
- I love archer!!! Always has and always will! Thanks for watching and submitting! again, I'm sorry if I didn't get to you. It means a lot you guys watch me.  keep it up!
- you dorkly are so nice and supportive ;-;

https://www.overleaf.com/project/5e6e79853f3f66000100b871

## 4 DISCUSSION

Overall, the twitter bot surprisingly provided a lot of output that could directly sound like the gamer individuals that we trained our program on. However, there were obviously ones that seem a little off or is not perfect. We sorted out most of the tweets and found that 50 percent of the output sounded feasible, while the rest were a little confusing. Another thing to point out to is that if you don't have prior knowledge to these communities and games, then most of these tweets and analyses should sound very foreign. That's our point though, that a lot of the world is separated by different cultures, with their own subcultures, and for these programs to integrate and sound like these niche cultures illustrates that this is something that can possibly be developed and improved.

Right now, our bot utilizes only tweets from two video game personalities, and they are both female. This is a bias that we wanted to directly implement partly due to lack of visibility for female gamers, and also, that they are very popular and focused into two of the most popular video games of Fortnite and League of Legends. We wish we used a way larger dataset of different streamers, and see if we could make replies to tweets and see if people thought it was fake or not, but that would have to be a future exploration. Overall we are happy with our results, and we will explain in the next section the specific limitations that we faced with our data and process.

## 5 LIMITATIONS

### 5.1 Tweets

Downloading tweets caused some issues when later trying to classify. We first tried to download various game-related hashtags, but much of the dataset contained spam or promotional tweets and didn't leave much to work with. This led us to changing our methodology and working off of tweets downloaded directly from Twitter users instead of sifting through hashtags.

When generating new tweets, we experienced a lot of duplicate tweets due to the relatively small size of our dataset. The generated tweets had to be reviewed and compared against the initial dataset in order to determine which tweets were novel.

### 5.2 Sentiment Analysis

Our trained sentiment analysis classifier was not ideal for our 'gaming' tweets. The accuracy was not what we had hoped. A large amount of negative tweets were classified as positive by our classifier. There are several reasons for this.

The language syntax and style of gamer tweets differed greatly from our training set. The gaming tweets featured lots of words that were never encountered in training. Our classifier also could not parse or understand emojis and other special characters. For example, special symbols such as ': (' would have been removed during pre-processing because we could not identify as as a form of sentiment identifier.

A quick glance at conventional tweets found in the training set and those found in the gamer tweets would reveal a large cultural difference in how people communicated. Our classifier at times failed to correctly classify tweets due to this. Ideally, we would have trained our classifier with a large set of gaming-styled tweets. However, we could not find a public set of labeled gamer tweets.

## 6 CITATIONS AND BIBLIOGRAPHIES

[1]Ali Hasan, Sana Moin, Ahmad Karim, and Shahaboddin Shamshirband. 2018. Machine Learning-Based Sentiment Analysis for Twitter Accounts. Mathematical and Computational Applications 23, 1 (2018), 11. DOI:http://dx.doi.org/10.3390/mca23010011

[2]Erik Cambria, Dipankar Das, Sivaji Bandyopadhyay, and Antonio Feraco. 2017. Affective Computing and Sentiment Analysis. A Practical Guide to Sentiment Analysis Socio-Affective Computing (2017), 1–10. DOI:http://dx.doi.org/10.1007/978-3-319-55394-$8_1$[3]*KavehBakhtiyari*, *MonaTaghavi*, *andHafizahHusain*.2014.*Hybridaffectivecomputing¯keyboard*, *mouseandtouchscreen*
*fromreviewtoexperiment*.*NeuralComputingandApplications*26, 6(2014), 1277˘1296.*DOI* : *http* : //*dx*.*doi*.*org*/10.1007/*s*00521–
014–1790–*y*[4]*GeetikaGautamandDivakarYadav*.2014.*Sentimentanalysisoftwitterdatausingmachinelearningapproachesandsemantican*
*http* : //*dx*.*doi*.*org*/10.1109/*ic*3.2014.6897213[5]*JacobEuteneuer*.2018.*Defininggames*, *designingidentity*, *anddevelopingtoxicity* :
*Futuretrendsingamestudies*.*NewMediaSociety*21, 3(*August*2018), 786˘790.*DOI* : *http* : //*dx*.*doi*.*org*/10.1177/1461444818809716[6]*Marcus*
*http* : //*dx*.*doi*.*org*/10.1109/*netgames*.2015.7382991[7]*RobertR*.*Provine*, *RobertJ*.*Spencer*, *andDarcyL*.*Mandell*.2007.*EmotionalExpression*
*http* : //*dx*.*doi*.*org*/10.1177/0261927*x*06303481