Deep fake

Natural language generation

Generating plausible text

A lot of decisions will be made, said Trump. And that will be a little bit less than a day before it hits.

"I was really amazed at the number of people who come to see me because they are so anxious to see me as this historic president," said Trump.

But as much as the crowd likes to bask in the glitz, it can't have failed to be a little bit scared. And it's hard not to imagine what they're going through during their journey to the Oval Office.

The crowd will be watching the news for updates on what Obama's up to in the White House. Trump will be looking over his shoulder trying to figure out the next move he can take from here.

How does it work?

Step 1: Collect *a lot* of text

- 8 million web pages:
  - Wikipedia
  - News papers
  - Speeches of politicians
  - Books: Lord of the Rings, Harry potter..
  - …

Step 2: "*Train*" a model on this text

- Show some text to the model, ask it to predict the rest
- Tell the model what was the correct continuation
- Allow the model to update itself
- Repeat this process until it's predictions are reliable

What does this model look like?

Version 1: n-gram models

$w_1\ w_2\ w_3\ w_4\ \ldots.$ → w5?

A lot of decisions ..

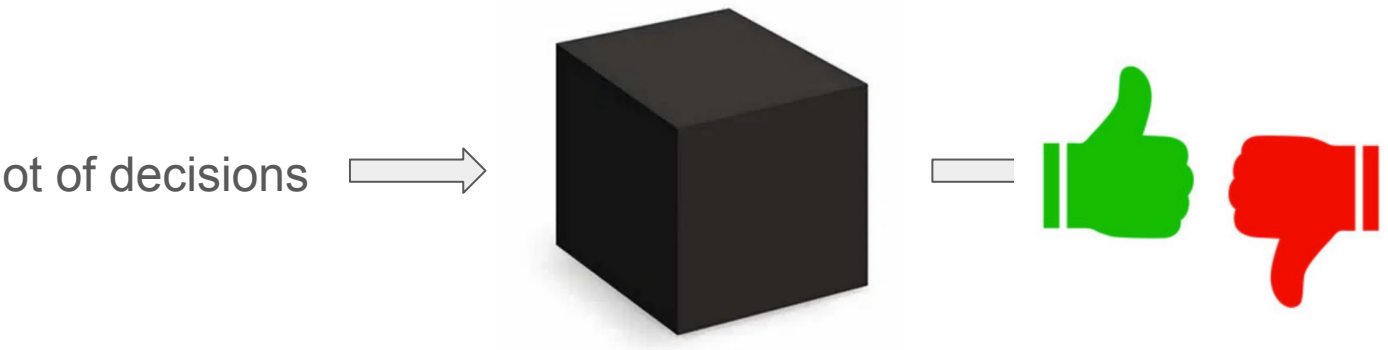
are	2458
that	521
about	210
certainly	2
make	139
….	

What does this model look like?

Version 2: Recurrent neural networks

$w_{-n}\ w_{-n+1} \ldots w_1\ w_2\ w_3\ w_4 \ldots. \rightarrow w_5?$

Arbitrary length history A lot of decisions



What do these black boxes look like?

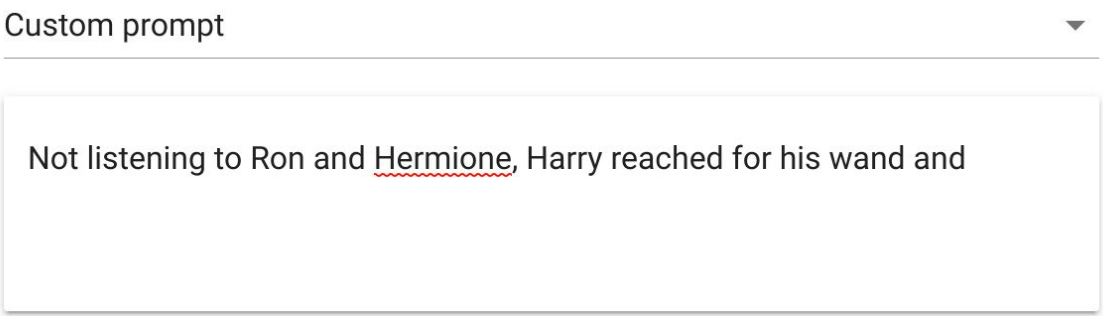
Blackbox models

Pro: we don’t have to hand-design what they do anymore, which means we don’t have to understand the underlying problem. We can just throw lots and lots of data at this system and it will discover regularities

Con: We have no idea what this model is doing, which makes it difficult to influence it when -- for instance -- it starts to generate racist language

Con: If a model that we don’t understand generates very realistic text, how can we keep it apart from real text?

So... are AI’s ready to take over the world?



So... re AI’s ready to take over the world?

Not listening to Ron and Hermione, Harry reached for his wand and spoke, "Expelliarmus." The dark magic hit the wall behind Hermione, and Harry's wand hit it, and the wall fell back and crumbled."

"Expelliarmus!" roared Hermione. She flew into Harry's arms and sobbed hysterically.

"Expelliarmus!" yelled Ron. "Hermione! Get out of the shower! We've got a problem!"

"What is it?" Ron demanded, turning to look at Harry. "It feels great!" He pointed towards the shower. "It feels good to be Ron's friend! He's my best friend!"

Harry blinked at Ron so he realized he missed the shower room.

Are AI’s ready to take over the world

Models can very convincingly

- generate grammatically correct, locally coherent sounding text

They cannot:

- Maintain coherence at the global level
- Conceptually understand the world they are talking about

Final slide?