NLP: Project Pickup

Mackenzie Bookamer and Thalia Koutsougeras

PROBLEM

STEM majors need a little extra help in the dating world...

OUR SOLUTION: *PICKUPMASTER 9000*

A CS/Math-themed pickup line generator that can help you think up your next nerdy pickup line and help you practice your skills!

- Input: Half of a sentence or a question
- Output: Completed sentence or question-answer format

DATA

613 CS pickup lines and jokes from 7 different websites
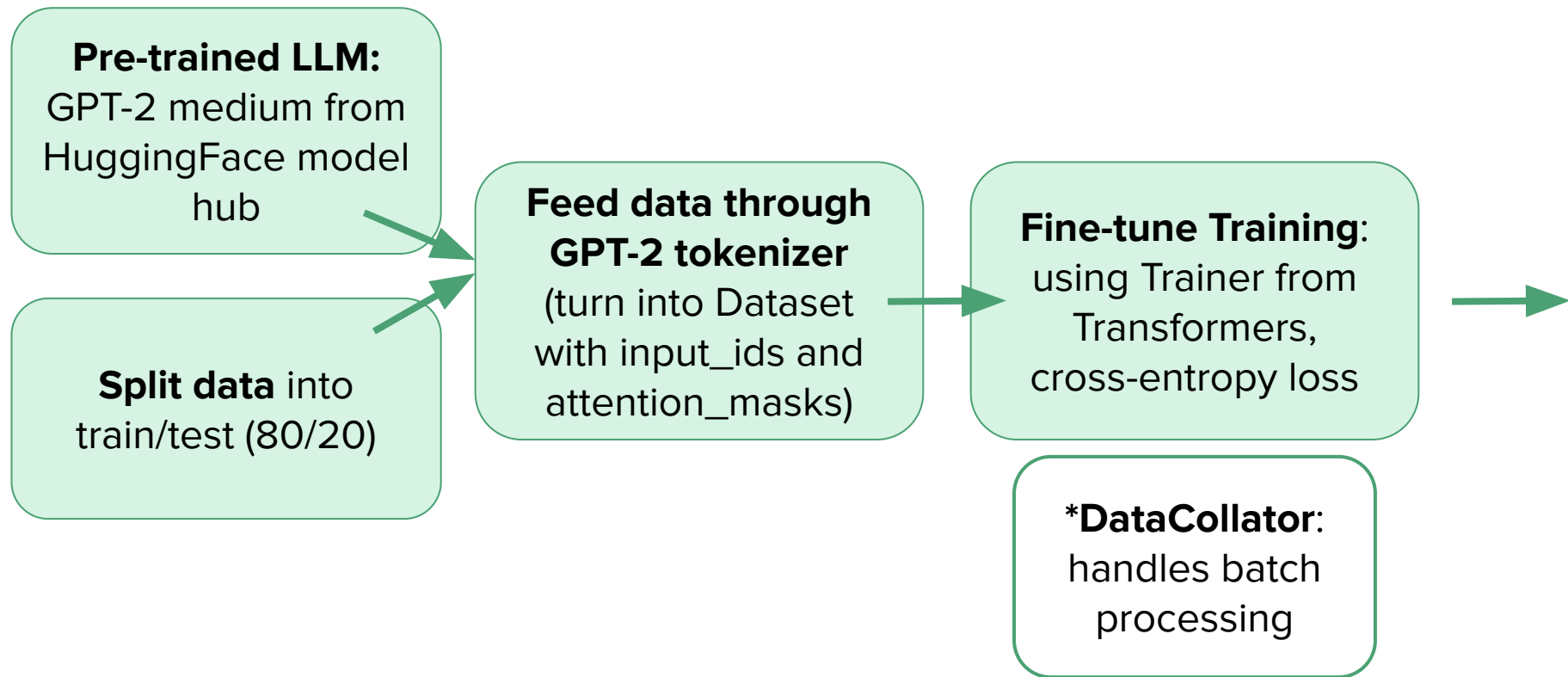
280 math pickup lines from 6 different websites

➜ 893 lines of data stored in 1 csv file

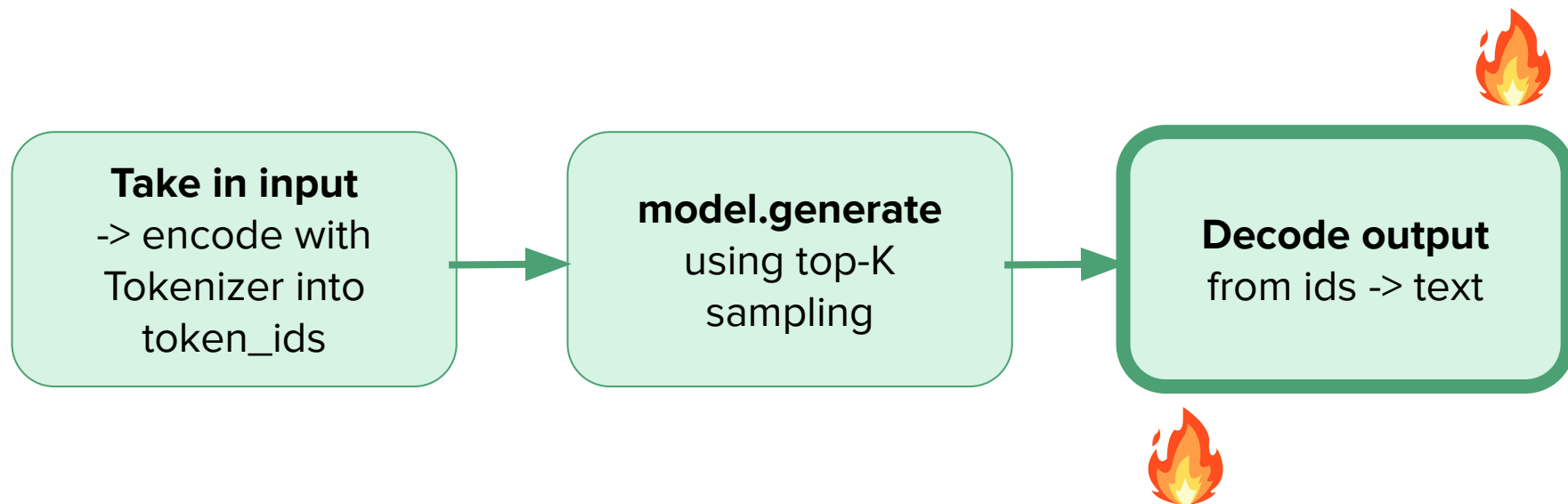**Funny Computer Programming Pick Up Lines**

- Are you a keyboard? Because you're my type.
- Are you a computer virus? Because my heart skips a beat every time I see you.
- Is your name Google? Because you have everything I've been searching for.

https://wittypro.net/pick-up-lines/computer-programming-pick-up-lines-and-rizz/

METHOD: TRAINING

Pre-trained LLM: GPT-2 medium from HuggingFace model hub

Split data into train/test (80/20)

Feed data through GPT-2 tokenizer (turn into Dataset with input_ids and attention_masks)

Fine-tune Training: using Trainer from Transformers, cross-entropy loss

*DataCollator: handles batch processing

METHOD: GENERATION

Take in input
-> encode with Tokenizer into token_ids

model.generate
using top-K sampling

Decode output
from ids -> text
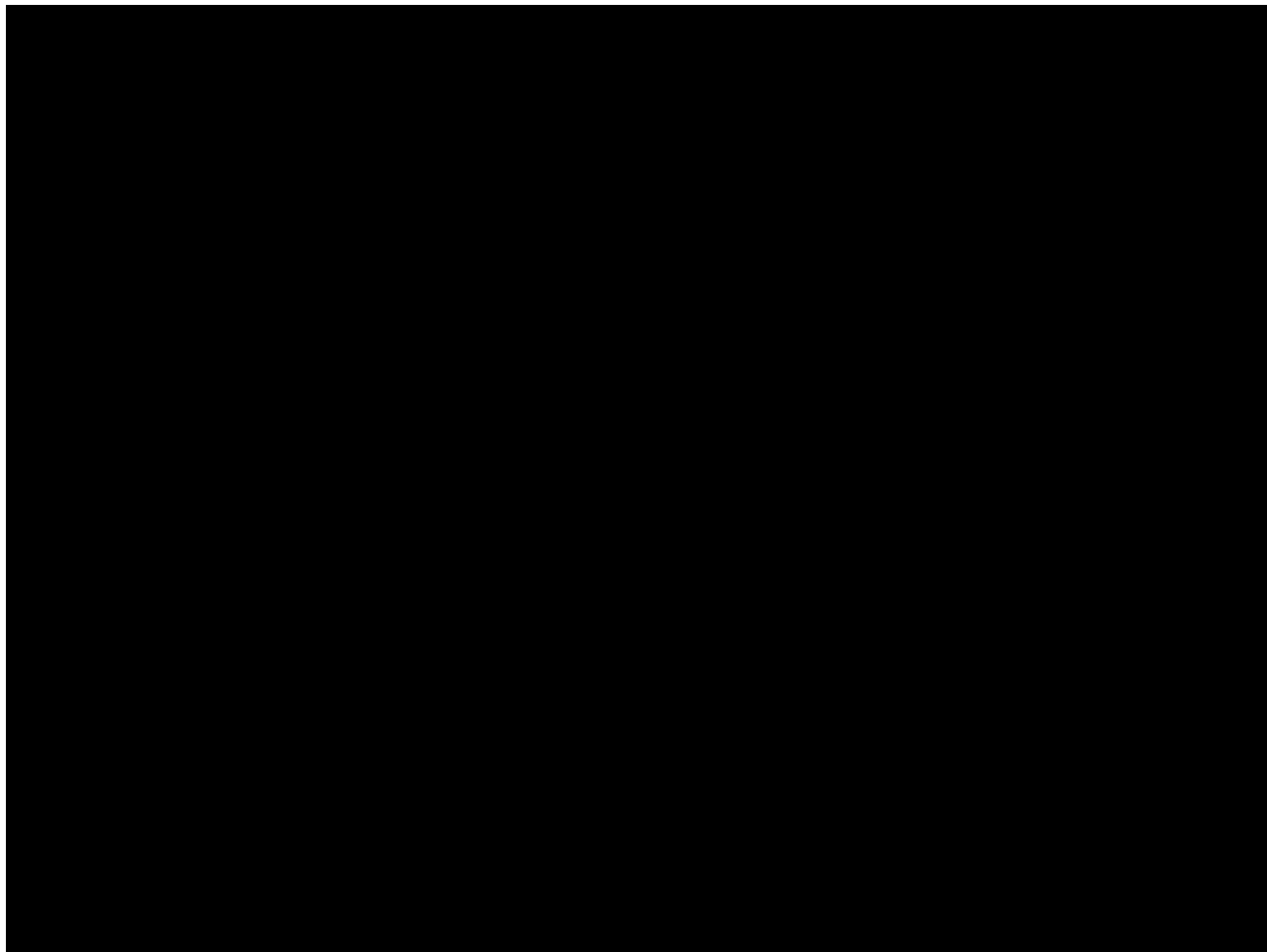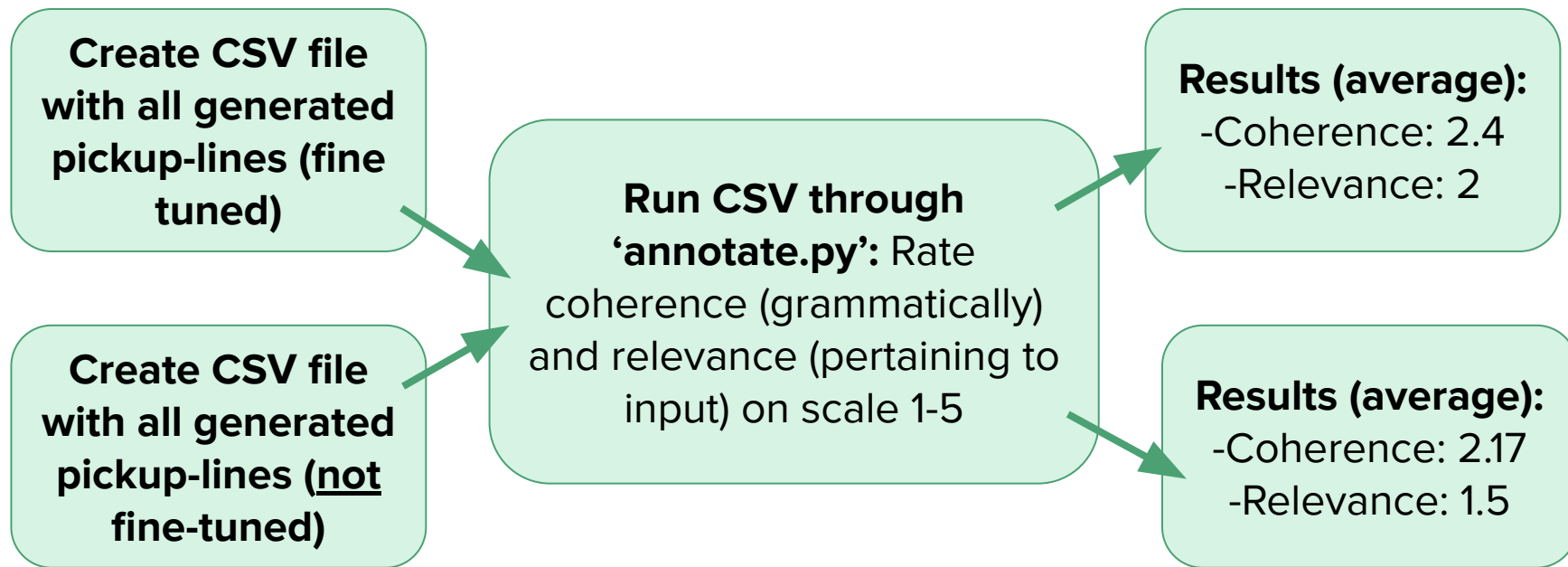
Demo

RESULTS: GENERATED LINES

- **Are you a graph? Because you've got my heart all connected. It's a pattern of mathematical equations.** Every time I see it, my heart goes a little faster.I don't know if it's
- Your must be a compiler because you make my code look like an assembly! <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <3 <
- Are you GitHub? Because I'm looking up your status! Because you're my favorite project repository! Because you're always working on the same thing, so you end up getting the most work done! Because you're like a computer, you can always
- My love for you is like a fractal because you have infinite growth, and I can never stop growing. You have my entire heart, and your beauty is a miracle. I cannot express my love for you without growing your love. Because you are
- Are you a Wi-Fi signal? Because you're getting more and more connected with me. I'm seeing you in my heart." I'm feeling connected with you."
- <u>Not fine tuned:</u> *Are you a graph? Yes. What kind of graph? A graph is a collection of data that is organized into a tree structure. How many trees does it have? It can be as many as*

RESULTS: HUMAN ANNOTATION

Create CSV file with all generated pickup-lines (fine tuned)

Create CSV file with all generated pickup-lines (<u>not</u> fine-tuned)

Run CSV through 'annotate.py': Rate coherence (grammatically) and relevance (pertaining to input) on scale 1-5

Results (average):
-Coherence: 2.4
-Relevance: 2

Results (average):
-Coherence: 2.17
-Relevance: 1.5

Questions?