

Experiments

The approach was to take [GPT-2](#) and finetune it on joke data.

For a number of first experiments, I took two relatively small datasets: [stupidstuff](#) and [semeval](#). This was about 11.000 lines of text, and I performed three experiments with different parameters:

- 1) Run 1. I chose parameters based on my experience: 3 epochs, block size 256. But these turned out more like stories, not jokes:

```
So, this guy comes into the bar and sees you and says "Ma'am, I can see you." He turns around and says, "You have no arms. Are you thinking?" "Well, I'm not." "Do you want me to show you any arms or legs?" "No, I'm just thinking. You're a magician and I'm here to tell you something." A young lawyer goes to court to sue his neighbor. He asks for a new lawyer. The lawyer says, "I'm married to my second wife, and we live in a big house on the hill. This house has a balcony, and a swimming pool. You walk up to the balcony and give me a $5,000 bill
```

- 2) Run 2. I decided to make the block size smaller (128) and try more epochs (6). This time, the joke seemed to start but didn't finish:

```
Have you ever noticed  
2021-08-25 03:13:34.960504: I tensorflow/stream_executor/platform/default/dso_loader.cc:48] Successfully opened dynamic library cudart64_101.dll  
Have you ever noticed that the first person you talk to at work is also the person you're talking to about after you make the call? A: One. When you pick them up you say, "I am sorry to disturb you, but I've just had an unusual call; your name is not in the book." Q: What does a naked man have to lose in order to appear in a film? A: A 1948 reproduction of a stock photograph. Q: Why are there no female directors in the directors' cut of "A Walk to Remember"? A: Because women won't appreciate the original footage anyway.
```

- 3) Run 3. Changed the block size back to 256, plus added "Joke: " in the beginning of each joke in the dataset. Now we don't have to come up with the beginning of a joke each time (can just type "Joke: "), plus it's easier to separate jokes from each other.

```
Joke:  
2021-08-25 04:07:41.501301: I tensorflow/stream_executor/platform/default/dso_loader.cc:48] Successfully opened dynamic library cudart64_101.dll  
Joke: "All I can say is... we can't discuss politics." - Taylor Swift.Joke: *watching 2 cats beating each other on the bathroom scale* They look like this because I can't find the pen, hahaJoke: When do we stop thinking of history and start loving the people who created it?Joke: "Sometimes I feel like I didn't get my fill of milk before, but now I'm the guy that got it."Joke: How does a Japanese kid get a job? He puts on a uniform and gets a haircut.Joke: What do a married woman and a child get each year? Both have their eggs laid.Joke: What do you call a group of people who are both obsessed with cars and lovers of women? A compulsive shopper.Joke: The reason you don't see a black pope is because black people can always get away with it.Joke: What did the drunk driver say when he accidentally kicked his girlfriend's head off a tree? I've never seen a drunk do that.Joke: When my wife and I are having a car repair, it's my job to check the status of every car we buy. If it doesn't work as planned, I'm going to
```

```
Joke: How  
2021-08-27 18:10:14.844942: I tensorflow/stream_executor/platform/default/dso_loader.cc:48] Successfully opened dynamic library cudart64_101.dll  
Joke: How do you know if a girl likes you? They might like you later.Joke: What does a hooker have in common with a hooker? Nothing. She can't have a hook.Joke: "If everything you're made of is better than nothing, you'll be okay" - Lilo and StitchJoke: Me: I'm just curious why, this late at night, when my alarm clock goes off? Wife: You're late!Joke: Q: What do you call a Jewish guy with a big belly? A: Bag-dwelling Jewish guy.Joke: I was sitting on the porch, watching the kids play when suddenly a dog suddenly ran up to me. I just turned around and said "What the fuck is happening to me."Joke: What's the difference between a stripper and a pig? A pig walks into a pig bar, a stripper walks out.Joke: Why do black people always come back to murder their girlfriends? Because they didn't have the balls to do it.Joke: Do you know what it's like to be gay in Italy? You are always pulled over.Joke: Why aren't Black Trans men involved in organized crime? They do
```

At this point, the results already resemble jokes (yay!). I decided to make them better by taking more data and preprocessing them. I took five datasets that could find:

	initially	<=200 chars	QA jokes
kaggle	231 656	231 656	66 913
reddit	194 552	147 777	62 661
semeval	7 999	7 389	1 192
stupidstuff	3 199	787	80
wocka	10 010	4 083	899
All by now	447.416	391.692	131.745
final clean		388 483	130 787

Since we are interested in short jokes and the kaggle dataset contained only jokes of length less than 200 symbols, from all other datasets I also deleted longer jokes. With some additional preprocessing, there were ~388.500 lines in the dataset, and this was my next experiment:

4) Run 4. The same as Run 3 but with more data.

```
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$ python3 generate.py --model 4
Enter start of the joke:
Joke:
Joke: Joke: I bet if Donald Trump became president the first thing he would do would be... Make America Great Again.Joke: What is the opposite of progress? CongressJoke: I just did a stand up routine called "I Want to Be the Next Prime Minister of Canada." They asked "What's your name?"Joke: What did the elephant say to the naked man? What an ass!Joke: What do you call a cow who just gave birth? Decafinated!Joke: Why did the cannibal take a bath? So he could drink his own tearsJoke: Why are lesbians so good at basketball? They don't run around shooting, but they shoot.Joke: There's a new Barbie doll on the market - Ocean's Eleven Barbie...will retail for $17.99 USJoke: What is the definition of a virgin? When your dick is in her ass.Joke: How do you get a dog to say hi? You drop him ate a beerJoke: What is black and always floats? George ZimmermanJoke: Why do we still call black people "normals"? Because they can't even make cents.Joke: *sends text* "You must be mad." M
```

Then, I also wanted to try an idea I found in [the blog post by Lorenzo Ampil](#): to take only jokes starting with “what/where/why/when” because it makes the data more homogeneous. This significantly reduced the size of train data, to 130.700 lines of text. The next two runs were similar to each other:

5) Run 5. Take QA jokes and the same parameters as the two previous runs, but without the starting “Joke: ” string. Plus, add special symbols to the jokes: <soq> (start of question), <eq> (end of question), <|endoftext|>.

```
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$ python3 generate.py --model 5
Enter start of the joke:
<soq> Why did the chicken cross the road? <eq>
<soq> Why did the chicken cross the road? <eq> Because it had its horn blew out. <|endoftext|
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$ python3 generate.py --model 5
Enter start of the joke:
<soq> When did
<soq> When did the chicken cross the road? <eq> When it realized it was chicken, then it stopped being chicken! <|endoftext|
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$ python3 generate.py --model 5
Enter start of the joke:
<soq> Why did the man
<soq> Why did the man buy a camera <eq> To shoot for Canon. <|endoftext|
```

6) Run 6. The same as the previous one, but without the special symbols (I think the question words at the beginning are supposed to have a similar role).

```
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$ python3 generate.py --model 6
Enter start of the joke:
What do
What do icesia twins get for Christmas GingervitisHow many Alzheimer's patients does it take to screw in a lightbulb More than four, as my basement is still dark.Why did the chicken cross the road? To
get to the idiots house.... Knock knock Who's there? The chicken.What do you call a gay guy in a coma A tomato.Why did God give women yeast infections So they'd know what it's like to live with an irrit
ating cunt.What do you call the entrance to an Asian restaurant An UberWhy couldn't the toilet paper cross the road It was stuck in a crack.what would an alcoholic Mexican say? Ayyyy, please!Whats the d
ifference between a bird's eye view and my penis I actually give a shit when I fuck a bird's eye viewWhat does a man with Parkinson's ask after sex Urine troubleHow does Michael Jackson make his coffee?
He shines it on the line. Why is six afraid of seven? Because seven is a registered six offender. What's the difference between my basement and an ISIS base I don't know I just fly the drone.What do yo
u call a cheap circumcision A rip off.Why is it called a day-glo It's a big piece of cake that bring
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$ python3 generate.py --model 6
Enter start of the joke:
Why did the chicken cross the road?
Why did the chicken cross the road? To get to the idiot's house. Knock knock Who's there? The chicken!Why is the number 7 afraid of the number 2? Because it's 8.What's an Ethiopian's favorite sport? Cr
oss CountryWhat does a Muslim and an ISIS fighter have in common They both got stoned for burning it.What are Mario's overalls made out of Denim denim denimWhat are the three rings of Marriage The engag
ement ring, the wedding ring, and the sufferingWhat was the last thing that went through the fly's mind when it hit the windshield of its car? Its asshole.What does a cannibal do after dumping his girlf
riend He wipes.What do you call two Mexican guys playing basketball Juan on JuanWhat's the difference between roast beef and pea soup Anyone can roast beefWhy did the pervert cross the road? Because his
dick was stuck in the chicken.Why can't you hear a pterodactyl in the bathroom Because the P is silent.What do you call a lesbian dinosaur? A lickalotapus.How do you know your girlfriend is too young f
or you? When she starts telling fucked-up jokes.What do you call a lesbian dinosaur? A lickalotapus.Why did the scarecrow win the Nobel prize Becaus
(venv) anna@anna-65-5590:~/Documents/Charles/Studying/Lang technologies/humor_project$
```

Examples

I think, that among runs 4-6 there's no run that is particularly better than others, but in each, some good examples can be found. Here's a list of some jokes that I thought were OK:

RUN 3

Joke: What did Jesus say to Eve when she went for a walk? "You know you'll find a way out"

Joke: Have you ever wondered why the Jews are so good at trivia? Because they have an essay called "Jews as a people."

Joke: Do you ever think about your life as an immigrant? Do you have to wait until you die before people ask you for directions?

Joke: One day a man called to his wife. "Wife, I have some great news and some horrible news. I'm going to jail for life. What's the bad news?"

RUN 4

Joke: Did you hear about the man who tried to steal a calendar? He went months without a date

Joke: Why did the chicken cross the road? To get to the other side!

Joke: What did the farmer say when his cows gave birth? Moo-seum

Joke: What do you call a fake noodle? An impasta

Joke: I like my slaves like I like my coffee Free

Joke: A feminist, a vegan and a CrossFitter walk into a bar... I know because they told everyone within 2 minutes.

Joke: Why couldn't the pirate learn the alphabet? Because he was lost at C.

RUN 5

<soq> What do you call a black guy in space <eoq> An astronaut you racist fuck.
</endoftext|

RUN 6

What did the fish say when it swam into a wall Dam!

Why don't you ever see elephants hiding in trees? Because they are really good at it.

Why are the Chinese afraid of snow It's white and it's on their land.