Deep fake

Natural language generation

Generating plausible text

www.talktotransformer.com

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

Generating plausible text

www.talktotransformer.com

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..
  - …

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
- Adapt the parameters of the model such that it improves this prediction
- 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?

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 ..

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
that
about
certainly
make
….

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 1: n-gram models

$w_1\ w_2\ w_3\ w_4\ \ldots.$ → w5?

A lot of decisions ..

are	2458	25%
that	521	5.2%
about	210	2.1%
certainly	2	0.02%
make	139	1.4%
….	...	...

What does this model look like?

Version 2: Recurrent neural networks

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

What does this model look like?

Version 2: Recurrent neural networks

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

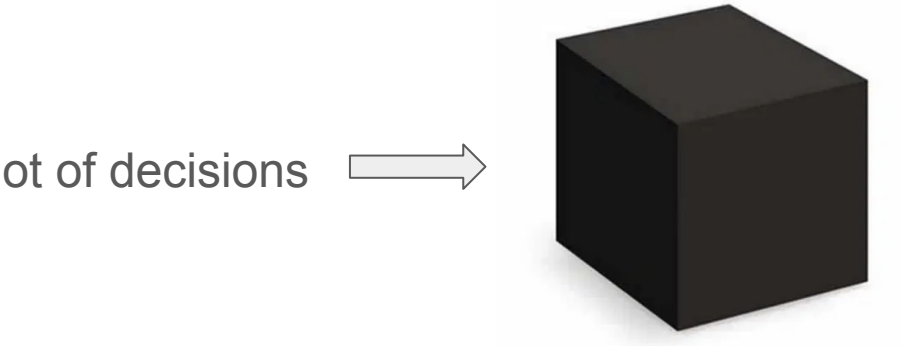
Arbitrary length history A lot of decisions

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

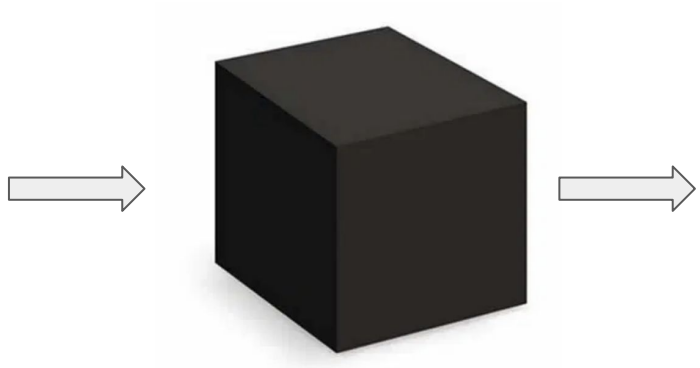


What does this model look like?

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

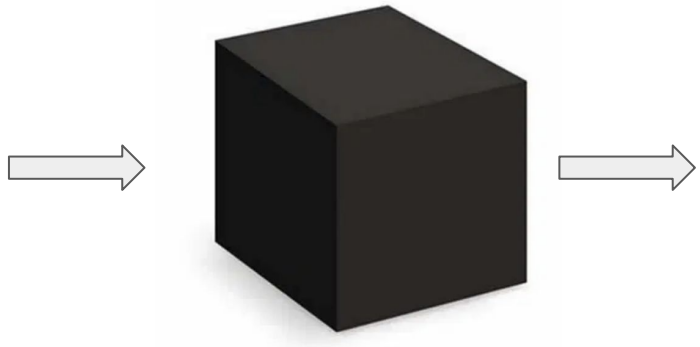


What does this model look like?

Version 2: Recurrent neural networks

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

Arbitrary length history A lot of decisions



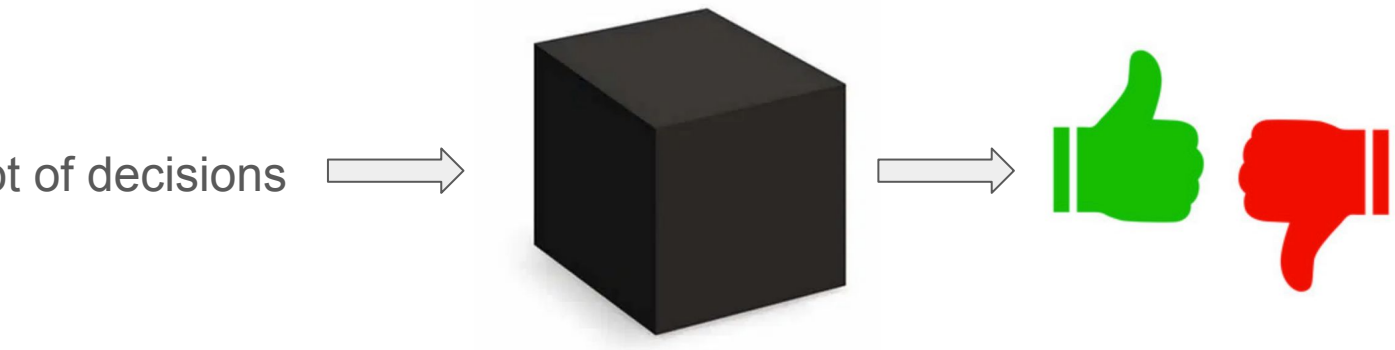
are

What does this model look like?

Version 2: Recurrent neural networks

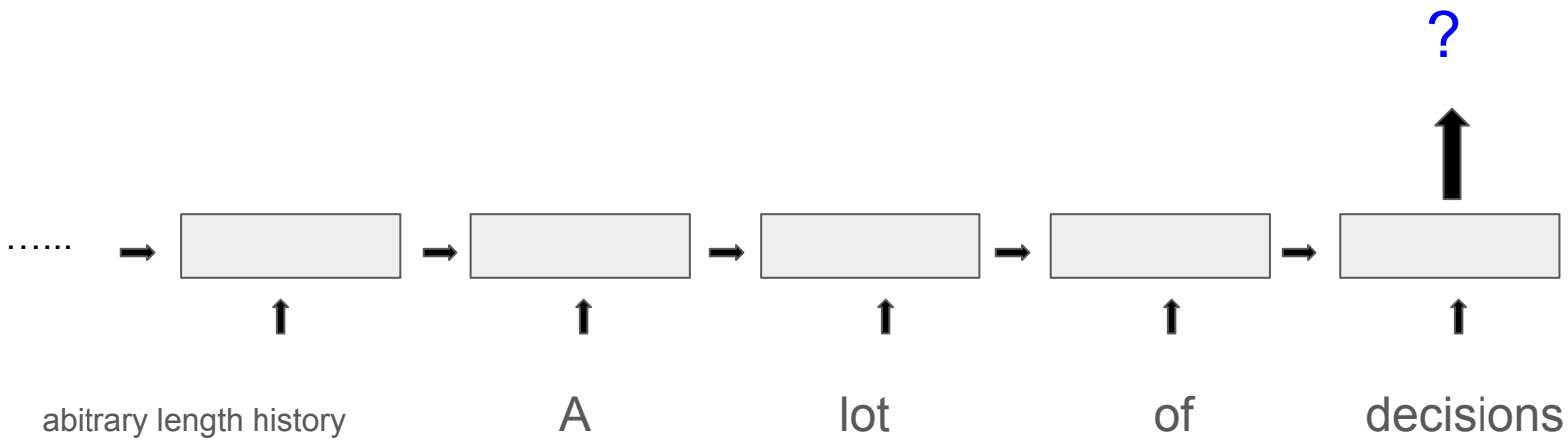
$w_{-n}\ w_{-n+1} \ldots w_1\ w_2\ w_3\ w_4 \ldots. \quad \rightarrow \quad w_5?$

Arbitrary length history A lot of decisions



What do these black boxes look like?

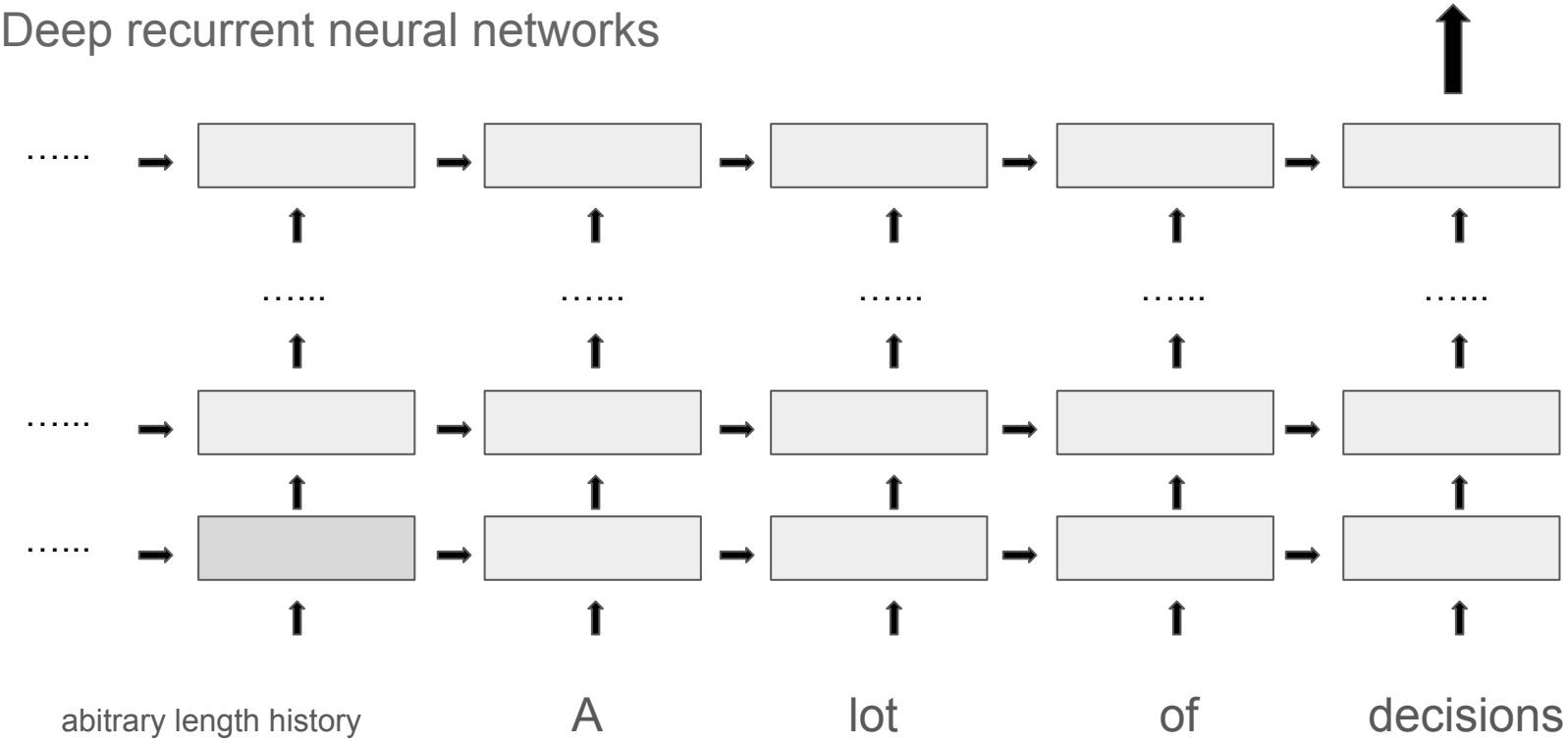
Recurrent neural networks



What do these black boxes look like?

?

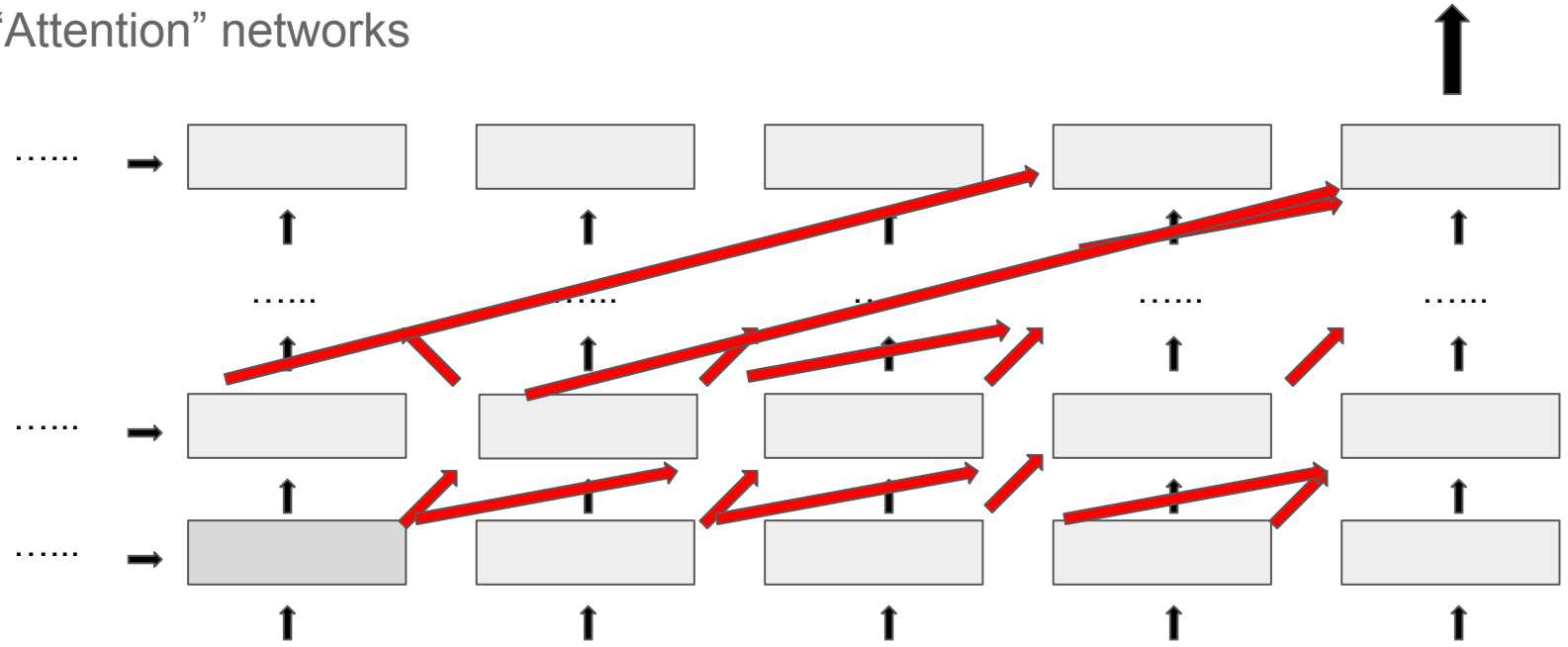
Deep recurrent neural networks



What do these black boxes look like?

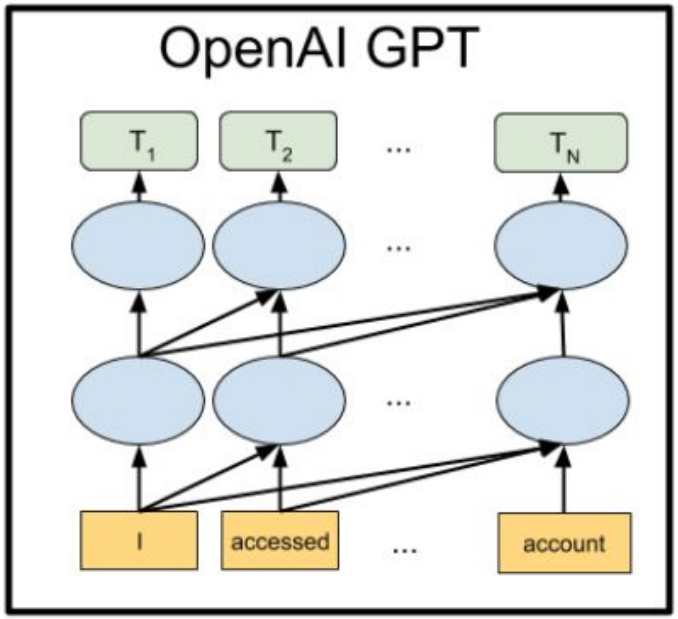
?

"Attention" networks



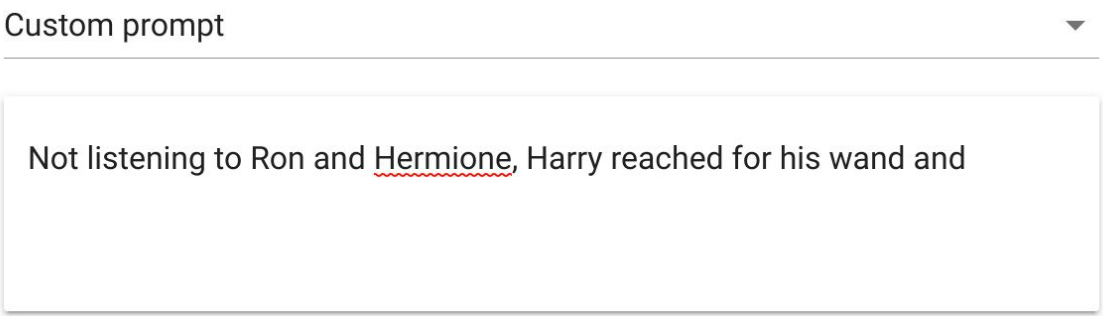
What do these black boxes look like?

"Attention" networks



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

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



So... are 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.

So… Are AI’s ready to take over the world

Models can very convincingly generate intriguing, grammatically correct and locally coherent text

But..

- Globally, texts are incoherent
- They do not conceptually understand the world they talk about
- They have no “common sense”