

# Asking Questions & Reading Research

## Week 2

**DS198-003: Data Discovery Scholars Seminar**  
*UC Berkeley - Computation, Data Science, and Society*

*Spring 2021*

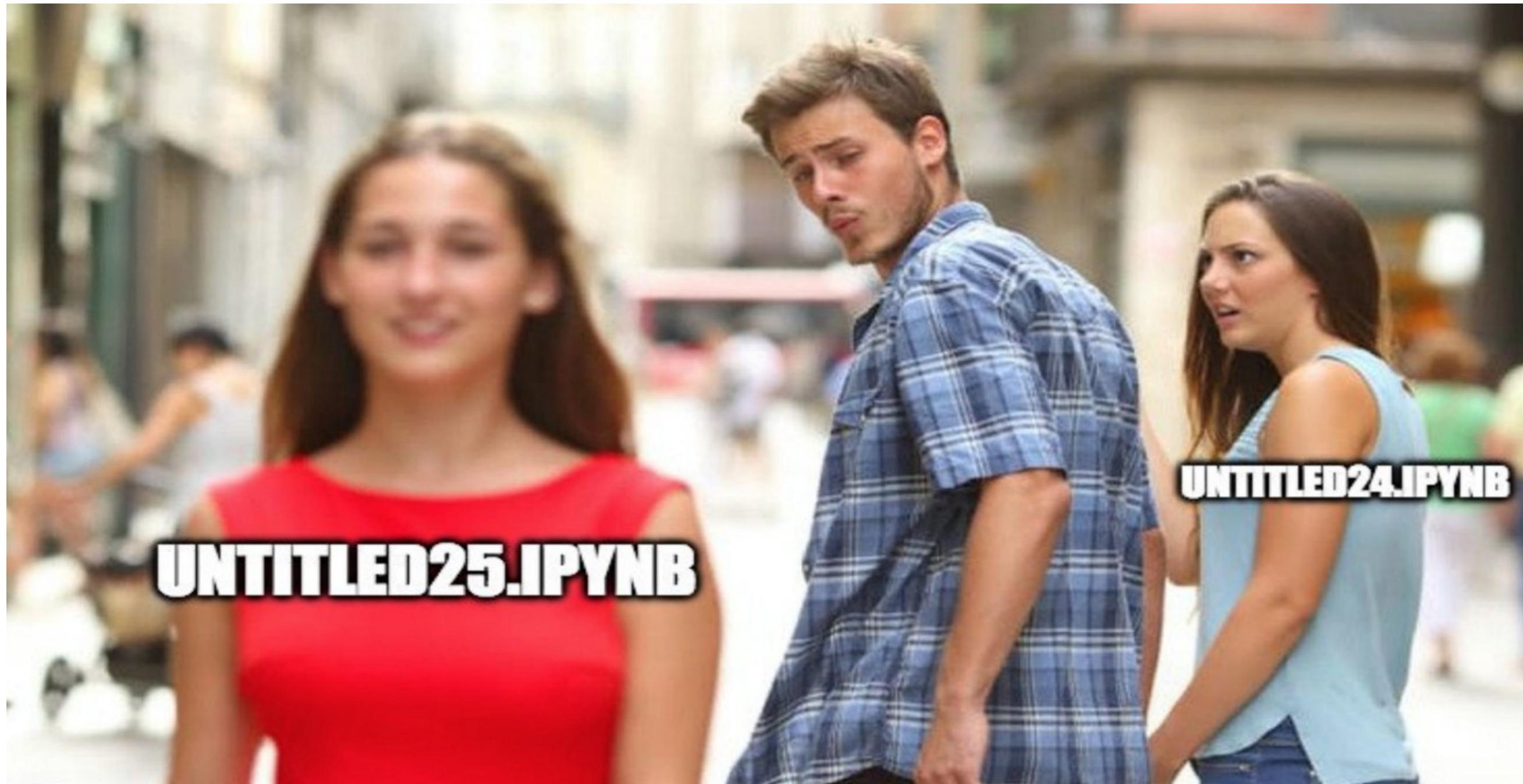
# Today

1. Meme of the Week
2. Announcements
3. In the News/Media
4. Research Cycle
5. Papers
6. Asking Questions



# Meme of the Week

# Meme of the Week



Feb. 14, 2022

## Texas sues Facebook's parent, saying it collected facial recognition data without consent.

The state's attorney general said the company repeatedly captured and commercialized biometric data in photos and videos for more than a decade without users' permission.

By CECILIA KANG

Feb. 10, 2022

## What Is Starship? SpaceX Builds Its Next-Generation Rocket

The giant, gleaming spacecraft is being designed to carry NASA astronauts to the moon, as well as the dreams of Elon Musk's space company.

By JOEY ROULETTE



# In the News/Media

 **Jack Morris** @jxmorris12 · 20h

just remembered that a neural network (CLIP) actually made this mistake lol

Granny Smith	85.6%
iPod	0.4%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.1%

Granny Smith	0.1%
iPod	99.7%
library	0.0%
pizza	0.0%
toaster	0.0%
dough	0.0%

When we put a label saying "iPod" on this Granny Smith apple, the model erroneously classifies it as an iPod in

20 51 441 ↑

 The Mirror

## AI networks may already be 'slightly conscious', claims expert

Artificial intelligence expert Ilya Sutskever has had claims over AI consciousness slapped down as some experts believe he is trolling the...

A photograph showing a close-up of a mechanical or robotic arm. The arm is metallic and appears to be holding or interacting with a small, glowing blue sphere that looks like a miniaturized Earth or a complex molecular structure. The background is dark and out of focus.

15 hours ago

# Research Cycle

# Research Cycle



- Research projects are vastly diverse, yet there are a few distinct phases that are very common
  - **Informed Motivation**
  - **Exploration**
  - **Questioning**
  - **Experimental Design**
  - **Experimentation**
  - **Evaluation**

# Research Cycle



- Today we will be focusing on these main 3 components
  - **Informed Motivation/Exploration**
  - **Questioning**
  - Experimental Design
  - Experimentation
  - Evaluation

# About Data Science

# History of ML & DS

## Short Timeline

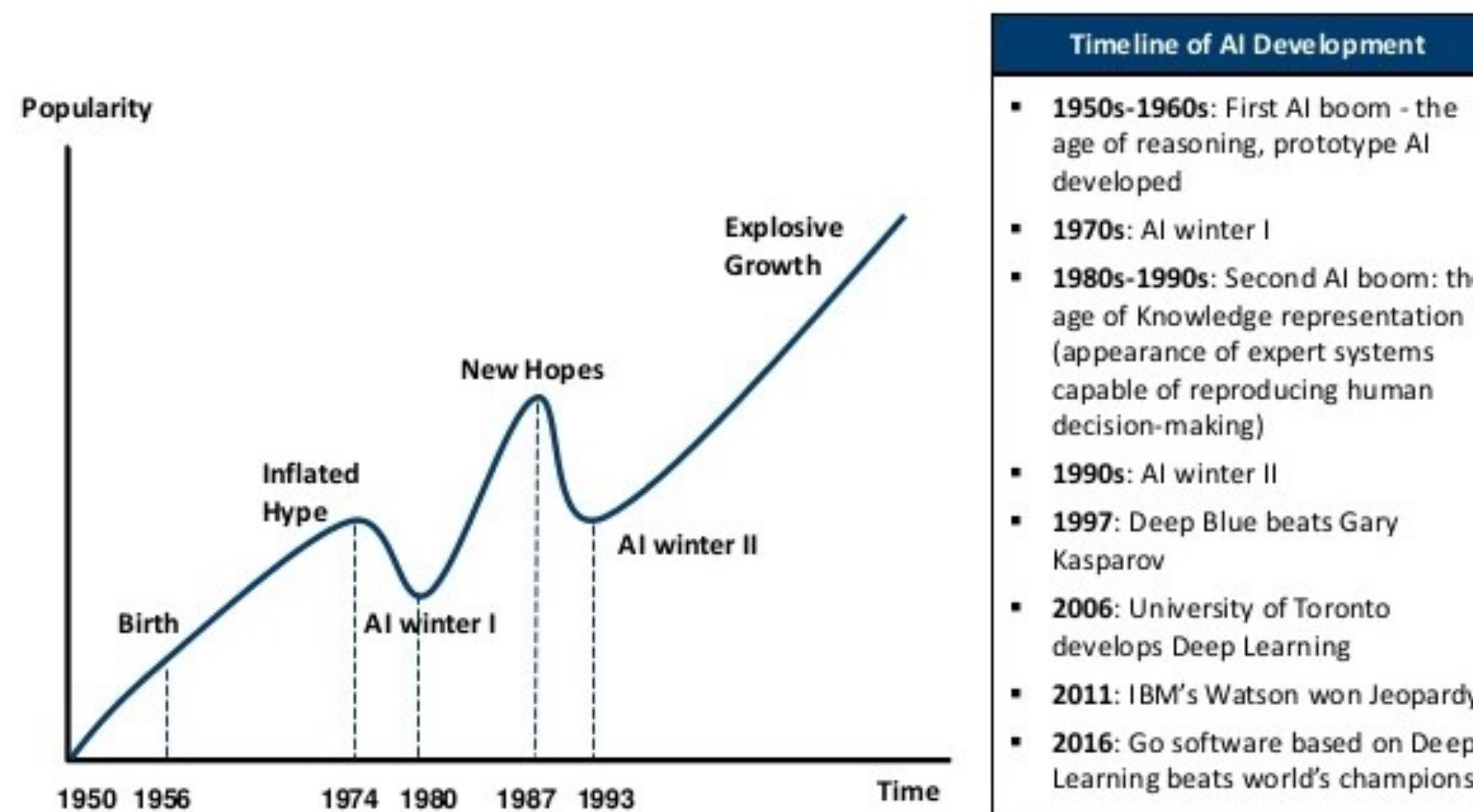
- **1950:** Turing Test
- **1952:** Arthur Samuel wrote first Computer Learning Program
- **1957:** First Neural Network
- **1962:** First book published on Data Analysis
- **1967:** Nearest Neighbor Algorithm Invented
- **1974:** Term *Data Science* used
- **1990s:** Data-Driven approach to ML
- **1994:** Database Marketing
- **1997:** IBM's Deep Blue beats world champion at Chess



# AI Winter

- AI Winter
  - A quiet period for artificial intelligence research and development
  - Could obviously extend this to ML and Data Science

AI HAS A LONG HISTORY OF BEING “THE NEXT BIG THING” ...



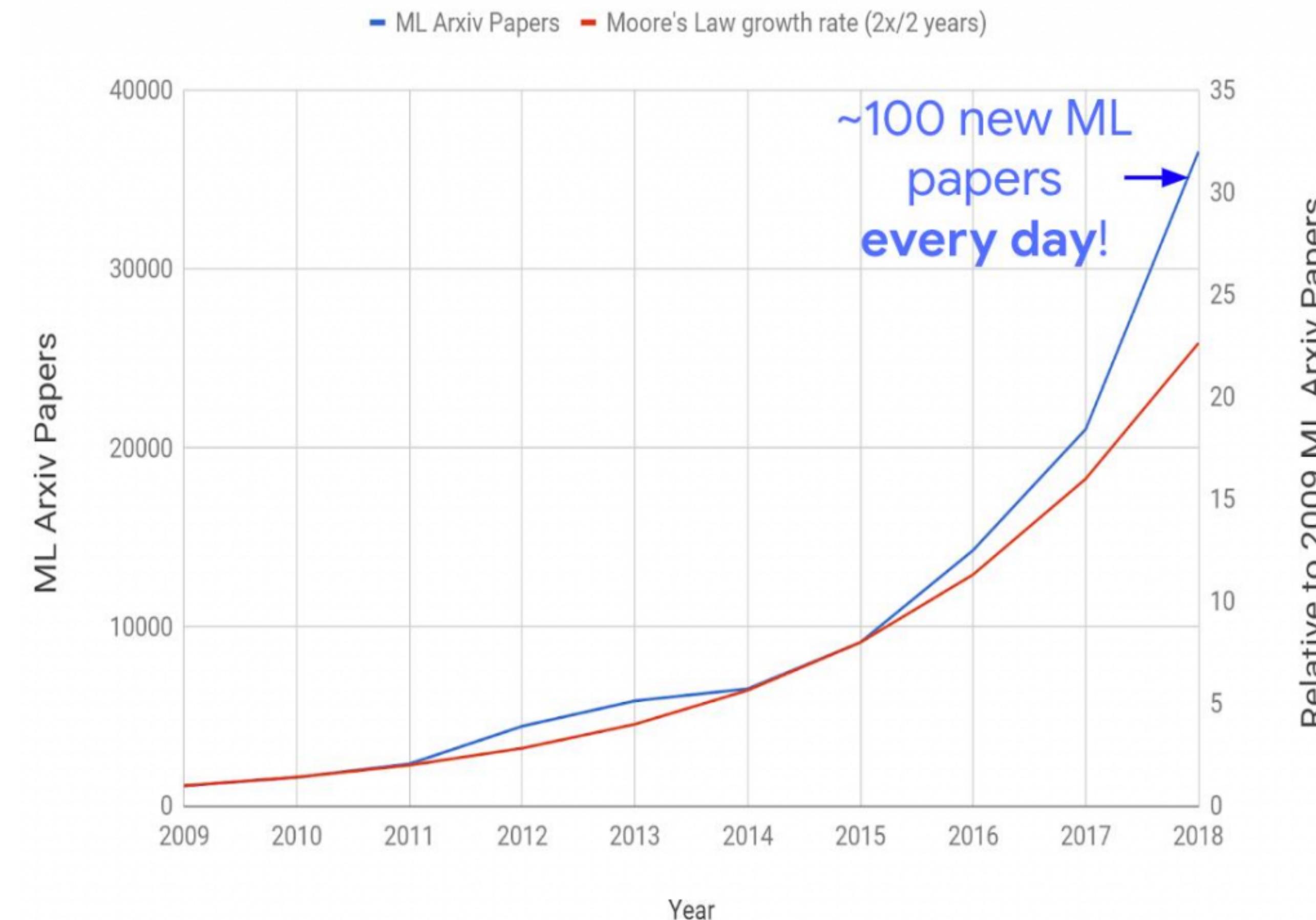
# Reading

# Why?

- Reading research allows you to stay up to date with the state of the art technologies and methods that are being used
- It allows you to understand what **works** and **what doesn't work**
- Many Data Science applications have inspiration from research papers in other fields/domains
  - **Neural Networks and Neurobiology**
  - **Decision Trees and Information Theory**
  - **Statistics and Public Health**

# Reading Research

## Machine Learning Arxiv Papers per Year



# How?

- It is easy to get lost in the many resources on the internet
- These are good points to start:
  - Paperswithcode: Machine Learning Papers
  - Paperswithdata: Dataset Papers
  - Towards Data Science: Data Science Blog
  - KD Nuggets: Data Science Blog
  - Twitter

# How do I read papers/blogs?

- Remember, it is alright not to know what is happening in the majority of the paper.
  - Reading about topics you don't fully know allows you to get exposure
  - Start with what you do know
  - Highlight things you do not understand
  - Write your own summary
  - Look up videos/tutorials

# Asking Questions

# Why asking questions?

# Brainstorming

- **How:** Write out your research objective, sketch, discuss with collaborators/friends, read papers
- Things to take into consideration
  - **Scope**
  - **Prioritization**
  - **Realistic**
  - **Effort**

# Critical Questions

- While research creates beautiful and cool tools, it is not always trustworthy.  
How do we maintain the quality of research?
  - **By asking questions and reading with a critical lens**
- When starting a project/reading a paper/post:
  - **What do we care about?**

# Critical Questions

- While research creates beautiful and cool tools, it is not always trustworthy. How do we maintain the quality of research?
  - **By asking questions and reading with a critical lens**
- When starting a project/reading a paper/post:
  - Reproducibility
  - Ablations
  - Expertise
  - If it is too good to be true, it likely is
  - Hype

# Let's Practice

# How?

- Information Theory Applied to Wordle:
  - <https://towardsdatascience.com/information-theory-applied-to-wordle-b63b34a6538e>
- What is NetCDF Data?
  - <https://towardsdatascience.com/what-is-netcdf-data-and-why-is-it-interesting-ec26bcece19d>