

IS813: Gen AI - Implementation

Mohannad Elhamod

Welcome!

What is this class about?

ChatGPT... Absolute Terror!

The Washington Post

TECH Artificial Intelligence Help Desk Internet Culture Space Tech Policy


INNOVATIONS

ChatGPT took their jobs. Now they want to be dogs and fix air conditioners.

Technology used to automate dirty and repetitive jobs. Now, artificial intelligence chatbots are coming after the jobs of humans.

By Pranshu Verma and Gerrit De Vynck

June 2, 2023 at 6:00 a.m. EDT





Forbes

FORBES > LEADERSHIP > EDUCATION

ChatGPT Will Not Replace Teachers But Could Act As A Teaching Assistant




Ray Ravaglia Contributor

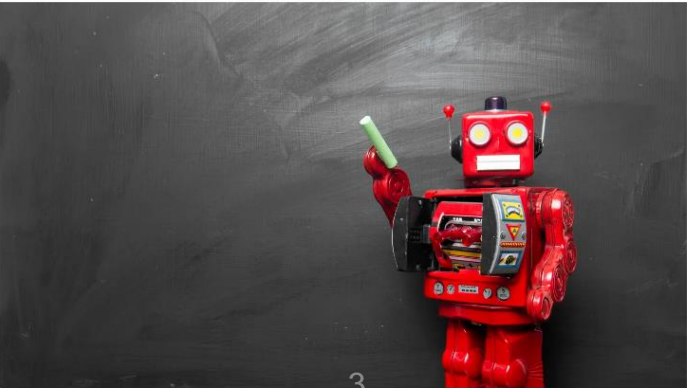
I write about education at the intersection of tech and innovation.

0

Mar 31, 2023, 01:15pm EDT

Listen to article 4 minutes





While Not Replacing Teachers Anytime Soon, GPTs Can Still Be Useful in the Classroom GETTY

ardian

with \$5 per month

The Guardian

on Sport Culture Lifestyle More

tics Business Tech Science Newsletters Fight for democracy

ChatGPT said I did not exist': how artists and writers are fighting back against AI



Forbes

3 > INNOVATION > ENTERPRISE TECH

Will ChatGPT Put Data Analysts Out Of Work?

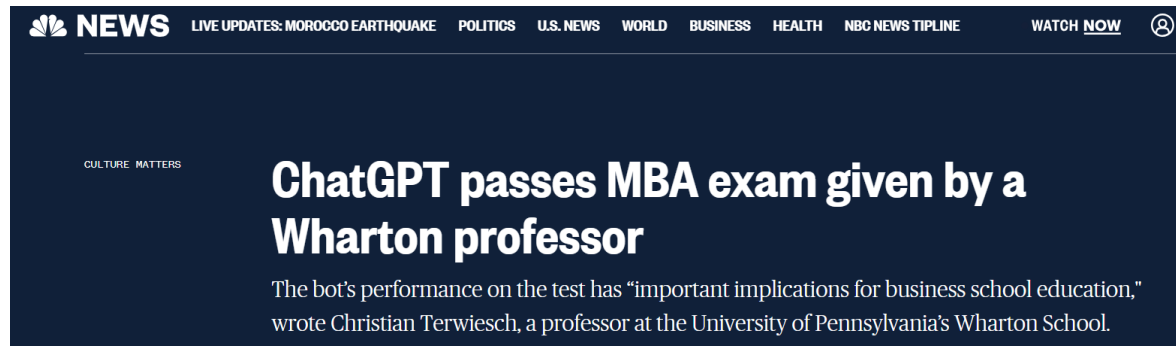
ard Marr Contributor

Follow

3

Feb 7, 2023, 01:40am EST


ChatGPT can do anything!





The New York Times
When Doctors Use a Chatbot to Improve Their Bedside Manner
Despite the drawbacks of turning to artificial intelligence in medicine, some physicians find that ChatGPT improves their ability to communicate empathetically with patients.





ChatGPT is amazing!


 Search


 Home


 My Network


 Jobs

 Messaging

 Notifications

 Me

 For Business

 Learning

Create your own newsletter

Start your own discussion with a newsletter on LinkedIn. Share what you know and build your thought leadership with every new edition.

Try it out





How ChatGPT Can Help You Ace Your Next Interview

 OLLMOO
7,453 followers

+ Follow


April 11, 2023


 


INSIDER


Newsletters Log in [Subscribe](#)


US MARKETS OPEN
In the news


 Dow Jones +0.13%

 Nasdaq +0.4%

 S&P 500 +0.42%

 META +1.98%

 TSLA +6.68%

 BABA -1.69%

HOME > TECH

Programmers are pumped by the rise of ChatGPT, because it makes their jobs easier and helps people to find a lucrative career in tech

Emilia David Mar 3, 2023, 5:00 AM EST


 Search Medium


Replace Grammarly Premium with OpenAI ChatGPT

How to use OpenAI's ChatGPT to replace Grammarly Premium

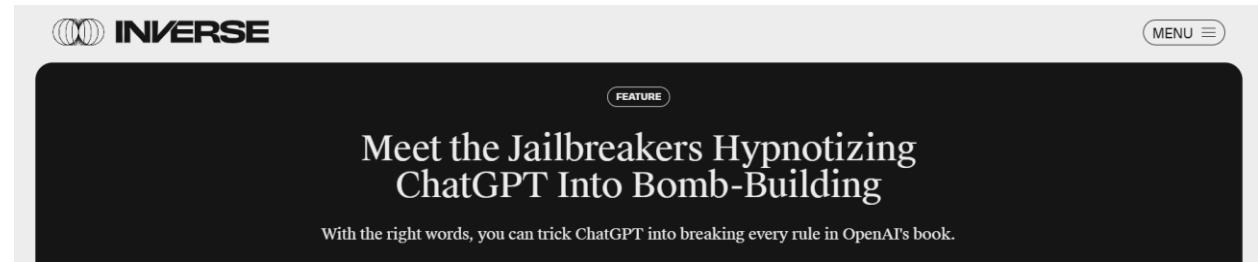
 Sung Kim · Follow

Published in Geek Culture · 9 min read · Dec 19, 2022

 420  7

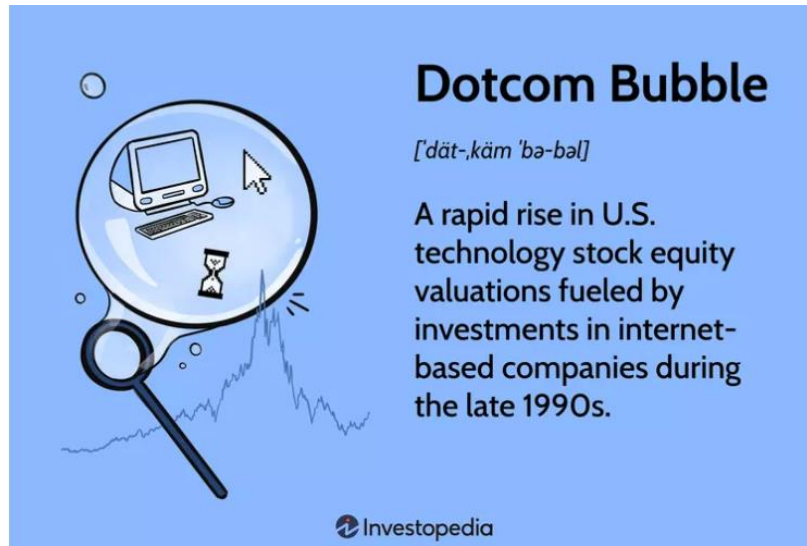
ChatGPT is horrible!



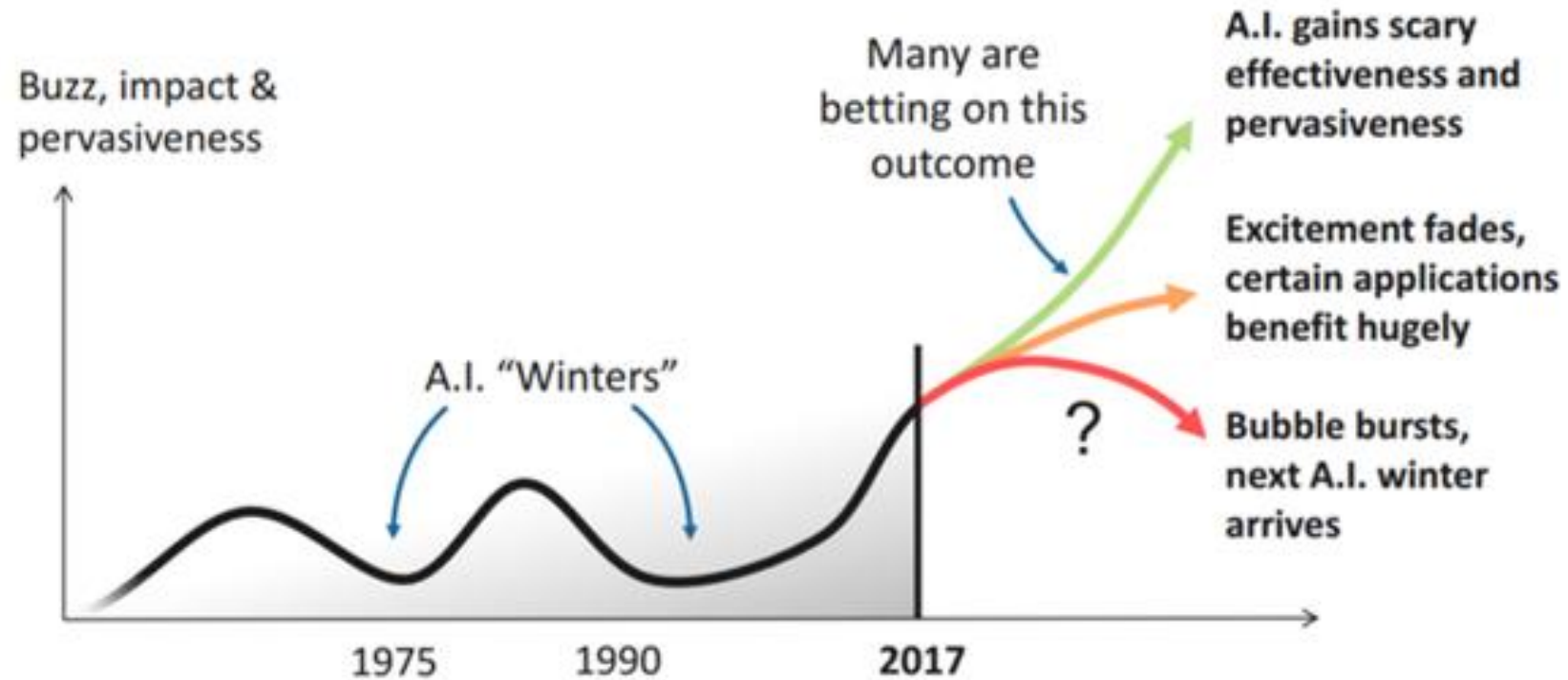
ChatGPT.. What is it really about though?

Let's play a bit!

Is this a hype?



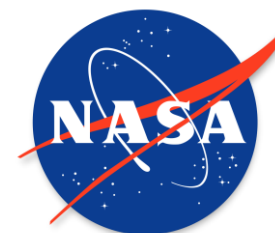
Is this a hype?



Borislav Agapiev

General Guidelines

Who is your professor?



Why are you here?

- I am sure it eventually leads to money....
- Maybe it is just a required class?
- Nonetheless, understanding how things work is essential for decision making and innovation.



Someone who had begun to [study] geometry asked Euclid, 'What shall I get by learning these things?' Euclid called his slave and said, 'Give him [some money], since he must make gain out of what he learns'.

(Heath, 1981, loc. 8625)



Euclid

Fundamentals are important!

- The more foundational knowledge you skip, the more fundamental errors you will make.
- Work hard. **Be patient!**



Zen Speaks: Shouts of Nothingness

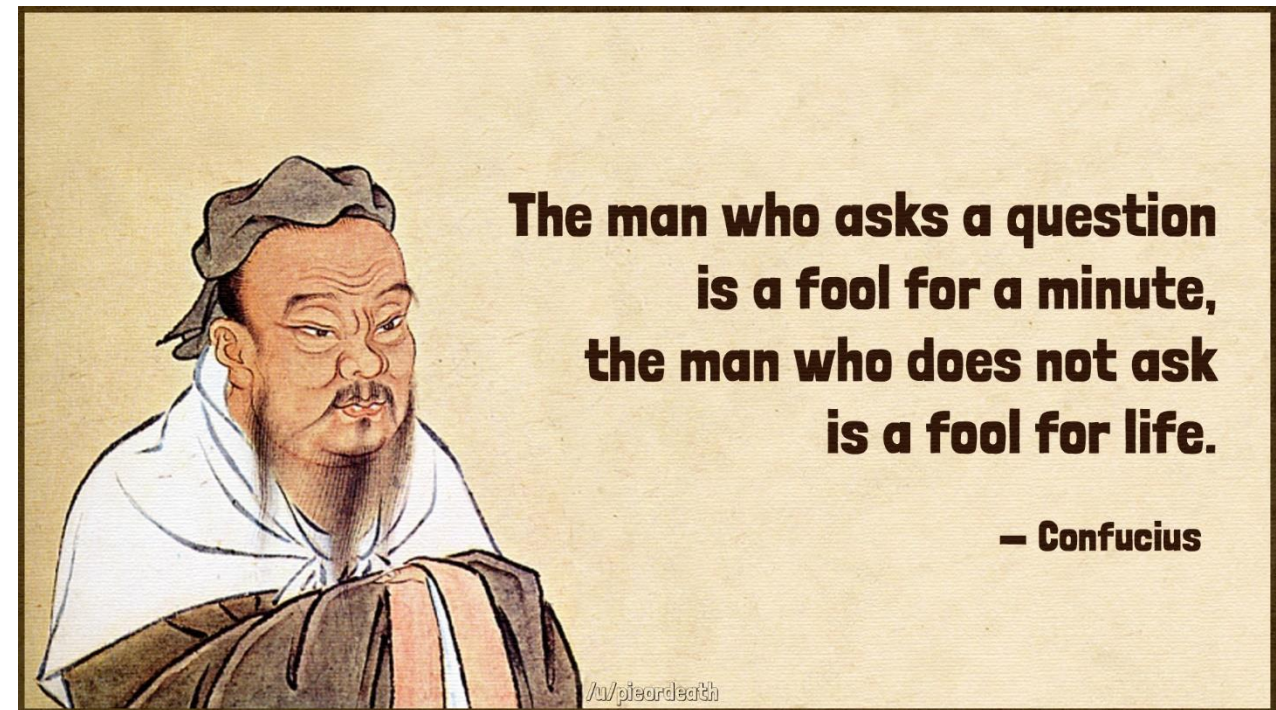
Your professor is not a God

- Deep Learning and Gen AI are fast-growing domains.
- The internet nowadays has all kinds of learning material.
- Your professor is **NOT** here as a walking encyclopedia. He is here to guide your learning experience and build you a solid foundation, so you could continue learning on your own later.



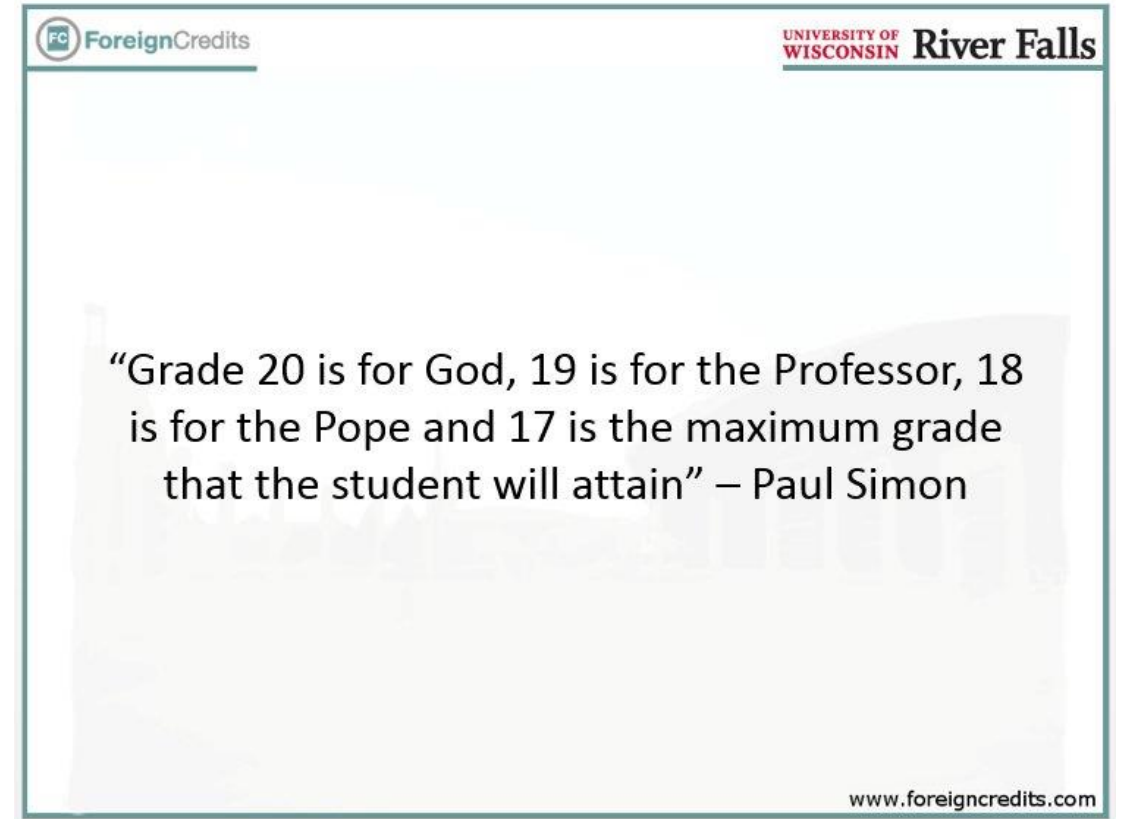
No Question is Foolish

- Other might have the same question.
- Knowledge is hierarchical.



Can I get an A?

- [Questrom policy](#)
- But you will get your fair chance.



Participation is Essential!

- 20% of your grade!
- The instructor reserves the right to cold-call.



Office Hours

- They are for you to take advantage of!
- However, to make the best of your and the TA's/instructor's time, do your homework before dropping in:
 - If you have a question about your project, make sure you have synched with your colleague in advance.
 - If you have question about assignments, make sure you have done your best and that your question is specific rather than “How do I solve this?”.



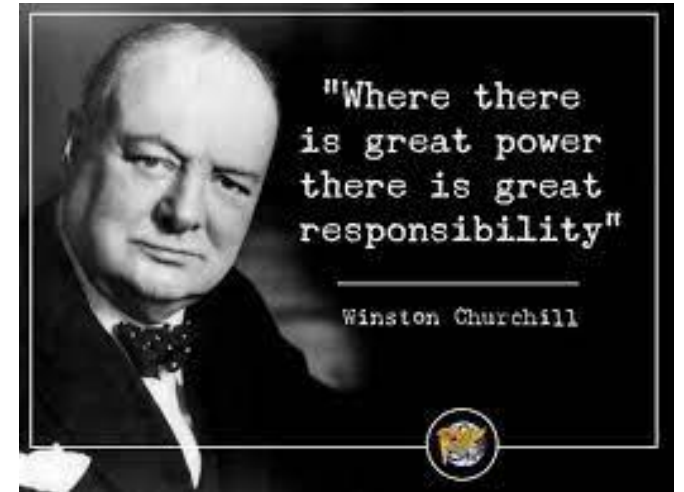
Pass through the Syllabus

- Have you actually read it?!
- Pay attention to Piazza for announcements!

General Objectives

Assignments are meant to simulate challenges in real-life.

- Follow instructions regarding whether ChatGPT is allowed for each assignment.
- You can also discuss with others.
- You **MUST** cite your resources!
- You cannot have someone do the work for you though.
 - No copy-paste of others' solutions.
 - Deviating from instructions leads to penalties.
 - **You must own and understand your work!**
 - **Zero-tolerance for cheating!**



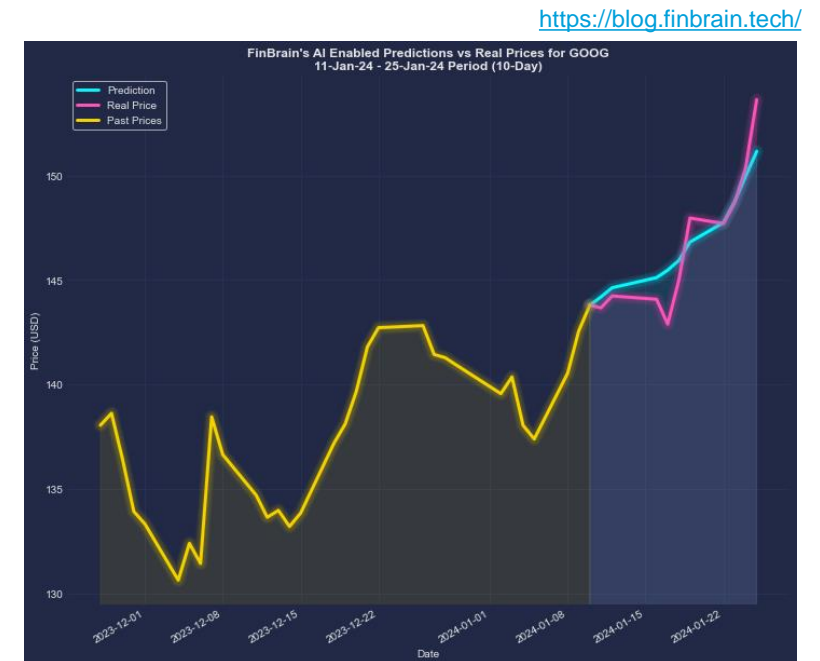
Intro to ML

What is Machine Learning?

- What challenges you find with this approach?

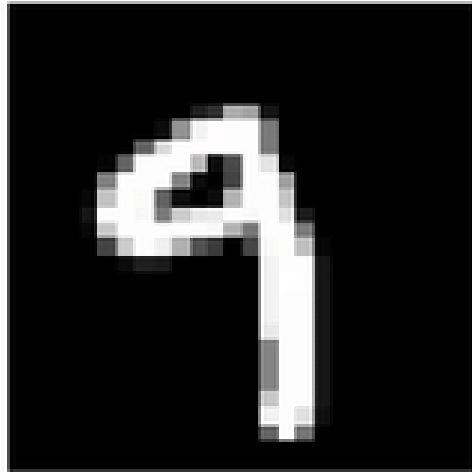


Knowledge-based modeling

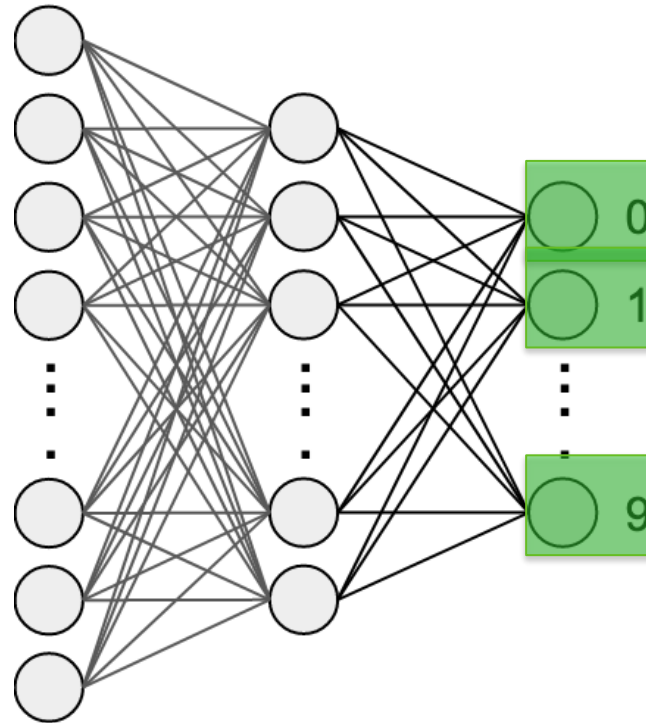


Data-driven modeling

Data-driven Modeling

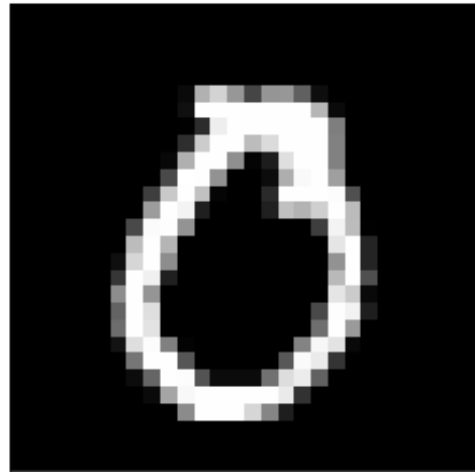


MNIST Dataset

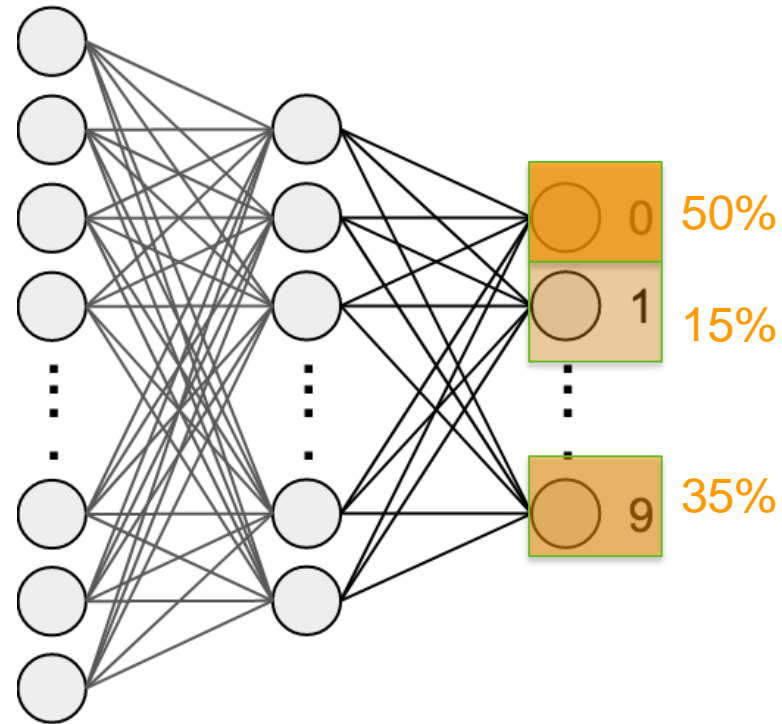


Phase 1: Training

Data-driven Modeling

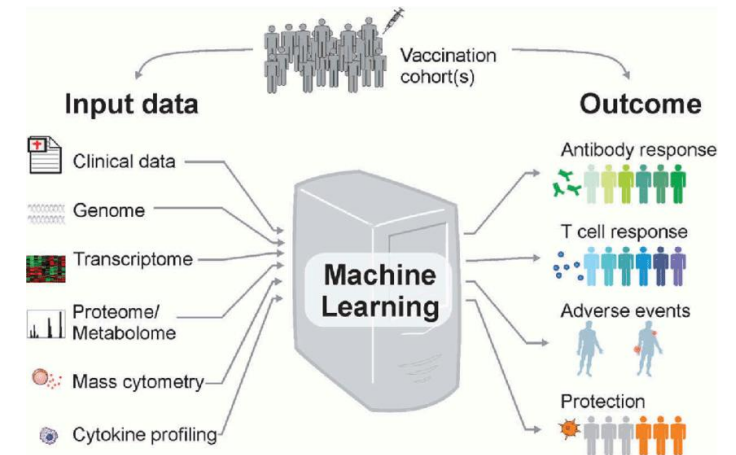
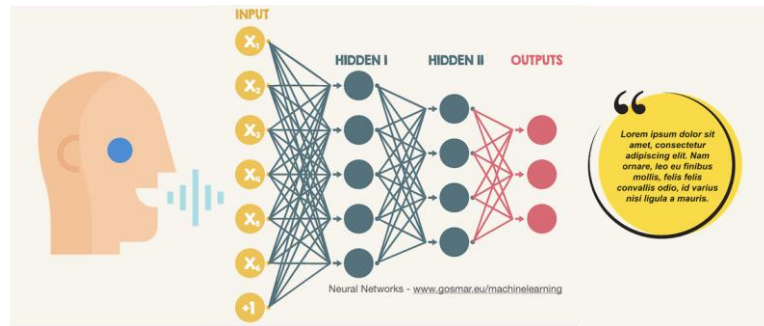


MNIST Dataset



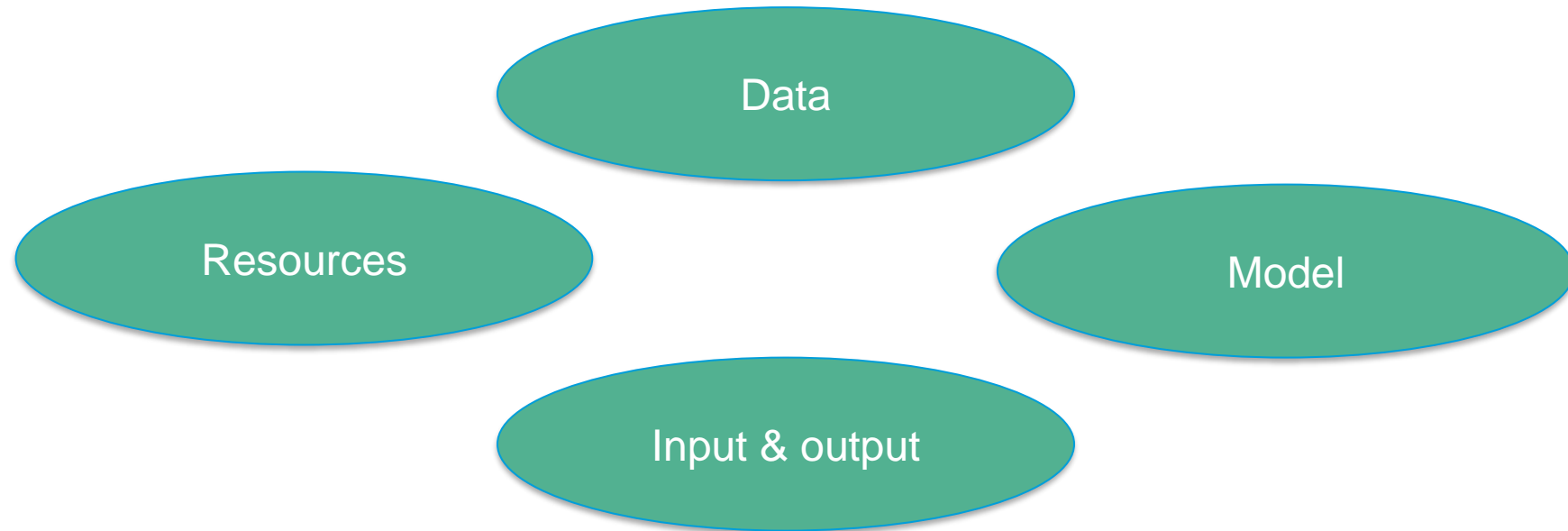
Phase 2: Validation/Testing

Applications



- <https://www.gosmar.eu/machinelearning/2020/05/25/neural-networks-and-speech-recognition/>
- <https://www.ufrjnautilus.com/post/vis%C3%A3o-computacional-e-carros-aut%C3%B4nomos>
- <https://www.nature.com/articles/d41586-019-03298-6>
- <https://www.forbes.com/sites/johnkoetsier/2023/04/14/generative-ai-music-platform-creates-forever-songs-with-artists-unique-sounds-melodies-and-beats/?sh=75a691eada47>
- Gonzalez-Dias, Patricia & Lee, Eva & Sorgi, Sara & Lima, Diógenes & Urbanski, Alysso & Silveira, Eduardo & Nakaya, Helder. (2019). Methods for predicting vaccine immunogenicity and reactivity. Human Vaccines & Immunotherapeutics. 16. 1-8. 10.1080/21645515.2019.1697110.

“Pillars” of Using ML



Data

- Is there enough of it?
- Does it need clean-up?



Damaged specimen

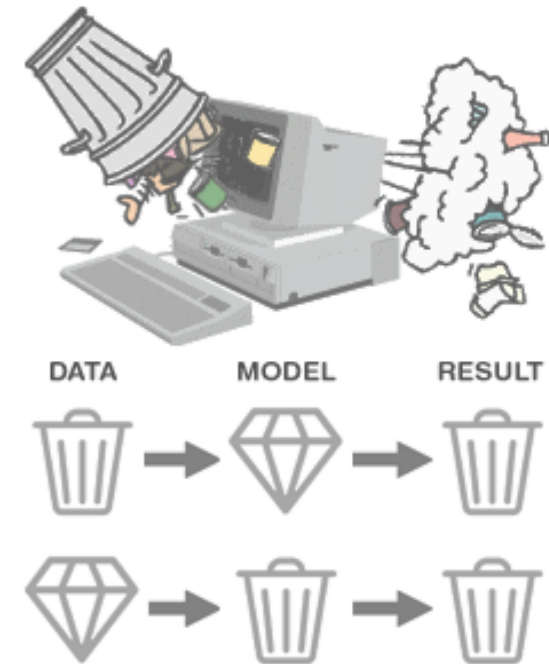


Missing Features



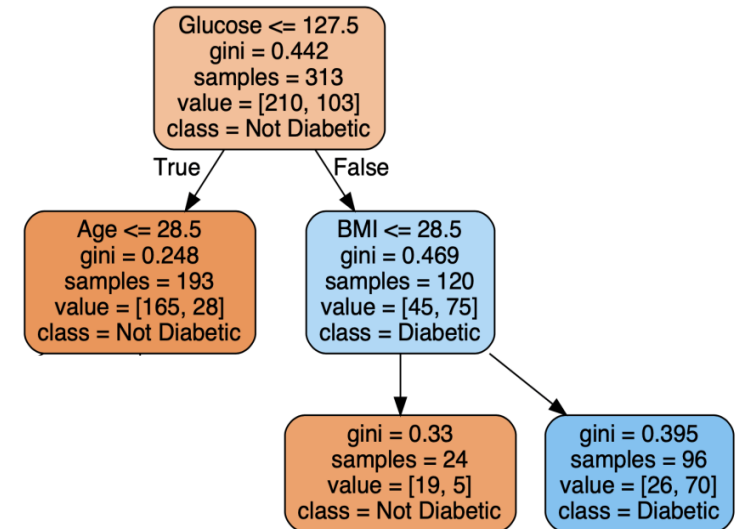
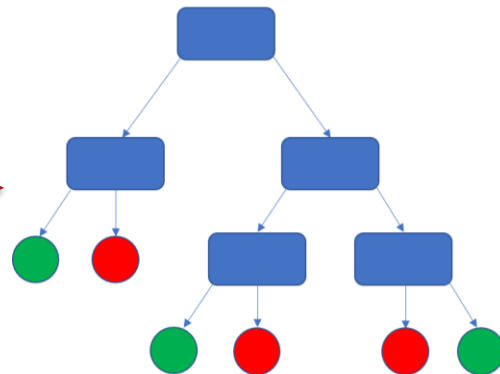
Occluded Features

Elhamod, M., Diamond, K. M., Murat Maga, A., Bakis, Y., Bart, H. L., Mabee, P., Dahdul, W., Leipzig, J., Greenberg, J., Avants, B., & Karpayne, A. (2022). Hierarchy-guided neural network for species classification. *Methods in Ecology and Evolution*, 13, 642–652. <https://doi.org/10.1111/2041-210X.13768>



Model

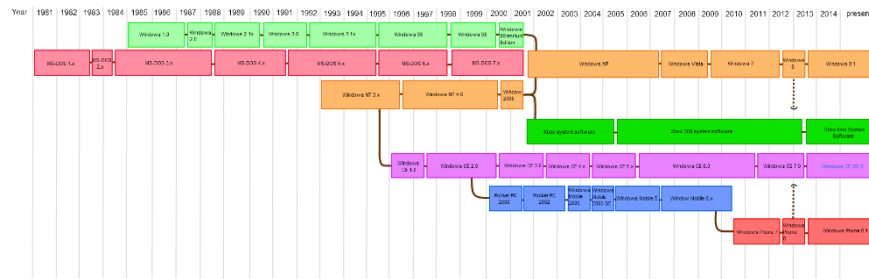
- What kind of model would be sufficient/suitable for modeling your data?



<https://statisticallyrelevant.com/decision-trees-in-python-predicting-diabetes/>

Resources

- What compute resources are available?
- What is the project's timeline?



In-Class Work

Intro to ML

Continued...

Model Generalization

- You build a model that *predicts a movie's box office revenue*. You have only a few movies to train on and for each movie you collect many features, including *whether the of the president at the time of release is a democrat*.
 - Since there are only a few movies for the model to train from, there is a chance that some noise exists.
 - **This is called overfitting!**

Model Generalization

- Now, you build another model that predicts a movie's box office revenue. You have lots of movies to train on. However, for each movie you only collect one feature: month of release.
 - While the month of release may have some signal, it is insufficient to provide an accurate prediction of the box office revenue.
 - **This is called underfitting!**

Model Generalization



<https://www.machinecurve.com/index.php/2020/11/16/how-to-easily-create-a-train-test-split-for-your-machine-learning-model/>

Model Generalization

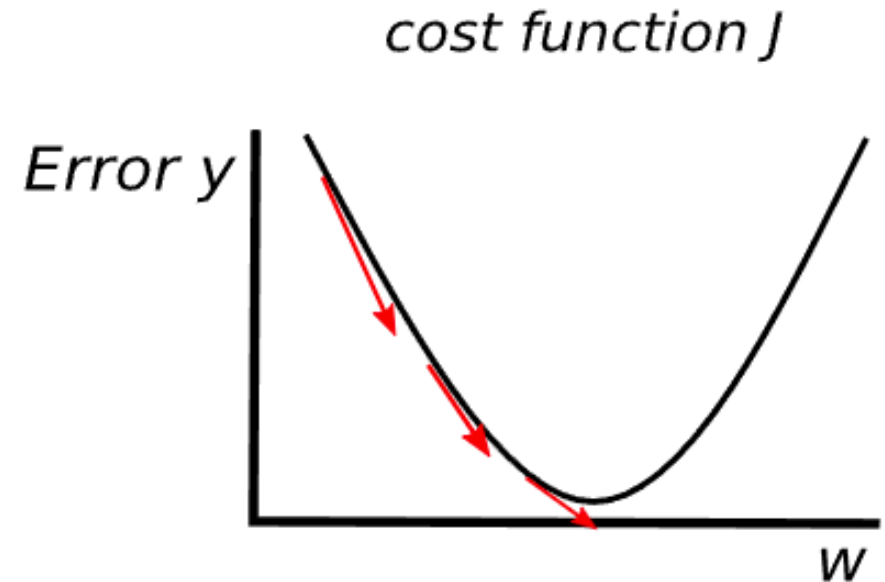


<https://medium.com/trusted-ai/explaining-ai-model-behaviour-with-ibm-watson-openscale-86515702c177>

Neural Networks

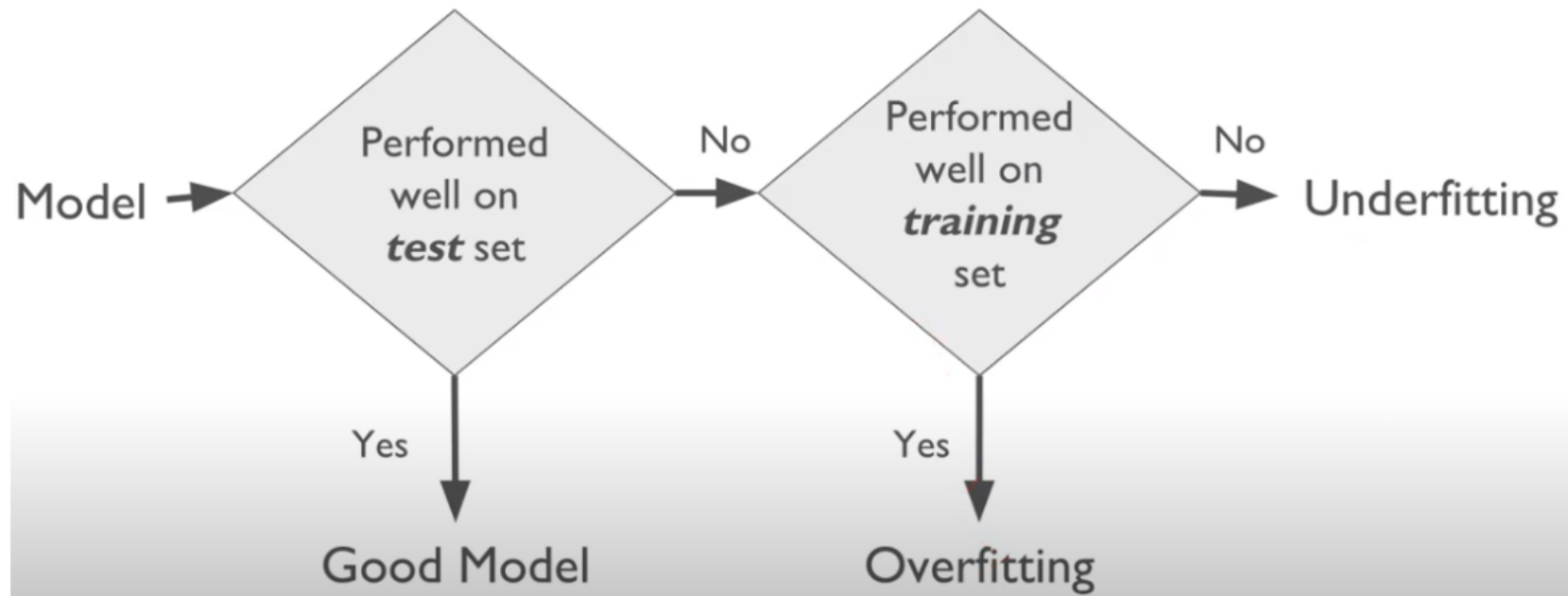
In a nutshell...

- A neural network has “weights” (or “parameters”).
- We want to assign these weights the values that lead to the lowest error.
 - Error \equiv loss \equiv cost function
 - Generally using gradient descent with backpropagation.



Elvira Siegel

Overfitting and Underfitting



<http://jcsites.juniata.edu/faculty/rhodes/ml/clusterAn.htm>

Are the results bad?

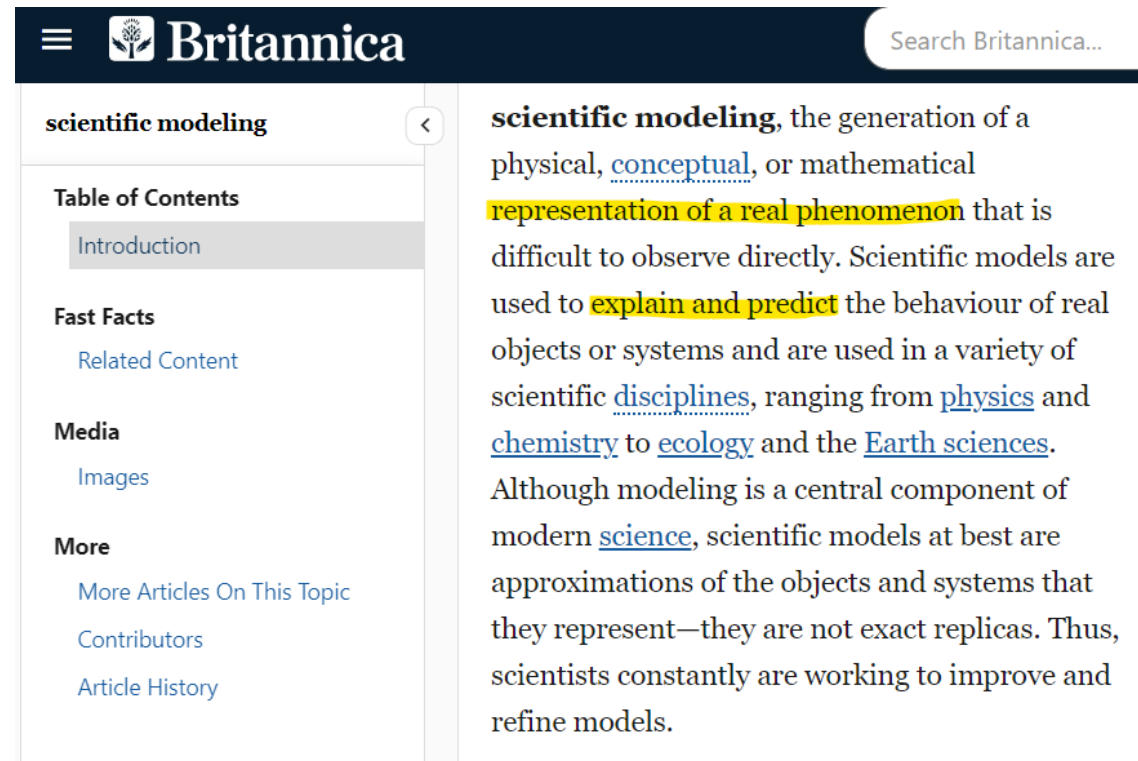
- Check against a benchmark!
 - paperswithcode.com
 - kaggle.com
 - **huggingface.com

How do I improve my results?

- Best way: Get more GOOD data
 - If not, clean-up existing data.
- Are you overfitting or underfitting?
 - Overfitting: get more data or use a less complex model.
 - Underfitting: get a more complex model.
- Keep it simple!
 - Start with a simple model, simple data, simple code.
 - Test by component
 - Test by example

Language Modeling

What is a model?



The screenshot shows the Britannica website interface for the article "scientific modeling". The left sidebar contains a "Table of Contents" with "Introduction" selected, and sections for "Fast Facts", "Media", and "More" with links to "Related Content", "Images", "More Articles On This Topic", "Contributors", and "Article History". The main content area defines scientific modeling as the generation of a physical, conceptual, or mathematical representation of a real phenomenon that is difficult to observe directly. It states that scientific models are used to explain and predict the behavior of real objects or systems and are used in a variety of scientific disciplines, ranging from physics and chemistry to ecology and the Earth sciences. It concludes that although modeling is a central component of modern science, scientific models at best are approximations of the objects and systems that they represent—they are not exact replicas. Thus, scientists constantly are working to improve and refine models.

scientific modeling, the generation of a physical, [conceptual](#), or mathematical [representation of a real phenomenon](#) that is difficult to observe directly. Scientific models are used to [explain and predict](#) the behaviour of real objects or systems and are used in a variety of scientific [disciplines](#), ranging from [physics](#) and [chemistry](#) to [ecology](#) and the [Earth sciences](#). Although modeling is a central component of modern [science](#), scientific models at best are approximations of the objects and systems that they represent—they are not exact replicas. Thus, scientists constantly are working to improve and refine models.

What is language modeling?...

Web search engine / ...

I saw a cat|

I saw a cat on the chair

I saw a cat running after a dog

I saw a cat in my dream

I saw a cat book

[Lena-votta](#)

Send

▼

To

Cc

Add a subject

Greetings,

I would like to

Tab

Mohannad Elhamod
Clinical Assistant Professor
Boston University | Questrom School of Business
elhamod@bu.edu

QUESTROM
MEANS
BUSINESS.

What is language modeling?...

I grabbed the **branch** and broke it.

I went to the **branch** and deposited some money.

Context matters!

What is language modeling?...

- I went to ____.
- I woke up at 7 am and went to ____.
- I woke up at 7 am, packed my book and notebook, and went to ____.

The more context, the more certain

What is language modeling?...

I went to the **branch** and deposited some money.

I went to the **bank** and deposited some money.

I went to the **ATM** and deposited some money.

Words which frequently appear in **similar contexts** have **similar meaning**.

[Lena-votta](#)

Natural Language Processing (NLP)

Includes text generation:

- Text completion.
- Text summarization.
- Question answering.

But there are also many other tasks such as Text classification: (e.g., Sentiment analysis, Reviews, Fake news) or word classification.

Formalizing our thoughts

- It seems we process language sequentially**.
- So, language modeling is the chaining of word probabilities.

$P(\text{I saw a cat on } \dots) =$

$P(\text{I}) \cdot P(\text{saw}|\text{I}) \cdot P(\text{a}|\text{I saw}) \cdot P(\text{cat}|\text{I saw a}) \cdot P(\text{on}|\text{I saw a cat}) \cdot \dots$

Probability of I saw a cat on

[Lena-volta](#)

- How do we calculate these probabilities?

counting...

$$P(\text{cat}) = \frac{N(\text{"cat" in corpus})}{N(\text{all words in corpus})}$$

$$P(\text{cat} | \text{my}) = \frac{N(\text{"my cat" in corpus})}{N(\text{"my" in corpus})}$$

Can you foresee any problem with this calculation?...

N-grams

Instead, let's just use a context of *fixed-length*.

$P(\text{I saw a cat on a mat}) =$

- $P(\text{I})$
- $P(\text{saw} \mid \text{I})$
- $P(\text{a} \mid \text{I saw})$
- $P(\text{cat} \mid \text{I saw a})$
- $P(\text{on} \mid \text{I saw a cat})$
- $P(\text{a} \mid \text{I saw a cat on})$
- $P(\text{mat} \mid \text{I saw a cat on a})$

- $n=3$ (trigram model): $P(y_t \mid y_1, \dots, y_{t-1}) = P(y_t \mid y_{t-2}, y_{t-1})$,
- $n=2$ (bigram model): $P(y_t \mid y_1, \dots, y_{t-1}) = P(y_t \mid y_{t-1})$,
- $n=1$ (unigram model): $P(y_t \mid y_1, \dots, y_{t-1}) = P(y_t)$.

[Lena-volta](#)

N-grams

Context is like a sliding window into the past.

Hugging Face is a startup based in New York City and Paris

$p(\text{word})$

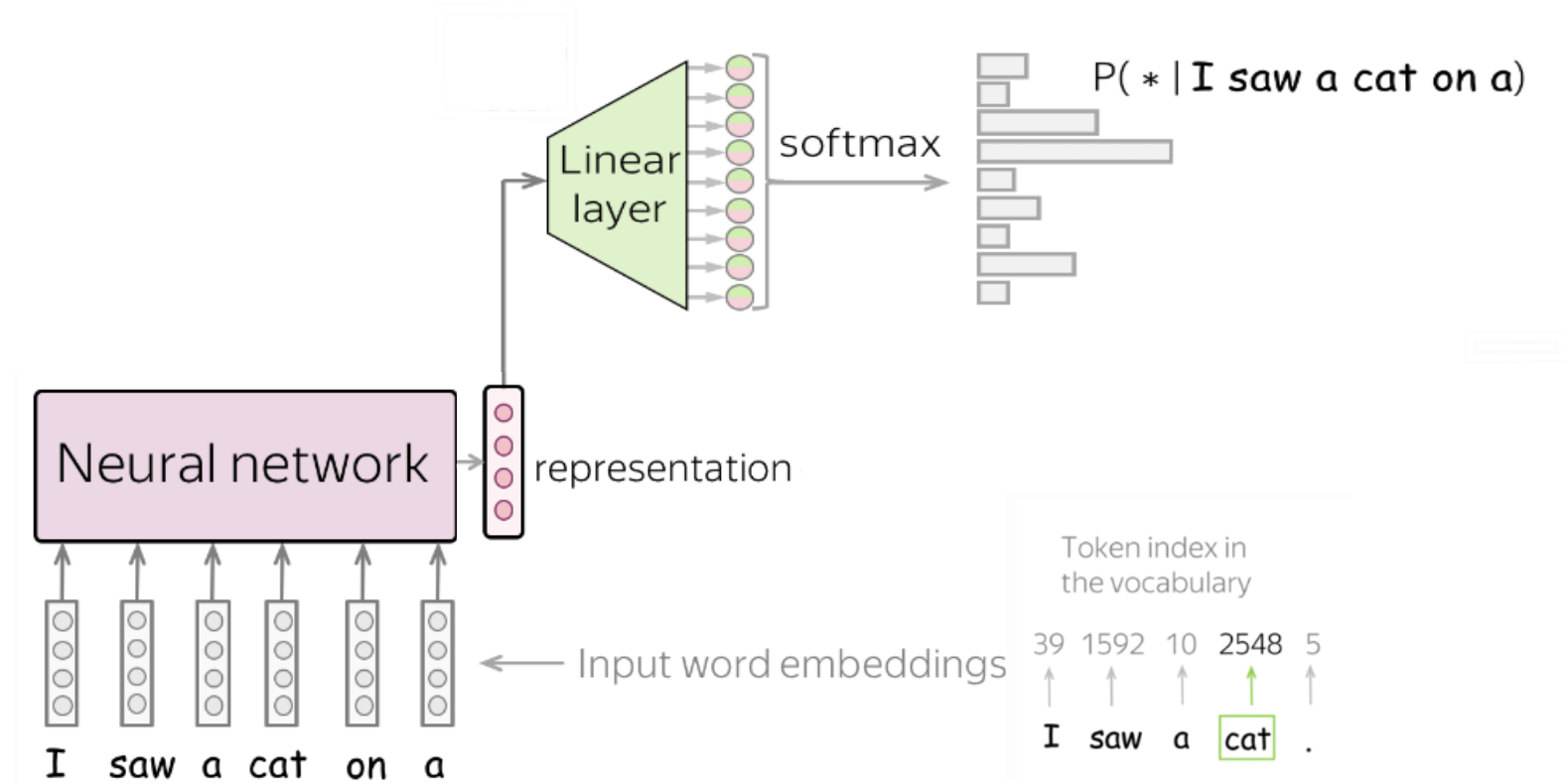
[Huggingface](#)

Context size

- I went to the beach...
 - My wife sat next to me. She was replying to some emails, and...
 - the bird stole our sandwich. Then...
 - it started raining suddenly and ____.
-
- Longer context: predictable outcome.
 - Shorter context: Too unpredictable.

Neural networks for language modeling

General Model Architecture



[Lena-volta](#)

Can you see any issue with inputting words in an NN?

What is an embedding?

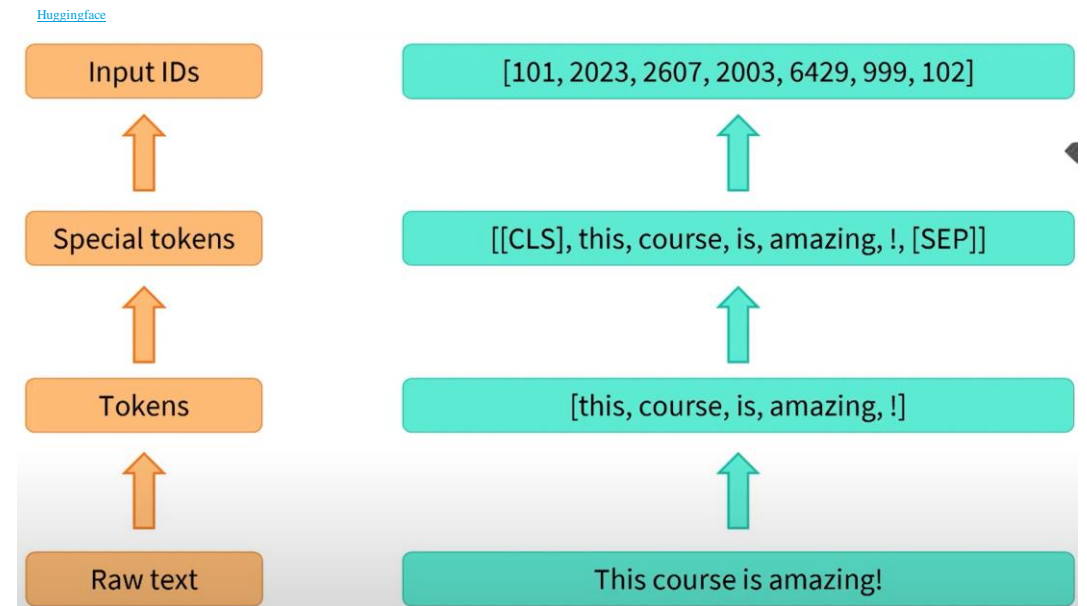
- embeddings = representation = features = latent space.
- It is a representation of your input.
- [Example for images.](#)

Word embeddings

- We ideally want words that have similar meanings to have smaller distances.
- [Demo](#)
- Examples:
 1. [Word2Vec \(Google\)](#)
 2. [GloVe \(Stanford\)](#)
 3. [Train your own!](#)

Tokenization

- Computers only understand numbers.
- We need to convert the text into tokens (e.g., words).



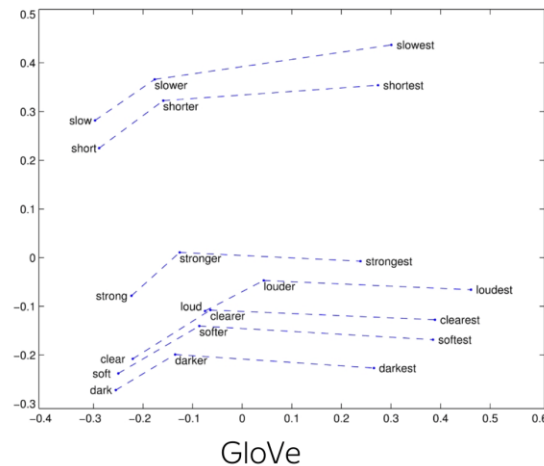
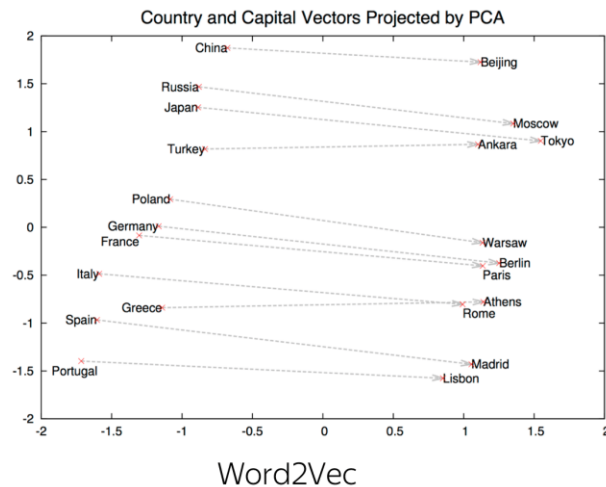
Word embeddings

Word embeddings can also be used to find directionality in the corpus:

- [Demo \(dimensionality\)](#)

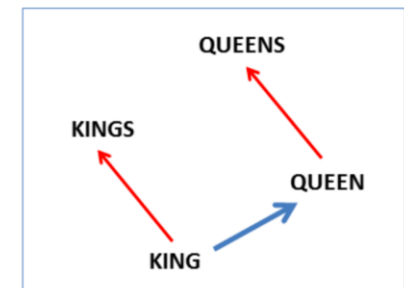
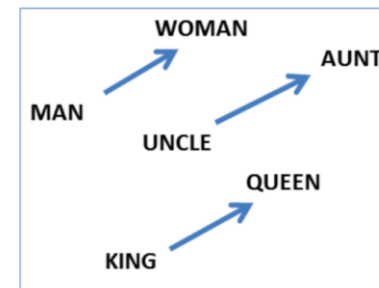
Word embeddings

Word embeddings can also be used find directionality in the corpus.



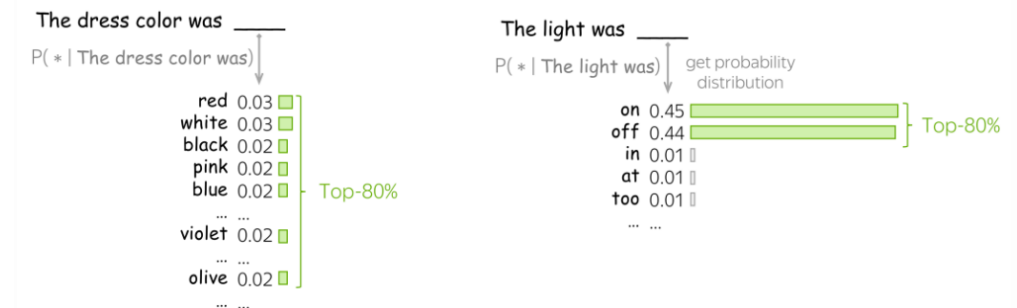
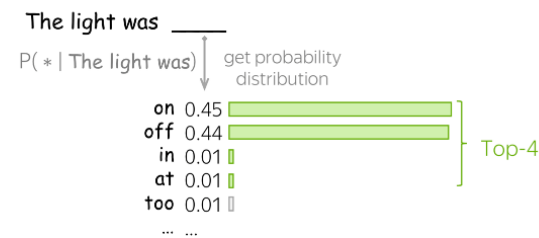
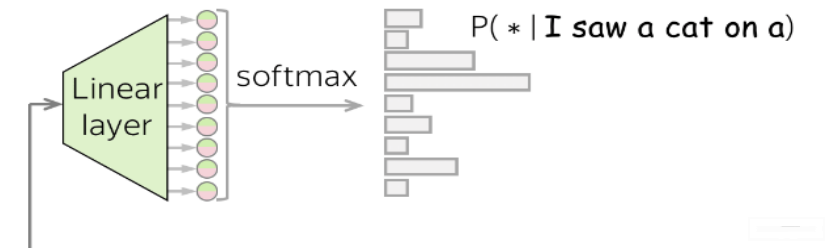
semantic: $v(\text{king}) - v(\text{man}) + v(\text{woman}) \approx v(\text{queen})$

syntactic: $v(\text{kings}) - v(\text{king}) + v(\text{queen}) \approx v(\text{queens})$



Sampling The Distribution

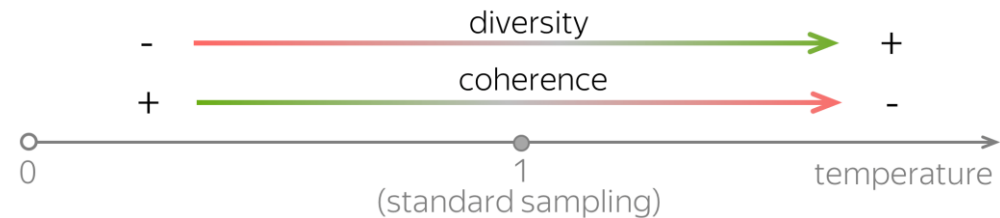
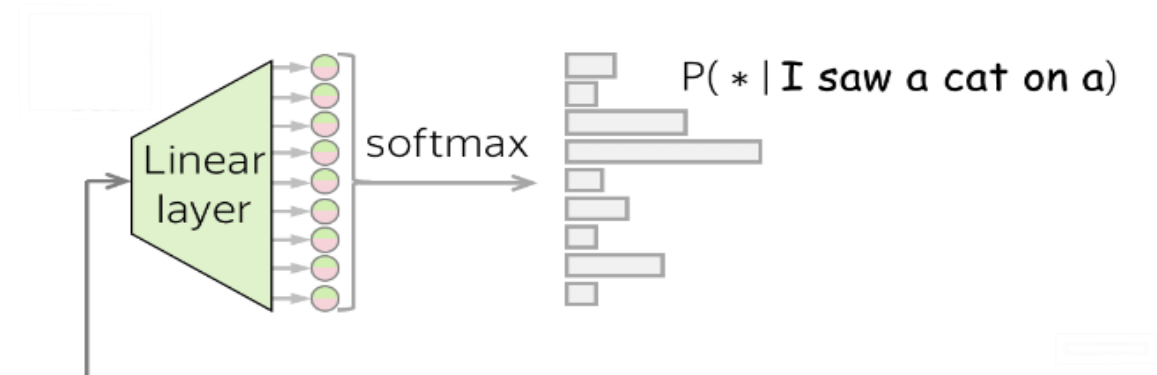
- Always take top probability?
 - That makes the model deterministic (no creativity).
- Alternative?
 - Top-k or top-p.



[Lena-voita](#)

Sampling The Distribution

- Some words have way higher probability than others.
- This can be manually tuned through **temperature**.
- [Demo](#)



[Lena-voita](#)

Measuring The Metric

- What are we looking for?
 - A model that is not surprised by the new text it seen.
- We use perplexity.
 - Takes values between 1 and number of possible tokens.
 - Smaller is better.
 - Demo

Resources

- [Math teachers stage a calculated protest](#)
- [It's time to believe in the AI hype](#)
- [Has the AI bubble burst?](#)
- [Did GPT4 really pass the bar exam?](#)
- [Amazon “AI” stores.](#)
- [Human oversight of Robotaxis.](#)
- [Gemini demo was faked.](#)
- [Video: A student trying to create a business with GenAI.](#)
- [Meaning and calculation of perplexity.](#)
- [Video: LLMs vs The Brain](#)