

Intro to Data Science

ChatGPT and the new frontier of data science

Prof. Bisbee

Vanderbilt University

Lecture Date: 2023/08/28

Slides Updated: 2023-08-16

Agenda

1. ChatGPT overview
2. Interactive examples with the homework

ChatGPT overview

- **L**arge **L**anguage **M**odel (LLM)
- Large:
 - Trained on the internet
 - Billions (trillions) of parameters
- Language model:
 - A "transformer"

ChatGPT overview

- Fill in the missing word: "**I drove my _ to the supermarket.**"
- How do we know this?
- Patterns!



ChatGPT overview

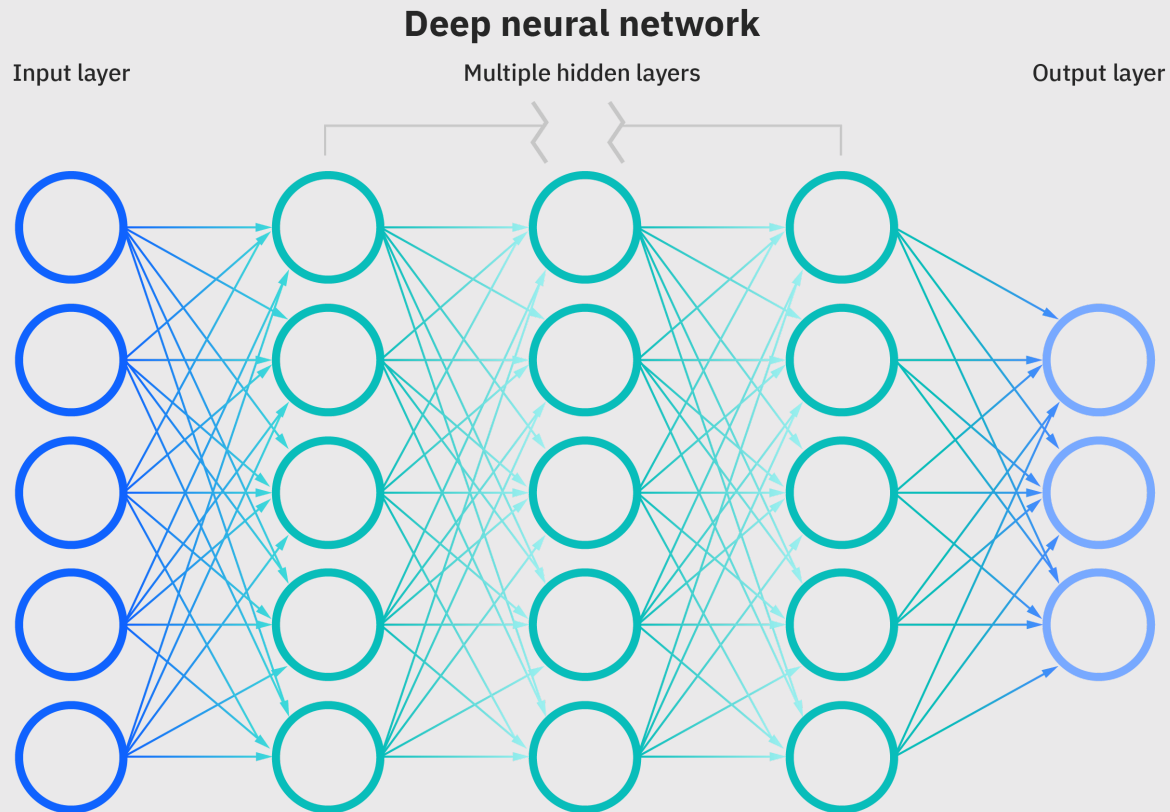
● ● ● ● ● ●

“I drove my _____ to the supermarket.”

○ Truck	●	●	●	●	●	●
○ Car	●	●	●	●	●	●
○ Bike	●	●	●	●	●	●
○ Horse	●	●	●	●	●	●
○ Crocodile	●	●	●	●	●	●

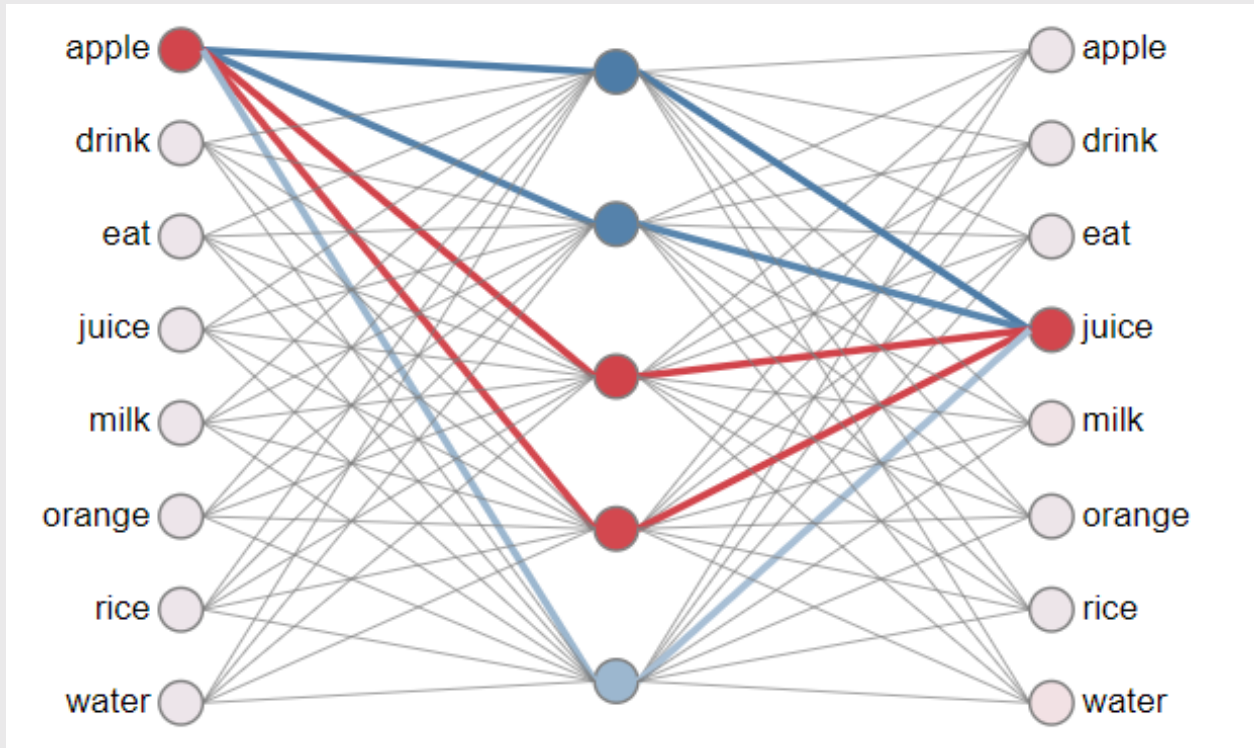
ChatGPT overview

- Researchers have been trying to emulate the brain's ability for years



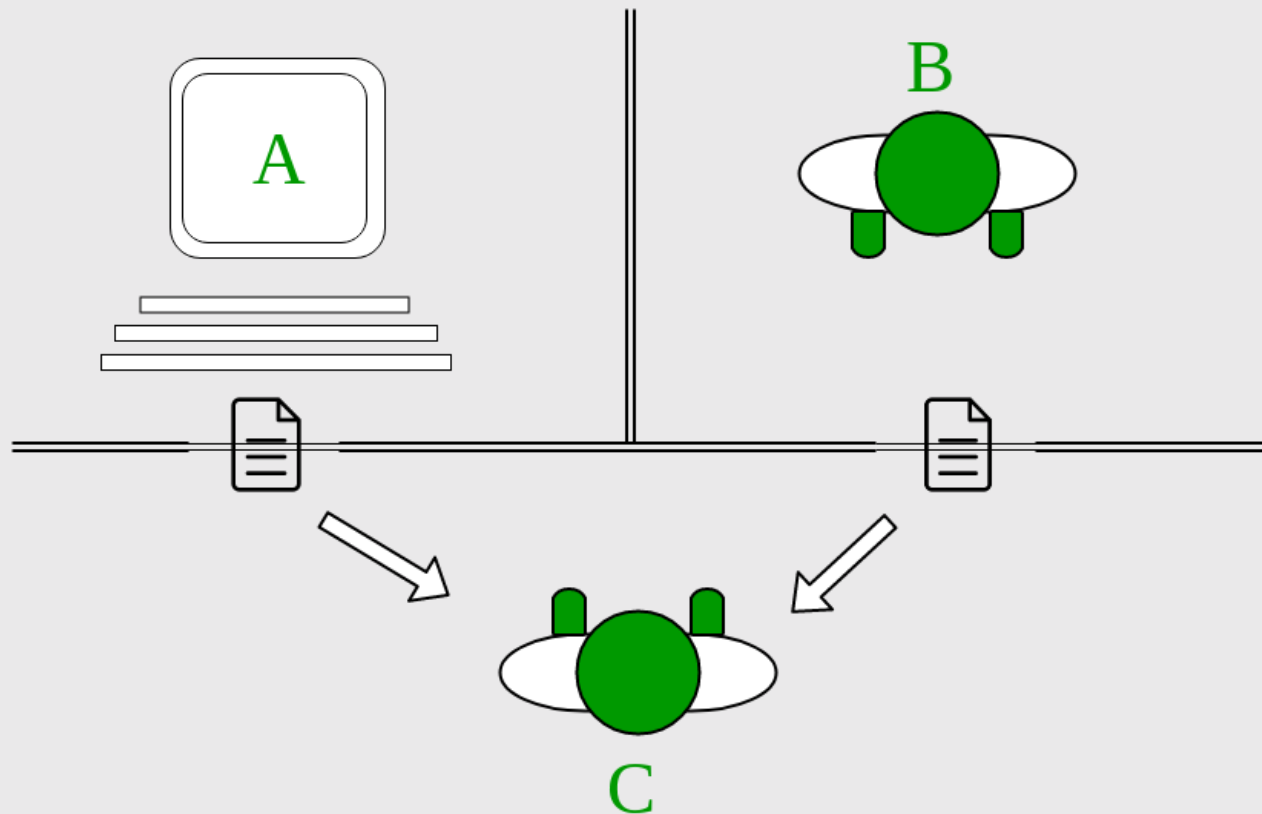
ChatGPT overview

- Approach is fundamentally to **predict the next word**



ChatGPT overview

- Artificial intelligence (AI) has improved enormously over the last 20 years
- But even in 2017, "chatbots" would fail the "Turing Test"



ChatGPT overview

- Artificial intelligence (AI) has improved enormously over the last 20 years
- But even in 2017, "chatbots" would fail the "Turing Test"

TURING TEST EXTRA CREDIT:
CONVINCE THE EXAMINER
THAT HE'S A COMPUTER.

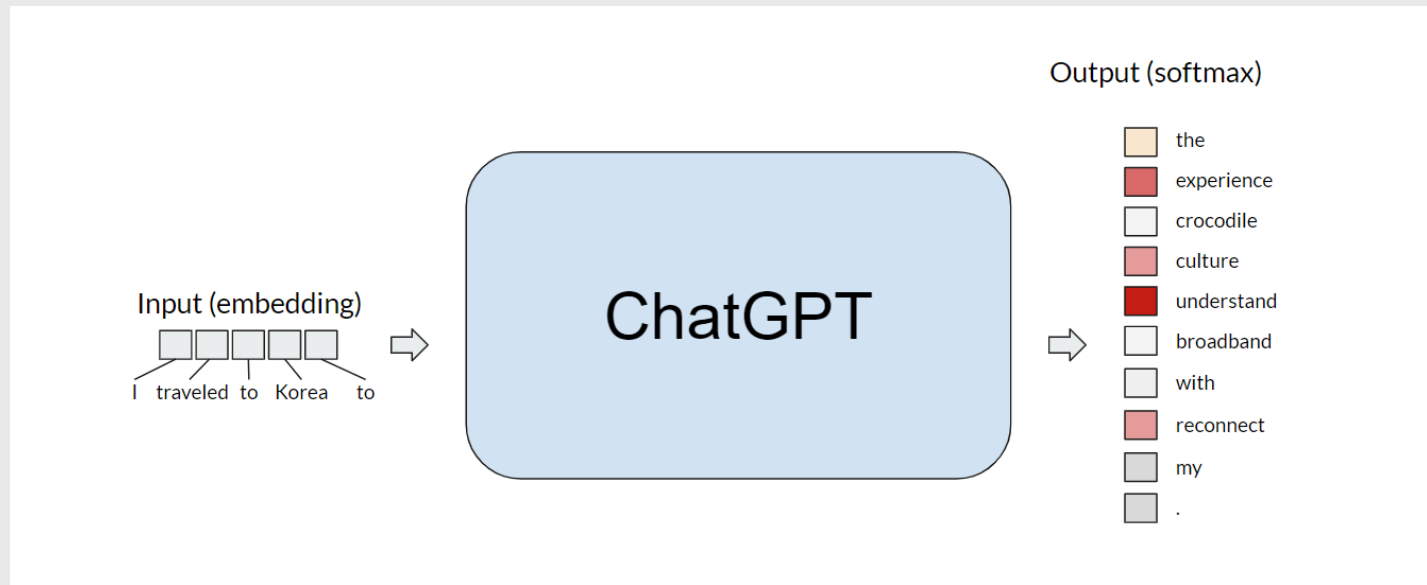
YOU KNOW, YOU MAKE
SOME REALLY GOOD POINTS.
/
I'M ... NOT EVEN SURE
WHO I AM ANYMORE.



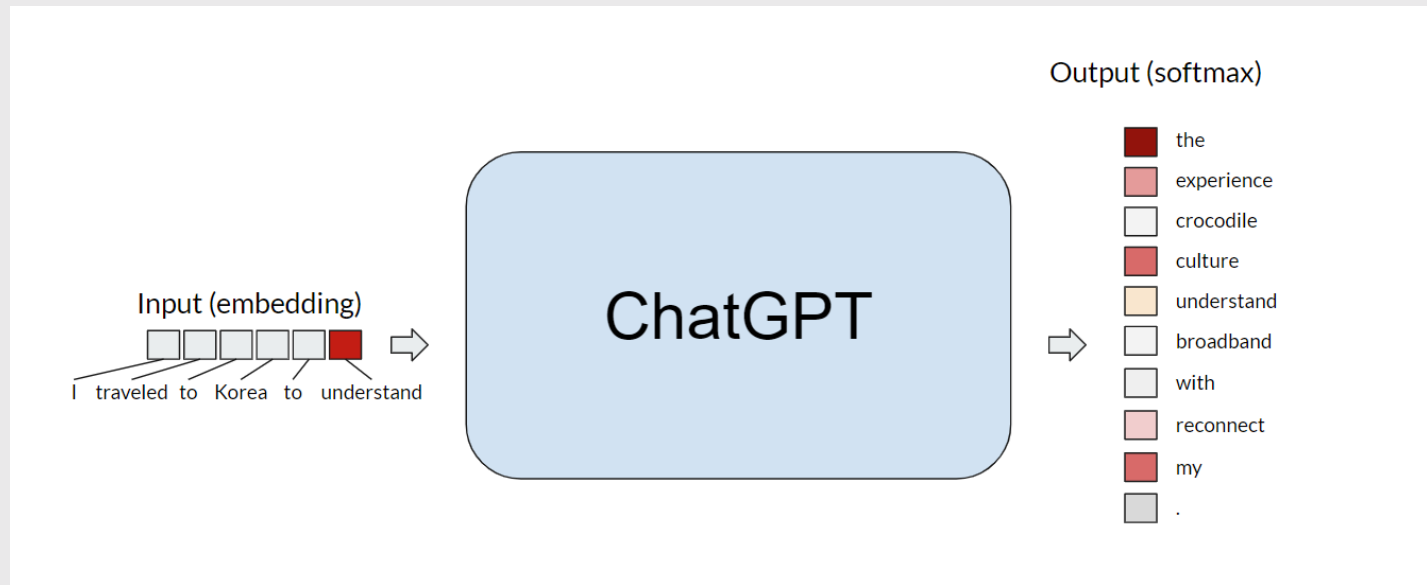
ChatGPT overview

- Innovations in 2017 allowed dramatic improvements in **G**enerative **P**re-trained **T**ransformers (GPT)
 - (Actually, dramatic improvements in all types of transformers)
- Interacting with ChatGPT can feel truly human
 - Passes the Turing Test
- But under the hood, it is still simply predicting the next word
 - Importantly, it is doing so **probabilistically**

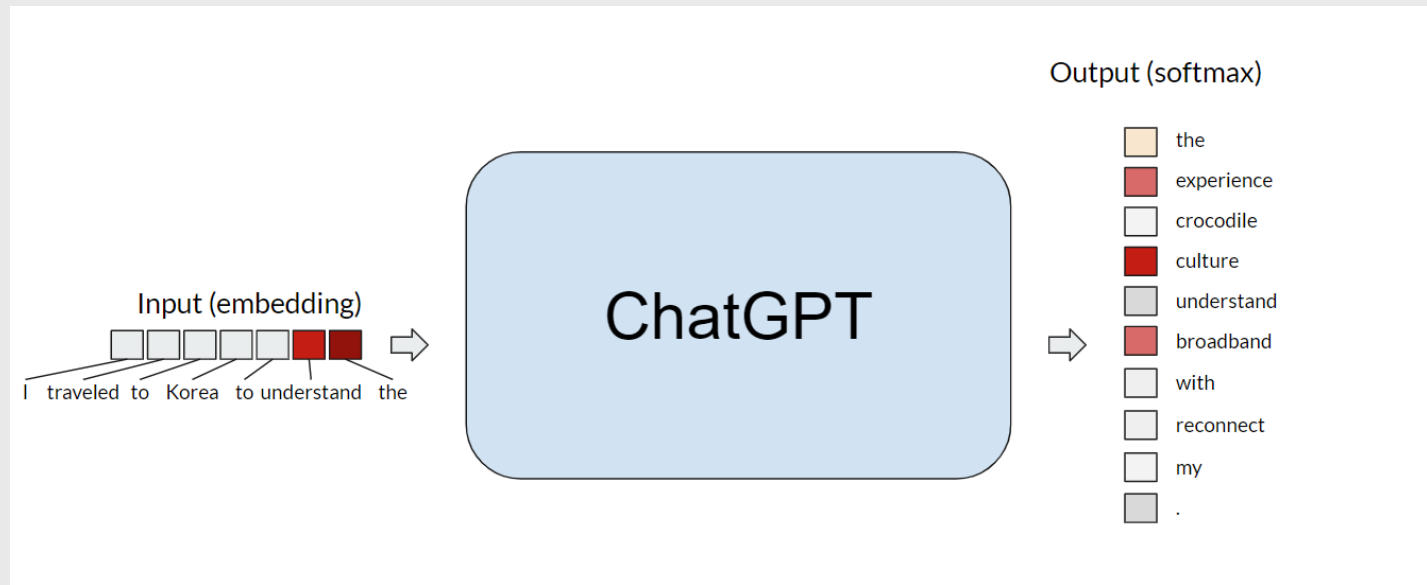
ChatGPT overview



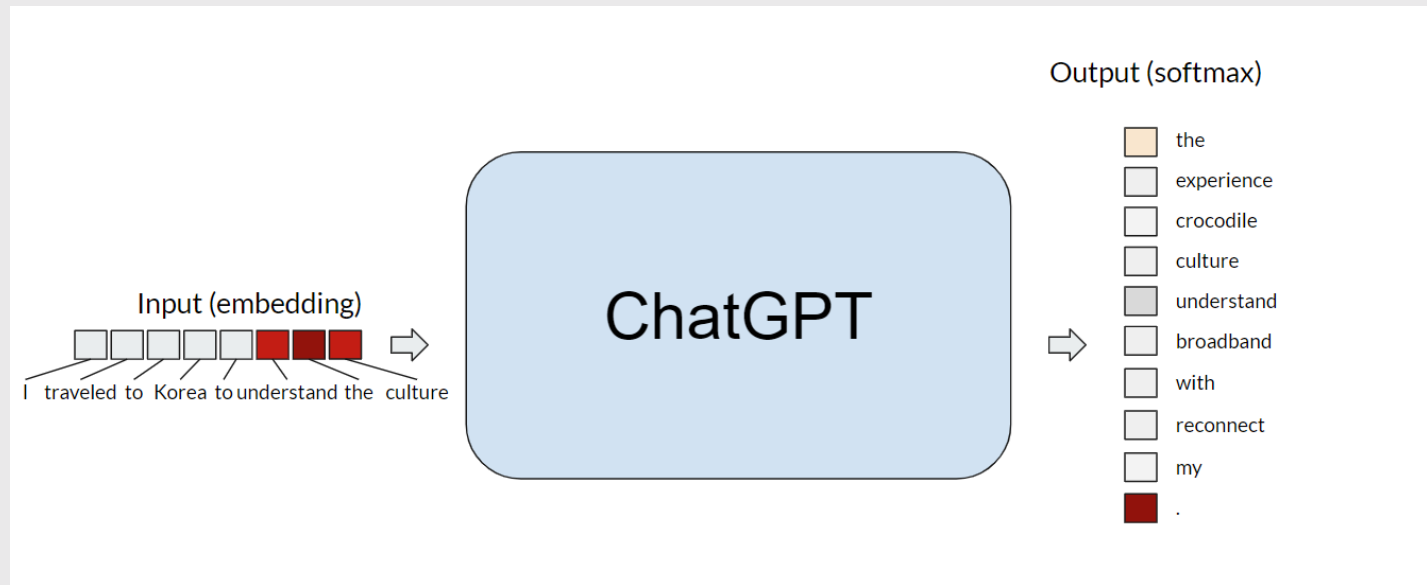
ChatGPT overview



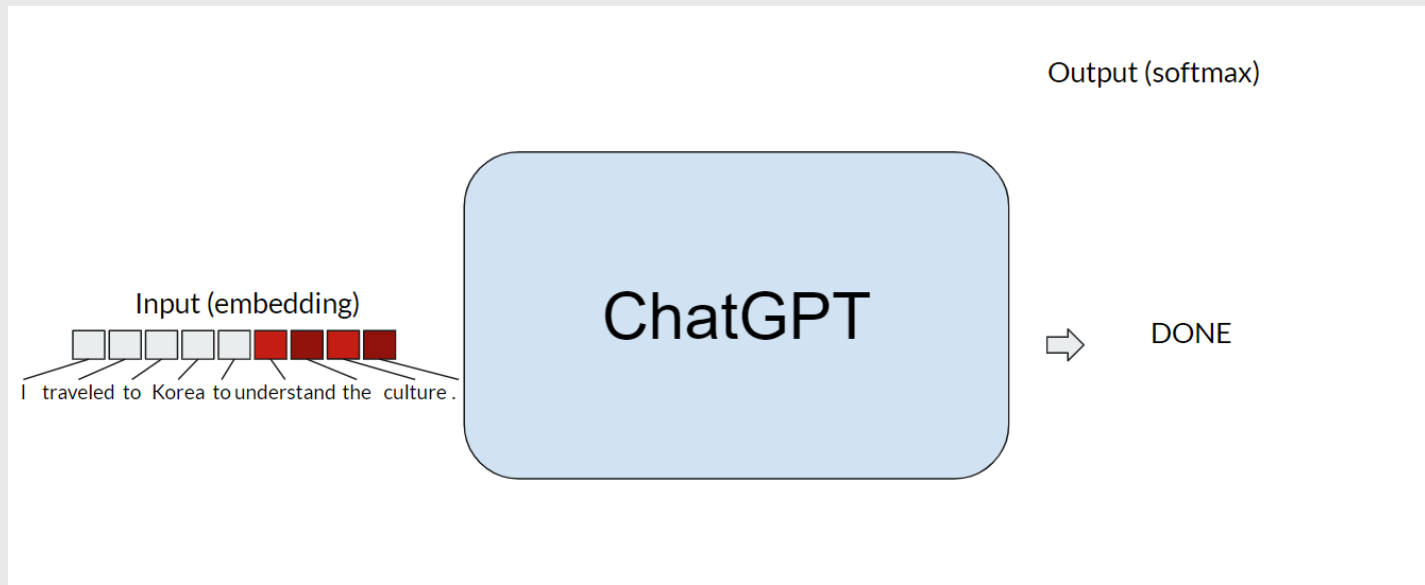
ChatGPT overview



ChatGPT overview



ChatGPT overview



So how can we use it?

- Interactive time!

Conclusion

- Go to Brightspace and take the **2nd** quiz
 - The password to take the quiz is ####
- Homework:
 1. Work through Intro_Data_Science_hw.Rmd
 2. Complete Problem Set 0 (on Brightspace)
 3. Create an OpenAI account (<https://auth0.openai.com/u/signup/>)