Assignment 7: Societal & Ethical Implications of Language Technologies

CSE 256: Statistical NLP: Spring 2022

University of California, San Diego

Released: May 30, 2022

Due: June 6, 2022

In this last assignment of the course, you will think about potential societal and ethical implications of powerful language technologies such as GPT-3, the subject of the second assigned paper for this course. Recently, a thorough article on this topic, with a focus on GPT-3, appeared in The New York Times (NYT), April 15, 2022, titled "A.I. Is Mastering Language Should We Trust What It Says?" by Steven Johnson. We will use this article as a reference.
The readings for this assignment are:

- Main reading (if you have subscription, you can access it directly on NYT website). For the purpose of this assignment, you will only need to read until page 6 of the linked PDF. However, the article is worth the read, we encourage you to read the entire 12 pages.
- Bonus points reading: a critique of the article written by Professor Emily Bender.

Note: The problem of fairness, bias, and other ethical considerations, is as much of a philosophical one as it is a technical one. Often there is no obvious "right thing to do", and it has even been shown mathematically that it is impossible for a machine learning classifier model to satisfy reasonable fairness criteria (Kleinberg et al., 2016). Nevertheless, we need to be aware of the real-world impact of the systems we build, and understand the relationship between ideas and consequences. In that spirit, a small part (2.5 points) of this assignment will ask for your opinion, when this is the case, the question will be marked with (*opinion*).

Additional notes:

1. Unless stated, questions in this assignment are asking for answers from the Johnson article.
2. *Italicized text blocks are direct quotes from the Jonson article.*
3. For most questions, we expect your answer to be about 1-3 sentences, but can be longer if desired.

Part 1. The Game (1 Point)

Q1 *... Mostly, it is playing a kind of game, over and over again, billions of times a second. And the game is called: Guess what the missing word is.*

Which NLP task is described by the author as a game of guessing the missing word?

Part 2: Opportunities and Challenges (4.5 Points)

Q2 According to the author, what kind of jobs could be lost in the future because of software like GPT-3? Name three such jobs *(1.5 Points)*. (*Opinion*) Which of the three jobs is the least likely to be replaced by GPT-3-like models anytime soon, and why? *(0.5 Point)*

Q3 *... But as GPT-3's fluency has dazzled many observers, the large-language-model approach has also attracted significant criticism over the last few years.*

According to the author, what are some specific criticisms that have been made against GPT-3 like models? Give three criticisms *(1.5 Points)*.

Q4 *Gebru and a group of co-authors declared that large language models were just "stochastic parrots": that is, the software was using randomization to merely remix human-authored sentences.*

Why is it a societal problem if language models are simply "stochastic parrots" *(1 Point)* .

Part 3: Open AI Remedies (1 Point)

Q5 According to the article, roughly, a fifth of Open AI is focused on "safety" and "alignment". What are some specific things they are doing to "align the technology with humanity's interests". Give two examples *(1 Point).*

Part 4: Intelligence or Lack of (4 Points)

Q6 *...One puzzling - and potentially dangerous - attribute of deep-learning systems generally is that it's very difficult to tell what is actually happening inside the model. You give the program an input, and it gives you an output, but it's hard to tell why exactly the software chose that output over others. This is one reason the debate about large language models exists. Some people argue that higher-level understanding is emerging, thanks to the deep layers of the neural net. Others think the program by definition can't get to true understanding simply by playing "guess the missing word" all day. But no one really knows.*

From what you have learned in the course and/or the Johnson article, what arguments can be made to support the thesis that "higher-level understanding is emerging, thanks to the deep layers of the neural net". Make two points. You can give specific applications/datasets/evaluations as examples or make more abstract arguments. *(2 Points).*

From what you have learned in the course and/or the Johnson article, what arguments can be made to support the thesis that "the program by definition can't get to true understanding simply by playing "guess the missing word" all day". Make two points. You can give specific applications/datasets/evaluations as examples or make more abstract arguments *(2 Points).*

Part 5: Regulation (2 Points)

Q7 *... It seemed as if the cycle of corporate consolidation that characterized the social media age was already happening with A.I., only this time around, the algorithms might not just sow polarization or sell our attention to the highest bidder - they might end up destroying humanity itself. And once again, all the evidence suggested that this power was going to be controlled by a few Silicon Valley megacorporations.*

Second quote:

... Wherever you land in this debate, the pace of recent improvement in large language models makes it hard to imagine that they won't be deployed commercially in the coming years. And that raises the question of exactly how they — and, for that matter, the other headlong advances of A.I. — should be unleashed on the world. In the rise of Facebook and Google, we have seen how dominance in a new realm of technology can quickly lead to astonishing power over society, and A.I. threatens to be even more transformative than social media in its ultimate effects. What is the right kind of organization to build and own something of such scale and ambition, with such promise and such potential for abuse? Or should we be building it at all?

(*Opinion*) Discuss your own thoughts with regard to the two questions at the end of the quote above (*2 Points*).

Part 6: "Resisting the Urge to be Impressed" (Bonus)

Q8 Provide three specific arguments against the Johnson article as pointed out in the critique piece in Professor Emily Bender*(1.5 Bonus Points)*

Submission Instructions

Submit your work on Gradescope.

- **Code:** There is no code for this homework
- **Report:** Submit your report, it should be **1-2 pages long, in pdf** (reasonable font sizes).