What is Machine Learning?

"Learning is any process by which a system improves performance from experience."

Herbert Simon

What is Machine Learning?

"Machine learning … gives computers the ability to learn without being explicitly programmed."

Arthur Samuel

What is Machine Learning?

- **Tom Mitchell:** Algorithms that
  - improve their **performance** P
  - at **task** T
  - with **experience** E

- A well-defined machine learning task is given by (P, T, E)

Example: Game Playing

- **Tom Mitchell:** Algorithms that
  - improve their **performance** P
  - at **task** T
  - with **experience** E

- T = playing Checkers
- P = win rate against opponents
- E = playing games against itself

Example: Prediction

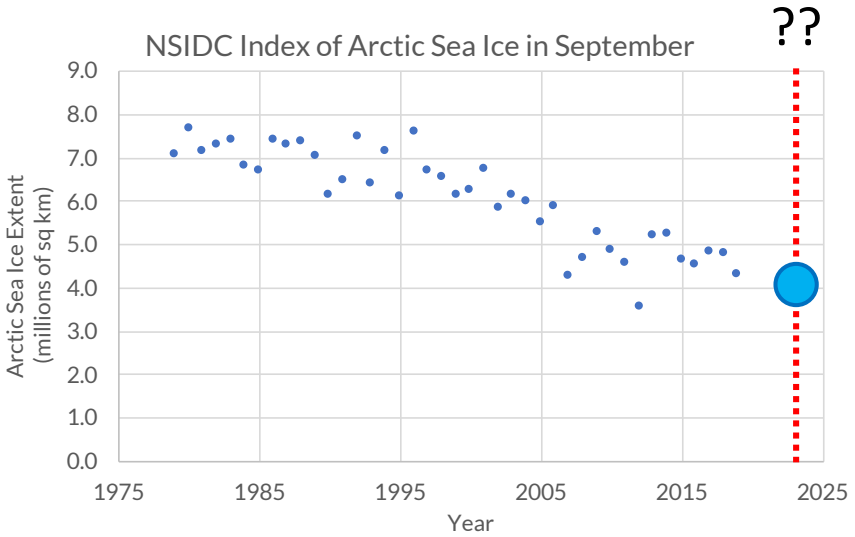


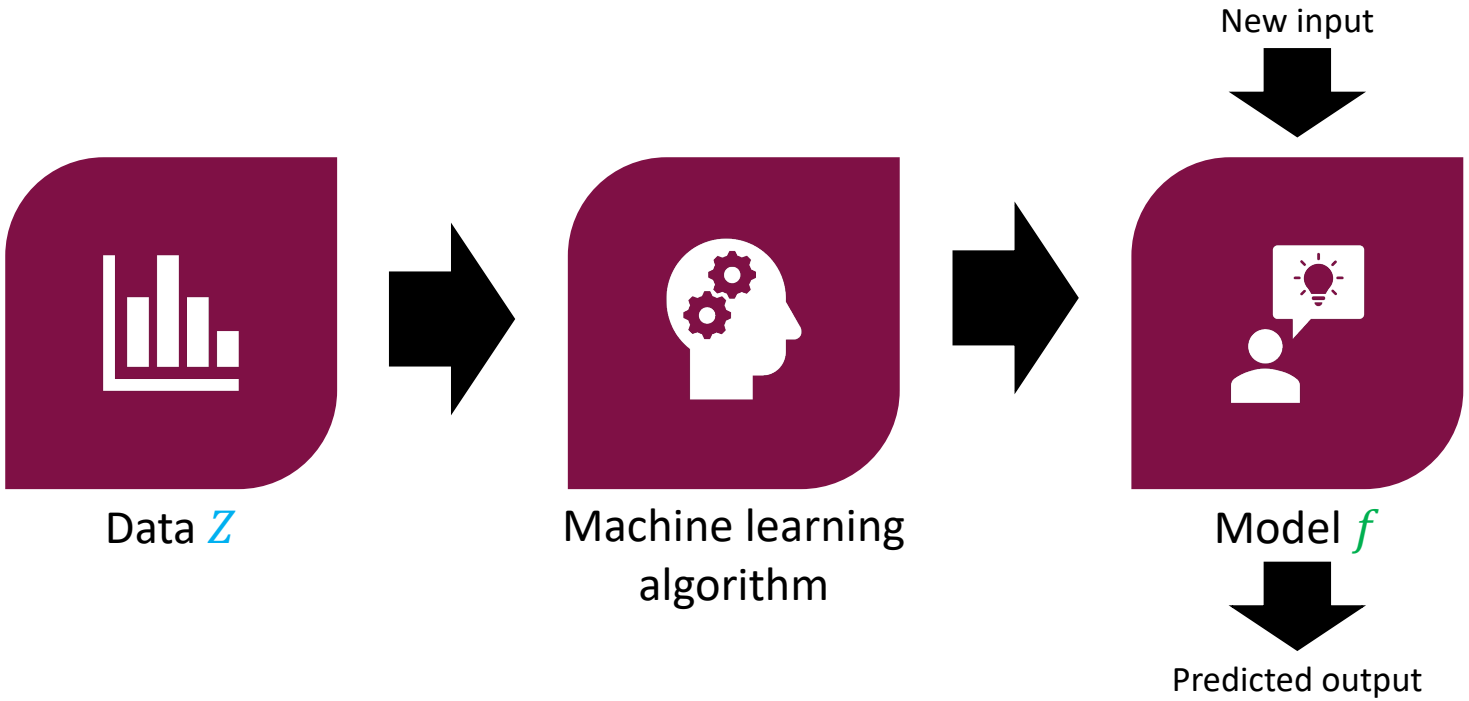
Image: https://www.flickr.com/photos/gsfc/5937599688/
Data from https://nsidc.org/arcticseaicenews/sea-ice-tools/

Example: Prediction

- **Tom Mitchell:** Algorithms that
  - improve their **performance** P
  - at some **task** T
  - with **experience** E

- T = predict Arctic sea ice extent
- P = prediction error (e.g., absolute difference)
- E = historical data

Machine Learning for Prediction



Example: Prediction

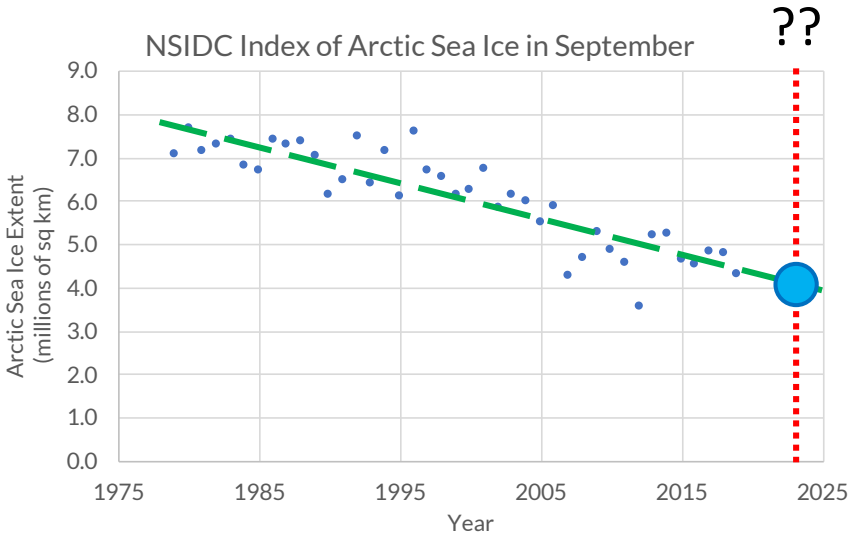
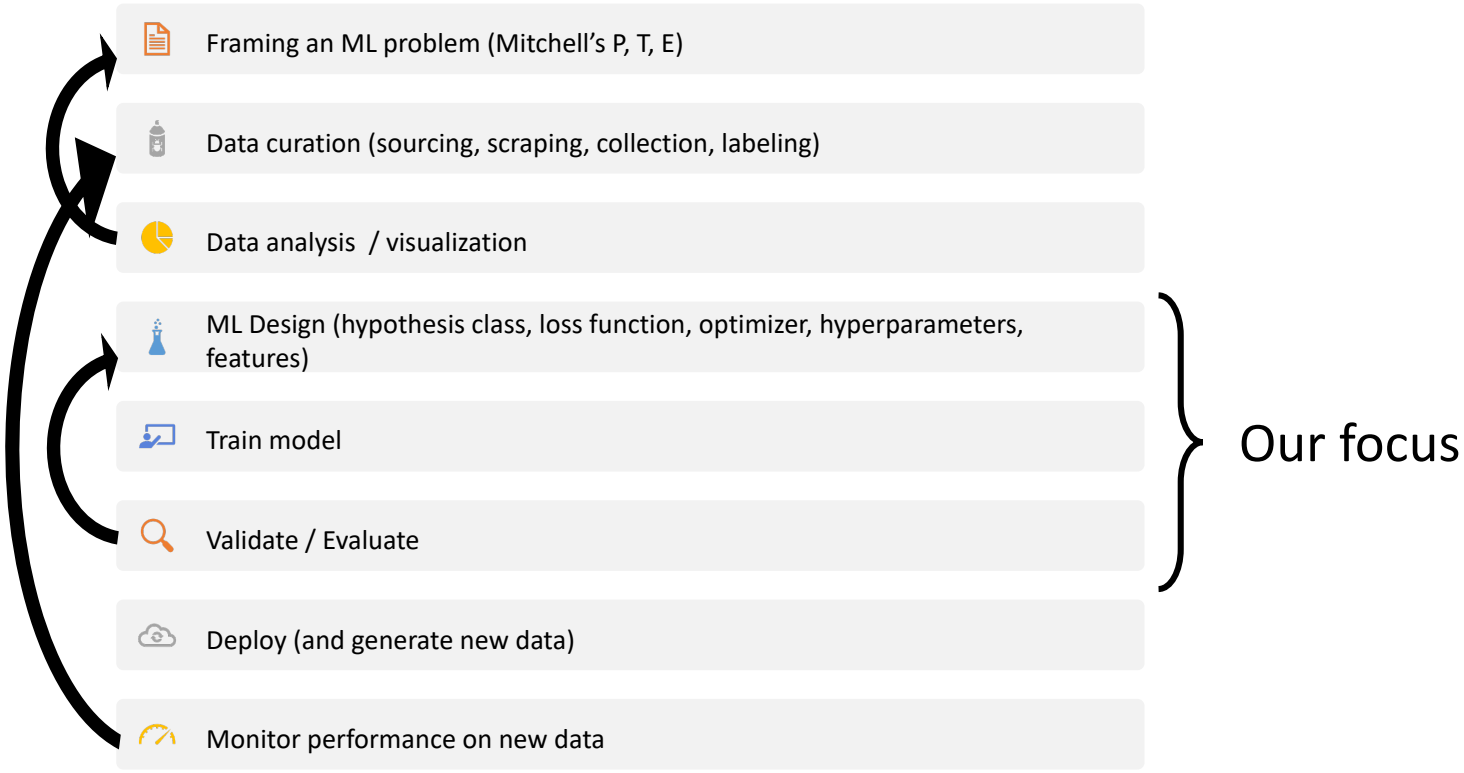


Image: https://www.flickr.com/photos/gsfc/5937599688/
Data from https://nsidc.org/arcticseaicenews/sea-ice-tools/

Machine Learning Workflow



Types of Learning

- **Supervised learning**
  - **Input:** Examples of inputs and outputs
  - **Output:** Model that predicts unknown output given a new input

- **Unsupervised learning**
  - **Input:** Examples of some data (no “outputs”)
  - **Output:** Representation of structure in the data

- **Reinforcement learning**
  - **Input:** Sequence of interactions with an environment
  - **Output:** Policy that performs a desired task

Supervised Learning

- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x

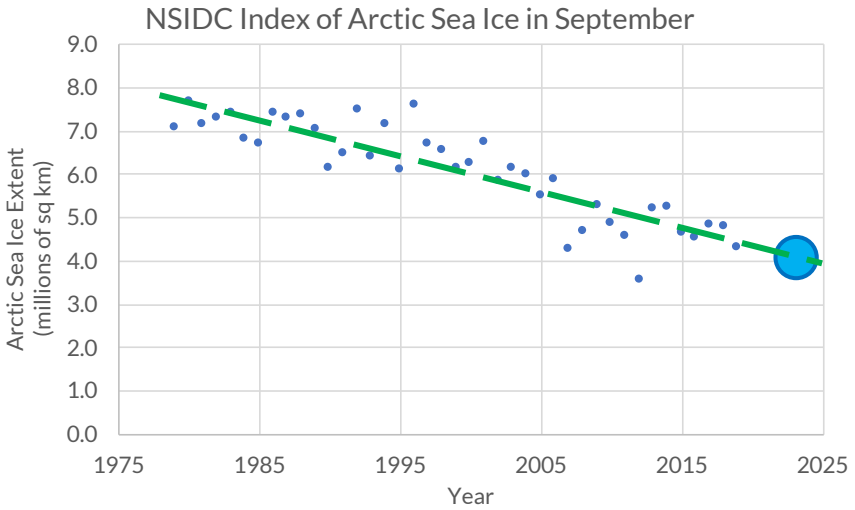


Image: https://www.flickr.com/photos/gsfc/5937599688/
Data from https://nsidc.org/arcticseaicenews/sea-ice-tools/

Supervised Learning

- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x
- **Regression:** Labels y are real-valued

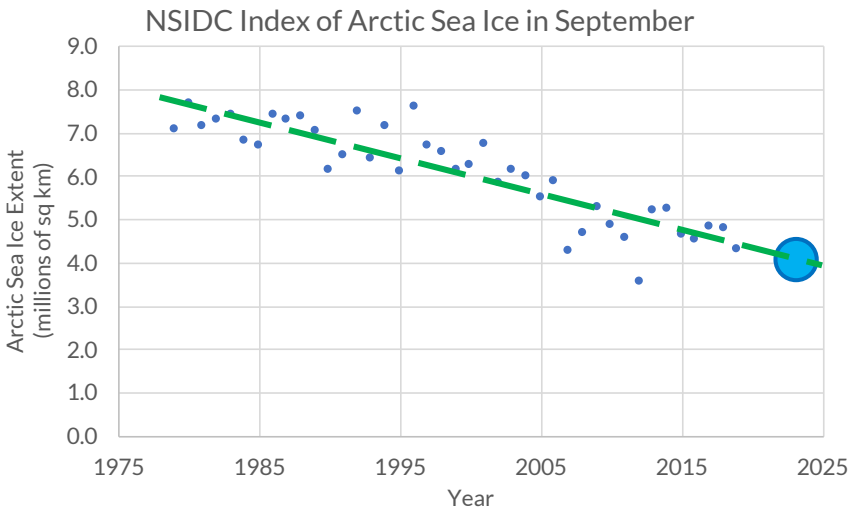


Image: https://www.flickr.com/photos/gsfc/59375599688/
Data from https://nsidc.org/arcticseaicenews/sea-ice-tools/

Supervised Learning

- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x
- **Classification:** Labels y are categories

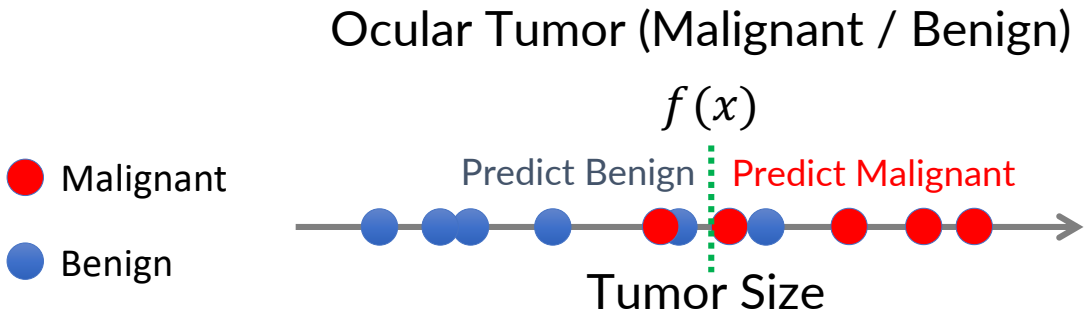
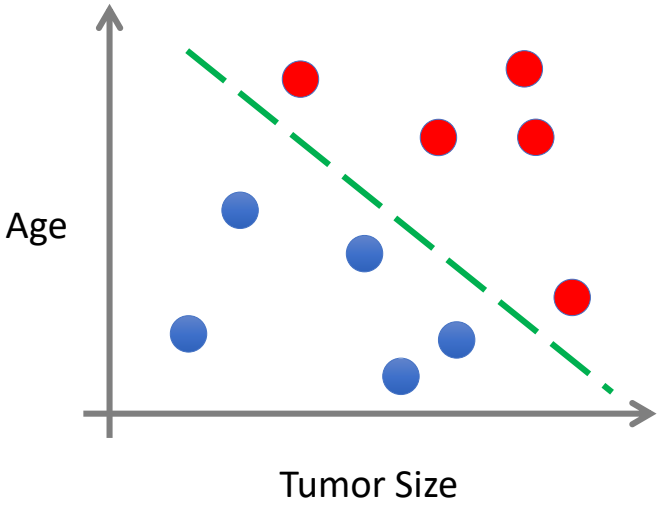


Image: https://eyecancer.com/uncategorized/choroidal-metastasis-test/

Supervised Learning

- Given $(x_1, y_1), \ldots, (x_n, y_n)$, learn a function that predicts y given x
- Inputs x can be multi-dimensional



- Patient age
- Clump thickness
- Tumor Color
- Cell type
- …

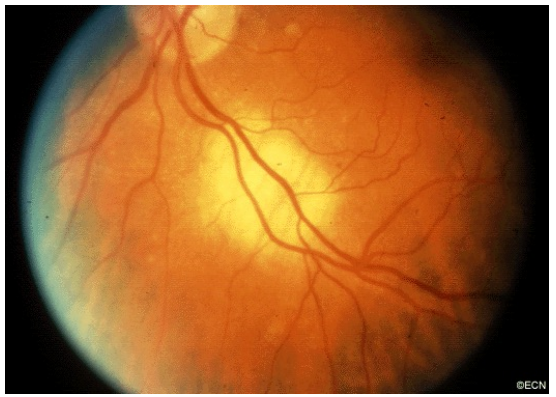
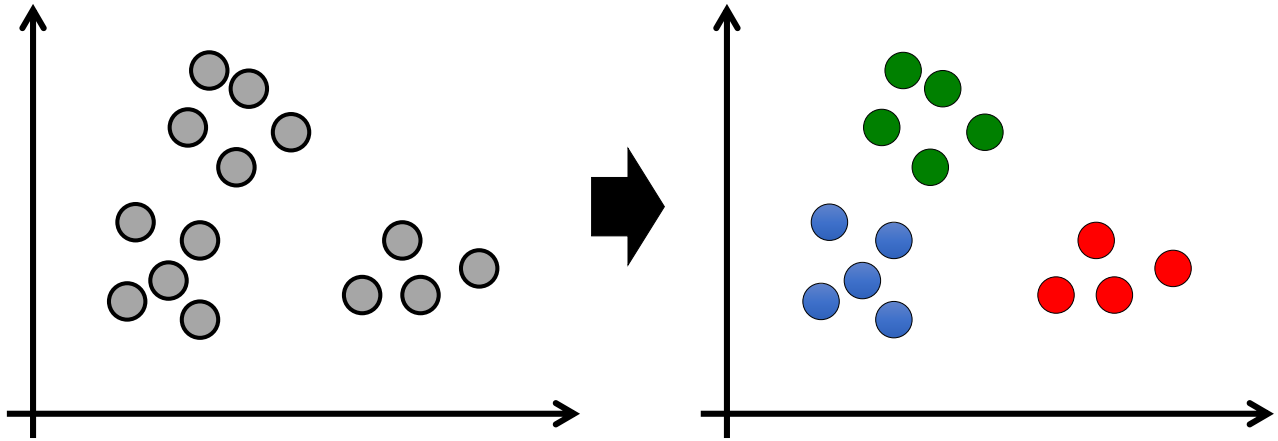


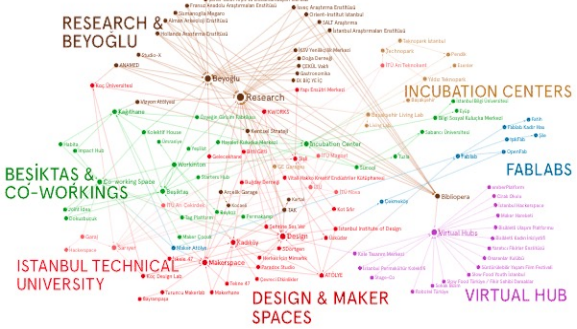
Image: https://eyecancer.com/uncategorized/choroidal-metastasis-test/

Unsupervised Learning

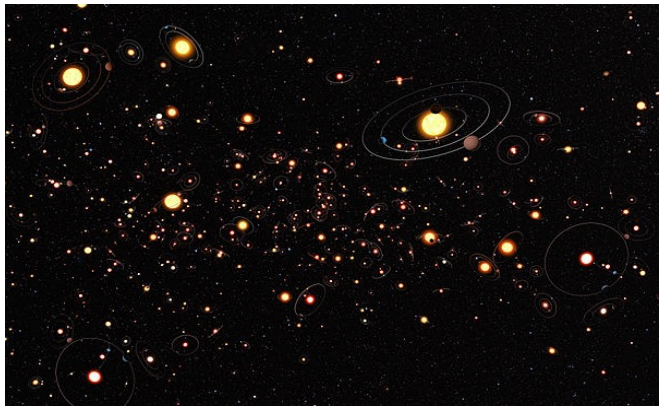
- Given $x_1, \ldots, x_n$ (no labels), output hidden structure in x's
  - E.g., clustering



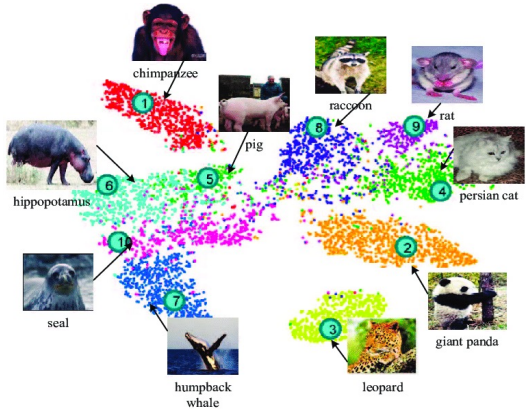
Unsupervised Learning



Find Subgroups in Social Networks



Identify Types of Exoplanets



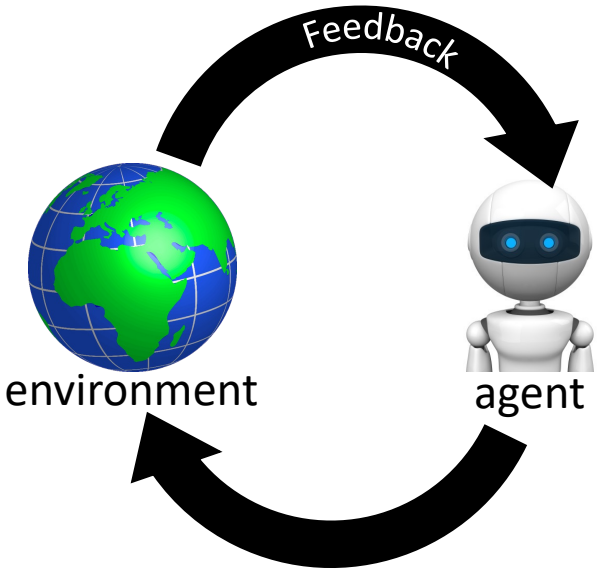
Visualize Data

Image Credits:
https://medium.com/graph-commons/finding-organic-clusters-in-your-complex-data-networks-5c27e1d4645d
https://arxiv.org/pdf/1703.08893.pdf
https://en.wikipedia.org/wiki/Exoplanet

Reinforcement Learning

- Learn how to perform a task from interactions with the **environment**

- **Examples:**
  - Playing chess (interact with the game)
  - Robot grasping an object (interact with the object/real world)
  - Optimize inventory allocations (interact with the inventory system)

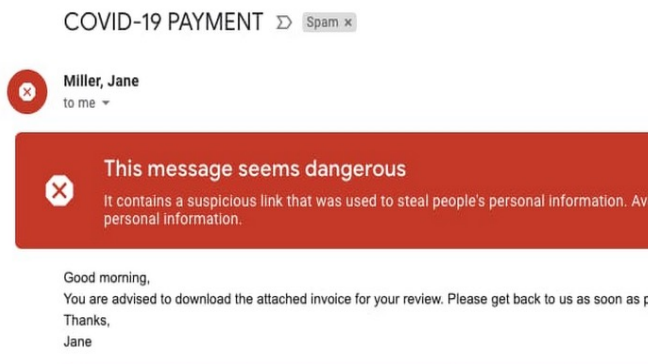


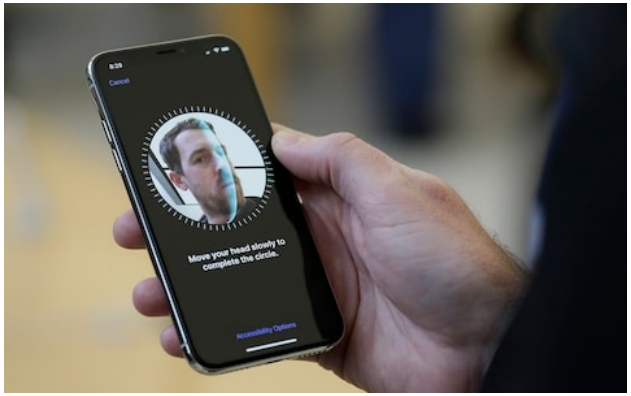
Reinforcement Learning

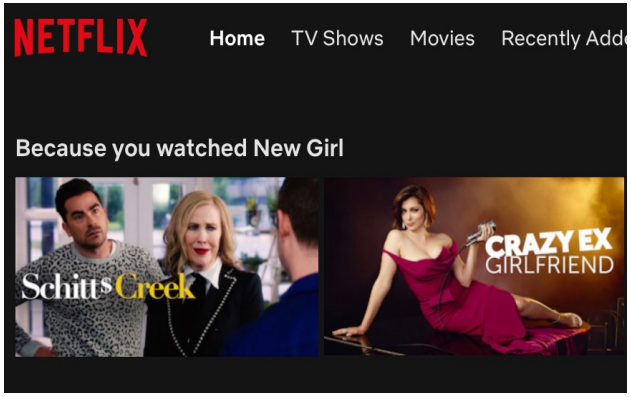
https://www.youtube.com/watch?v=iaF43Ze1oeI

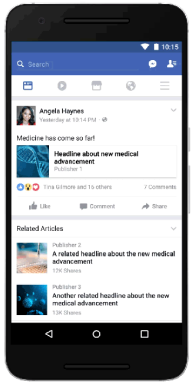
Applications of Machine Learning

Everyday Applications



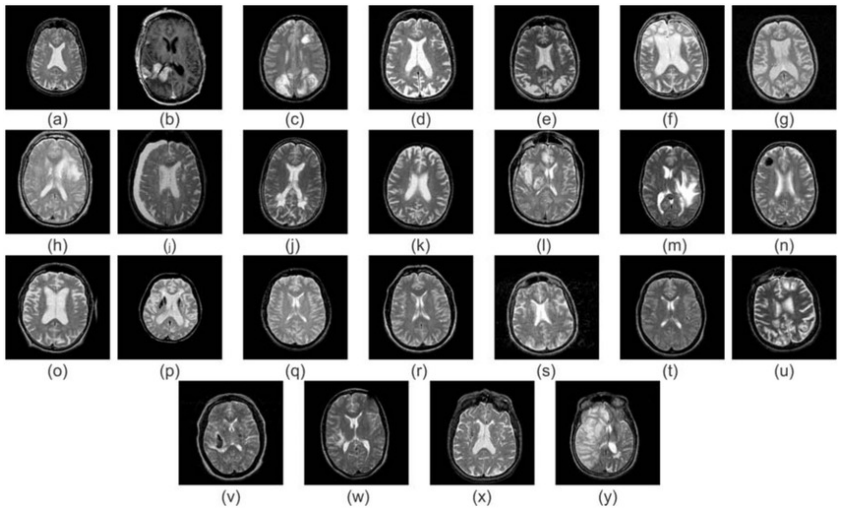






Radiology and Medicine

Input: Brain scans



Output: Neurological disease labels

Machine learning studies on major brain diseases: 5-year trends of 2014–2018

Koji Sakai[1] · Kei Yamada[1]

Applications of machine learning in drug discovery and development

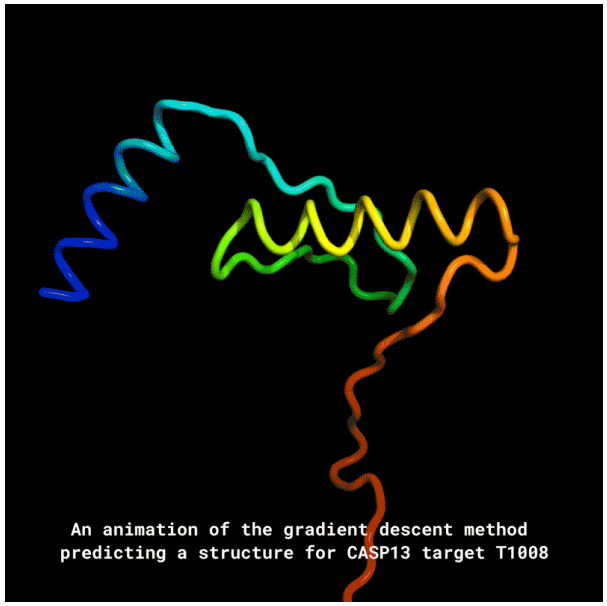
https://www.nature.com/articles/s41573-019-0024-5

Deep learning-enabled medical computer vision

Andre Esteva ✉, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean & Richard Socher

https://www.nature.com/articles/s41746-020-00376-2

Scientific Discovery



https://deepmind.com/blog/article/AlphaFold-Using-AI-for-scientific-discovery

https://www.jpl.nasa.gov/edu/news/2019/4/19/how-scientists-captured-the-first-image-of-a-black-hole/

Creating Images & Text

https://thispersondoesnotexist.com/

SYSTEM PROMPT (HUMAN-WRITTEN)

Recycling is good for the world.

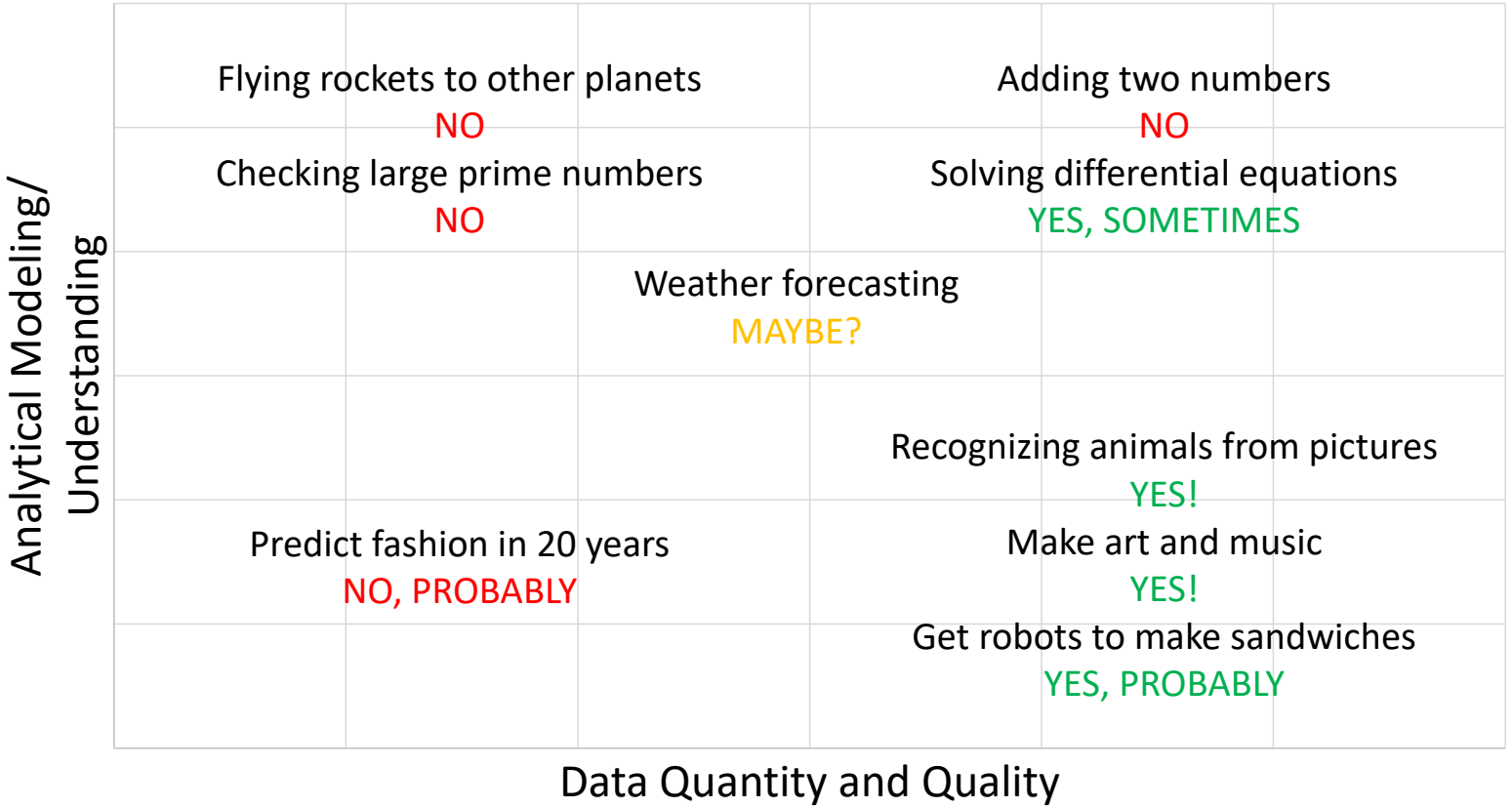
NO! YOU COULD NOT BE MORE WRONG!!

MODEL COMPLETION (MACHINE-WRITTEN, 25 TRIES)

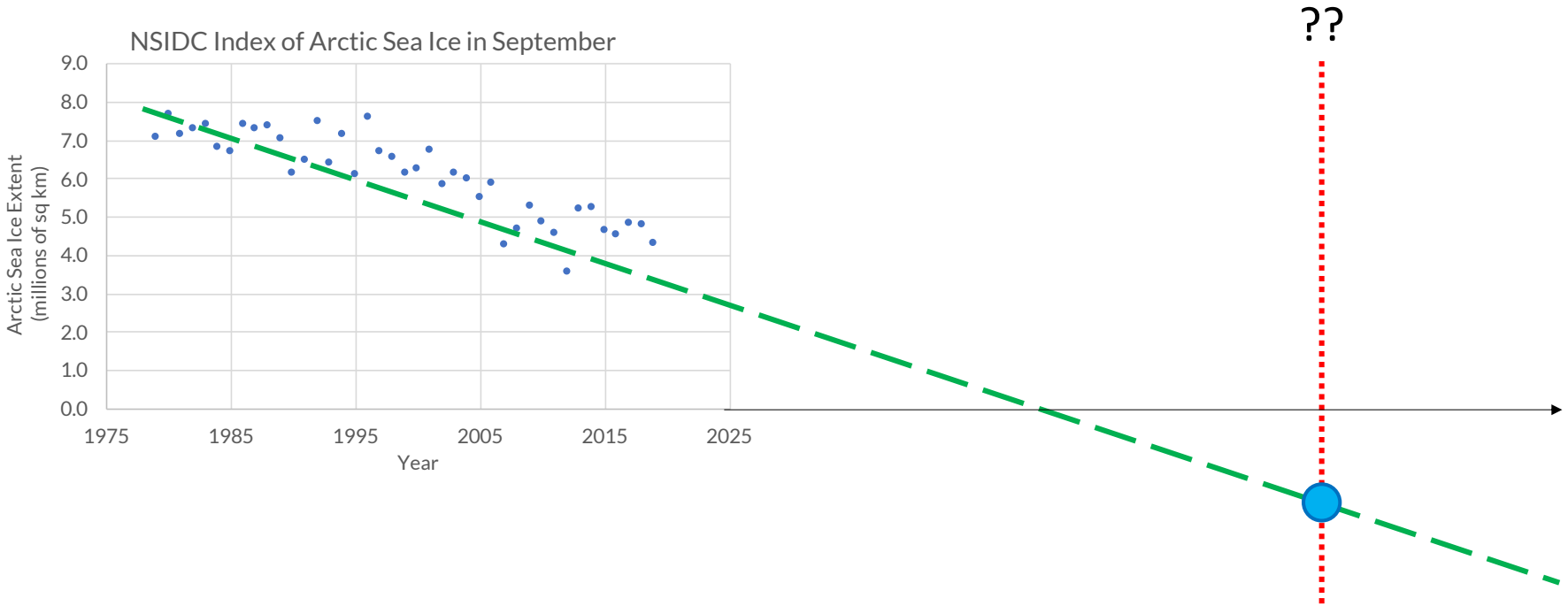
Recycling is NOT good for the world. It is bad for the environment, it is bad for our health, and it is bad for our economy. I'm not kidding. Recycling is not good for the environment. It is destructive to the earth and it is a major contributor to global warming. Recycling is not good for our health. It contributes to obesity and diseases like heart disease and cancer. Recycling is bad for our economy. It increases the cost of a product, and in turn, the price of everything that is made with that product. Recycling is not good for our nation. We pay a

https://transformer.huggingface.co/doc/gpt2-large

When should we use machine learning?



Danger of Out-of-Domain Machine Learning



Any time you are evaluating on data “far” from your training data, beware!

Ethical Considerations

“The Pennsylvania Board of Probation and Parole has begun using machine learning forecasts to help inform parole release decisions. In this paper, we evaluate the impact of the forecasts on those decisions and subsequent recidivism.”

An impact assessment of machine learning risk forecasts on parole board decisions and recidivism

Richard Berk

“In 2013, the University of Texas at Austin’s computer science department began using a machine-learning system called GRADE to help make decisions about who gets into its Ph.D. program”

The Death and Life of an Admissions Algorithm

“Videos about vegetarianism led to videos about veganism. Videos about jogging led to videos about running ultramarathons. It seems as if you are never ‘hard core’ enough for YouTube’s recommendation algorithm. It promotes, recommends and disseminates videos in a manner that appears to constantly up the stakes. Given its billion or so users, YouTube may be one of the most powerful radicalizing instruments of the 21st century.”

YouTube, the great radicalizer

THE NEW YORK TIMES / ZEYNEP TUFEKCI / MAR 12