Data Science: What's hot? What's not?

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

Spring 2022

Date	#	Description	Assignments
4/12	7	What’s hot in DS & Presenting	
4/19		Check-In II	
4/26	🏁	Final Presentation	Capstone

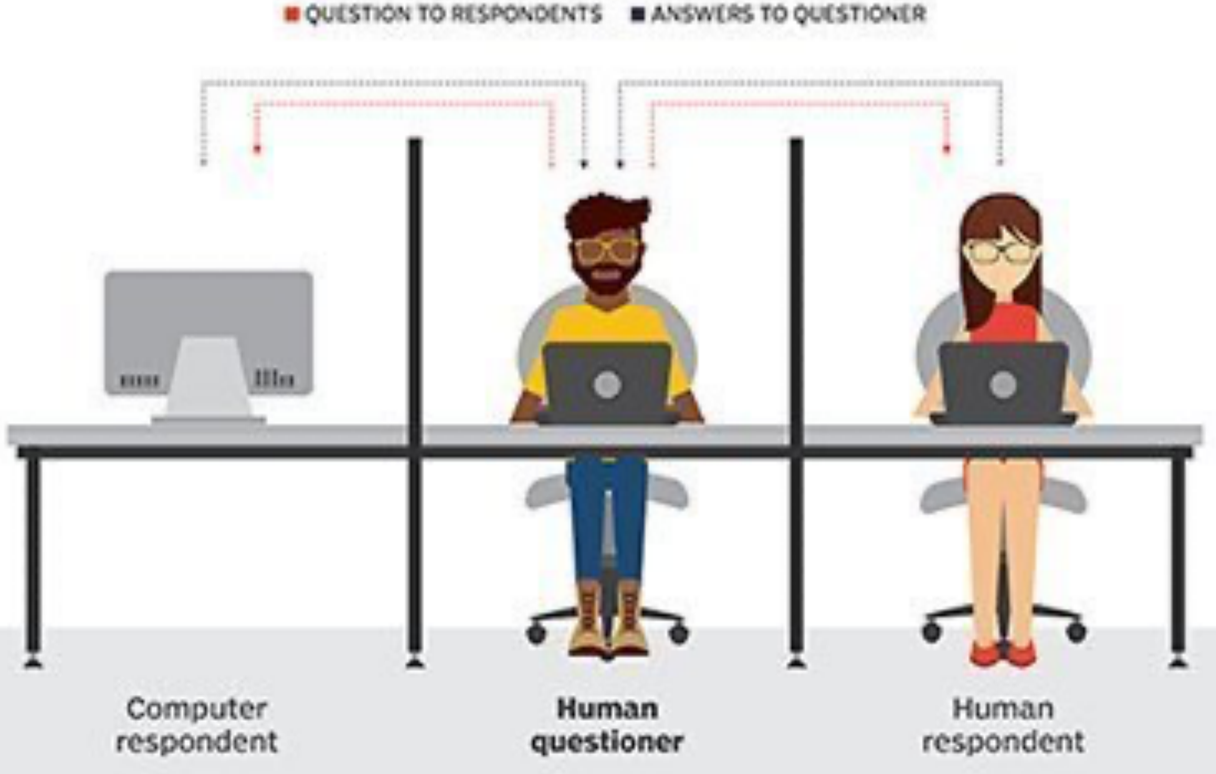
Meme of the Week


Seth Rosen
@sethrosen
Them: Can you just quickly pull this data for me?
Me: Sure, let me just:
SELECT * FROM some_ideal_clean_and_pristine.table_that_you_think_exists
1:42 PM · 20 Apr 20 · Twitter Web App
3,161 Retweets 20.5K Likes

History

Turing test

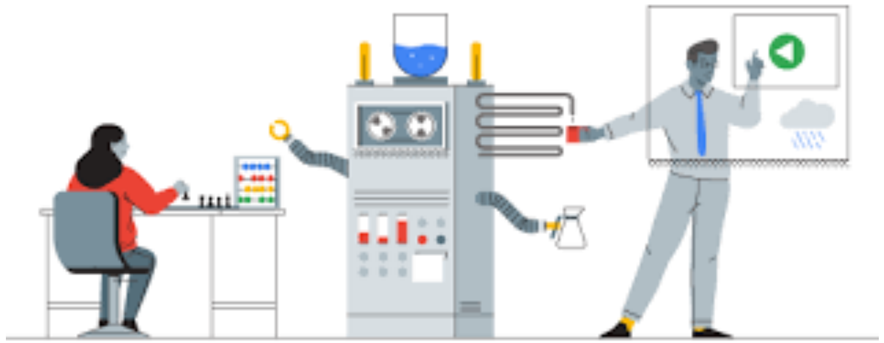
During the Turing test, the human questioner asks a series of questions to both respondents. After the specified time, the questioner tries to decide which terminal is operated by the human respondent and which terminal is operated by the computer.



History of ML & DS

Short Timeline

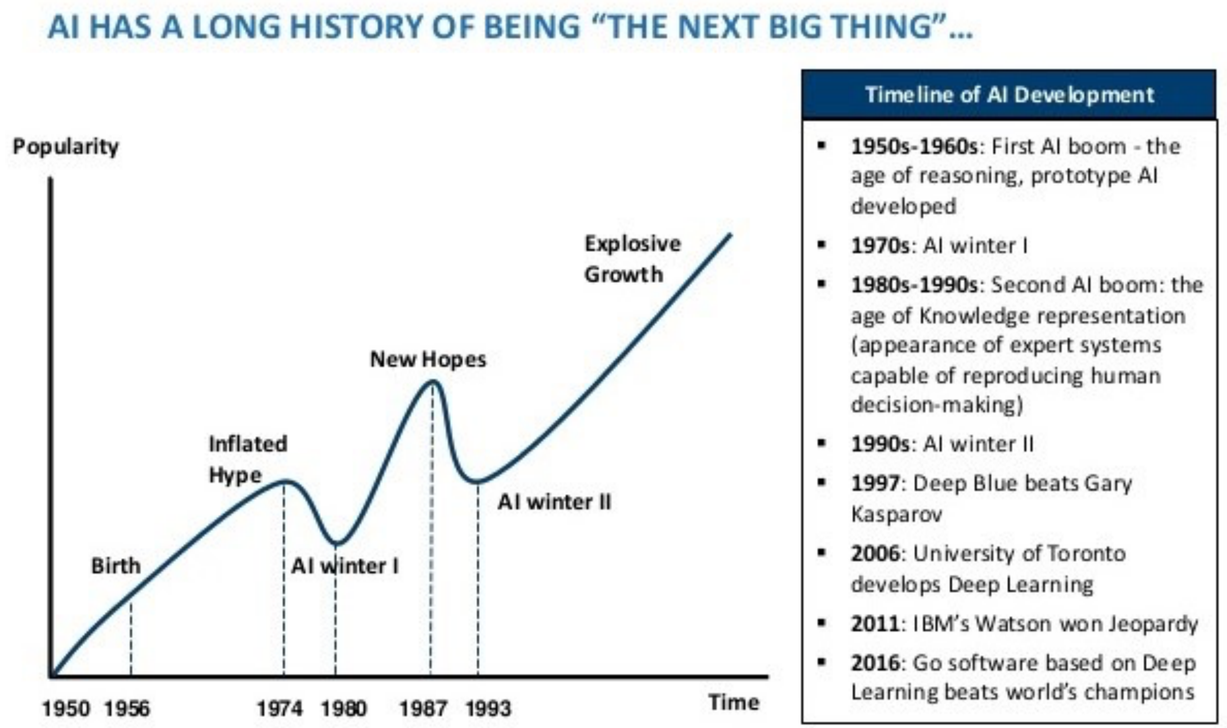
- **1950**: Turing Test
- **1952:** Arthur Samuel wrote first Computer Learning Program
- **1957:** First Neural Network
- **1962:** First book published on Data Analysis
- **1967:** Nearest Neighbor Algorithm Invented
- **1974:** Term *Data Science* used
- **1990s:** Data-Driven approach to ML
- **1994:**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



AI Winter

AI Winter

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



AI Winter

- **AI Winter I (1970s)**
  - Machine-Translation: Would require too much data
  - Perceptron: Can only solve linearly separable information
- **Revival**
  - ‘Expert System’: Use a bunch of if/else rules
  - Knowledge Representation
- **AI Winter II**
  - These machines were too brittle and inefficient
- **Deep Learning & Data-Centric Design**
  - ImageNet, Natural Language Processing, AlphaGo/Zero

AI Winter

- **Factors that contribute to AI winters**
  - Hype
  - Economic Features
  - Computing Capability
  - Empty Pipeline
  - Failure to Adapt
  - Data Availability

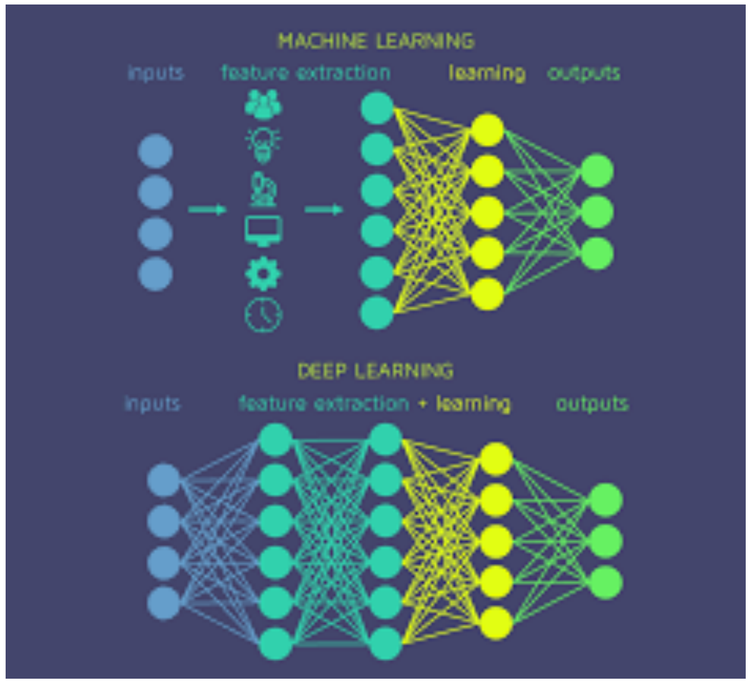
- **When is the next AI winter?**

Check out

https://bdtechtalks.com/tag/demystifying-ai/

What’s hot?

That’s so 2020



That’s so 2020…

- Deep Learning
  - Image Recognition/Detection
  - NLP: Speech-to-text & Machine Translation
  - Vision/RL: Autonomous Driving
  - Generally, prediction tasks that require much data

What’s hot

- **Auto-encoders**
  - How do reconstruct data has been fed in
- **Unsupervised/Self-Supervised Learning**
  - How do we learn with much less data?
    - AlphaGo/AlphaFold
- **Attention & Transformers**
  - Enabled Natural Language Sequences
- **GANs**
  - DeepFakes (and many other applications)

Opportunities

- **Meta-Learning**
  - Learning how to learn, how to tune parameters, how to augment data
- **Deploying & Tracking ML/DS Algorithms** (As discussed last week)
  - How do we track performance of algorithms in practice?
- **Privacy-Preserving Learning**
  - Federated Learning
- **Unsupervised Learning**
  - How do we use DL to get insights about our world?

A BRIEF HISTORY OF MACHINE LEARNING & A.I.

1st Generation: The Backend

LARGE DATA SETS

- Fraud Detection
- Search Algos
- Supply Chain Mngmt

2nd Generation: The Human Side

DATA ABOUT HUMANS

- Rec. Systems
- Social Media
- Commerce + Ads

3rd Generation: Modern Machine Learning

PATTERN RECOGNITION

- Speech Recog.
- Computer Vision
- Translation

?

1990 2000 2010 2020

Presenting

Presentations

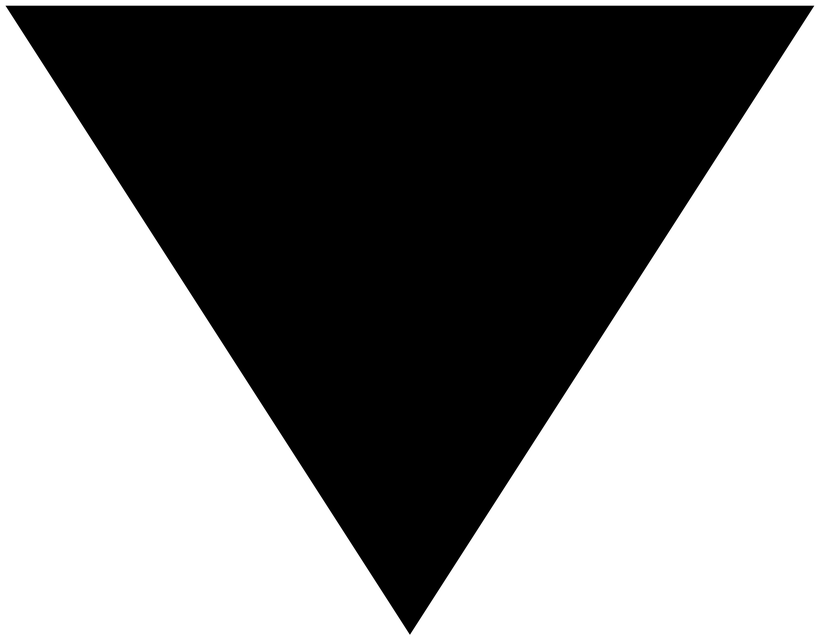
Game Plan

- How long is your talk?
- What kind of audience is going to show up?
- Is it formal/non-formal?
- Is it more like a talk or an actual presentation?
- What is your goal of your presentation?

Presentations

Game Plan

- Structure your presentation according to your (1) audience, (2) the objective of your talk by taking into account:



Attention

Building Intuition

Details

Presentations

Game Plan

- Depending on your outline, you might want to write out a script for your presentation.
  - Personal Preference
- Slide Decks can be helpful to structure talk
  - **Do not overcrowd your slides**
  - **One thought/idea per slide**
  - **Use visualizations/bolding**
  - **Mimic flow of thought**

Speaking with Purpose

- **Why** is the problem you are solving worth solving?
- What is the **core difference** between your method and all those that came before? This is really a two-part question (which most speakers screw up by answering only the second part).
  - What does your method **accomplish** that no previous method accomplishes?
  - What algorithmic or methodological idea enables your method to accomplish more?
- What is the **evidence** that your method is better in some circumstances? (And what are those circumstances?)
- What is the **one big idea** that you want people to leave your talk with? If you try to get across five ideas, you will usually impart none. If you choose one main idea and focus on advertising it, you will usually succeed. “Give them something to take home.”

https://people.eecs.berkeley.edu/~jrs/speaking.html

Group Discussion

Group Discussion

Let’s discuss the following questions (identical to the ones in your reflection). We will discuss what you talked about with the whole class afterwards.

- Do you think there will be an AI/Tech winter? Why is that the case?

- What next challenge/opportunity do you think is going to be next big one? What makes your chosen technology so special?

- What are the benefits and risks of hype/speculation, fluctuating availability of compute and economic trends on the development and advancement of technology?

Midsemester Check In II

Midsemester Check-In II

See link in slack

Discovery Scholar Capstone Assignment

Discovery Scholar Capstone Assignment

What did you learn this semester? What did you wish you had known at the beginning of the semester? How can we recreate a better programming/DS tutorial? What were the findings of your project?

In this assignment, you will be writing a blog post, creating a vlog/video, making a tutorial, useful illustration or whatever format you wish in which you share useful advice, your findings, walk through technical skills that will be helpful for the DS community or even future scholars.

At the end, you will have something to share with your LinkedIn network, Twitter or future employers/schools. If we have your permission, we will also share this with the future Data Discovery Scholars by showcasing it on a new website.

DUE: April 26th (By 5pm)