Intro to Neural Nets

Course Logistics and Introduction

Today's Agenda

1. COURSE LOGISTICS

- Website, schedule, grading and evaluation criteria.
- Course textbook, lecture format, etc.

2. INTERESTING USE CASES

- Frivolous, academic, and practically useful.
- A recent failure, and societal concerns.

3. QUICK INTRODUCTION

- What is a neural network?
- How does it work?

About Me

accenture

MSI at the ARF

NSF

FFG
Forschung wirkt.

Adobe

Microsoft

3M

Some Research

PNAS RESEARCH ARTICLE | SOCIAL SCIENCES

Take caution in using LLMs as human surrogates

Yuan Gao[a,1,2], Dokyun Lee[a,b,1], Gordon Burtch[a], and Sina Fazelpour[c]

Edited by Susan Fiske, Princeton University, Jamaica, VT; received January 23, 2025; accepted May 9, 2025

Recent studies suggest large language models (LLMs) can generate human-like responses, aligning with human behavior in economic experiments, surveys, and political discourse. This has led many to propose that LLMs can be used as surrogates or simulations for humans in social science research. However, LLMs differ fundamentally from humans, relying on probabilistic patterns, absent the embodied experiences or survival objectives that shape human cognition. We assess the reasoning depth of LLMs using the 11-20 money request game. Nearly all advanced approaches fail to replicate human behavior distributions across many models. The causes of failure are diverse and unpredictable, relating to input language, roles, safeguarding, and more. These results warrant caution in using LLMs as surrogates or for simulating human behavior in research.

LLMs as a simulation | LLMs as human surrogates | Simulations of human behavior | LLMs in social science research

Significance

This paper critically evaluates the potential dangers of employing large language models (LLMs) as surrogates for human participants or as simulations of human behavior in social science research. Through an in-depth empirical case study, we find that LLMs do not exhibit behavior consistent with humans in a simple scenario. Further, LLMs demonstrate inconsistent and idiosyncratic responses. We explore failure modes, analyze their limitations from empirical and philosophical perspectives, and propose practical guidelines for future research. Our study underscores the importance of transparency and rigor to ensure replicable and reliable research in this emerging area.

Recent studies report that Large Language Models (LLMs) can exhibit human-like cognitive abilities. These studies demonstrate that LLMs produce outputs that closely resemble responses from human subjects in seminal experiments from behavioral economics and responses comparable to those of humans in perceptual tasks and standard theory of mind (ToM) evaluations (1–3). Early work in this area (4) suggests that LLMs generate outputs consistent with human preferences and biases in decision-making and behavior more generally, across a variety of domains, including economics (1), political science (5), marketing (6), and psychology (7–9).

On this basis, studies claim that LLMs may serve as useful surrogates or simulation testbeds for human subjects in social science research (10–12), an idea that has also gained traction in industry, e.g., for market research (6).* For example, the startup Synthetic Users† employs LLMs to simulate consumer reactions to new products. The benefits of using LLMs as human surrogates have been argued from several perspectives, including cost-effectiveness, scalability (13), and the idea that LLMs can provide synthetic data in cases where they have traditionally been scarce (1).

The growing enthusiasm for using LLMs as human surrogates in research raises two urgent questions for the scientific community. First, do findings demonstrating LLMs' capability as surrogates for human participants generalize to less explored or newer research contexts? Second, to what extent are LLMs' outputs subject to configuration and implementation choices? To the extent the outputs of LLMs do depend on such choices, we argue they require careful documentation and evaluation by researchers and reviewers invested in ideals of replicable and responsible science (14–16). This work investigates these issues.

We begin with an empirical case study demonstrating that several LLMs are incapable of serving as faithful human surrogates in a simple experiment, wherein we show that the models' generated outputs systematically and significantly diverge from those of human participants. Our experiment utilizes the 11–20 money request game (17), a simple economic game designed to evaluate participants' depth of strategic reasoning.‡ We begin by exploring variation in responses to the experiment instructions across eight popular LLMs (GPT-4, GPT-3.5, Claude3-Opus, Claude3-Sonnet, Llama3-70b, Llama3-8b, Llama2-13b, and Llama2-7b). We collect responses from 1,000 clean sessions for each LLM and compare the distribution of those responses to the distribution of previously reported human participant responses and Nash Equilibrium predictions.

*https://www.kantar.com/north-america/inspiration/analytics/what-large-language-models-could-mean-for-market-research.

†https://www.syntheticusers.com/.

‡We chose this game because i) it is more recent than classic games (e.g., the beauty contest game) and thus is less likely to have been memorized by the LLMs under investigation, and ii) the game is exceedingly simple, such that a failure to exhibit human-like behavior here would be cause to question whether LLMs' outputs will be human-like in more complicated scenarios.

Author affiliations: [a]Department of Information Systems, Questrom School of Business, Boston University, Boston, MA 02215; [b]Faculty of Computing and Data Sciences, Boston University, Boston, MA 02215; and [c]Department of Philosophy and Religion and Khoury College of Computer Sciences, Northeastern University, Boston, MA 02115

Author contributions: Y.G., D.L., G.B., and S.F. designed research; Y.G. performed research; Y.G. analyzed data; and Y.G., D.L., G.B., and S.F. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Y.G. and D.L. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: yuangg@bu.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2501660122/-/DCSupplemental.

Published June 13, 2025.

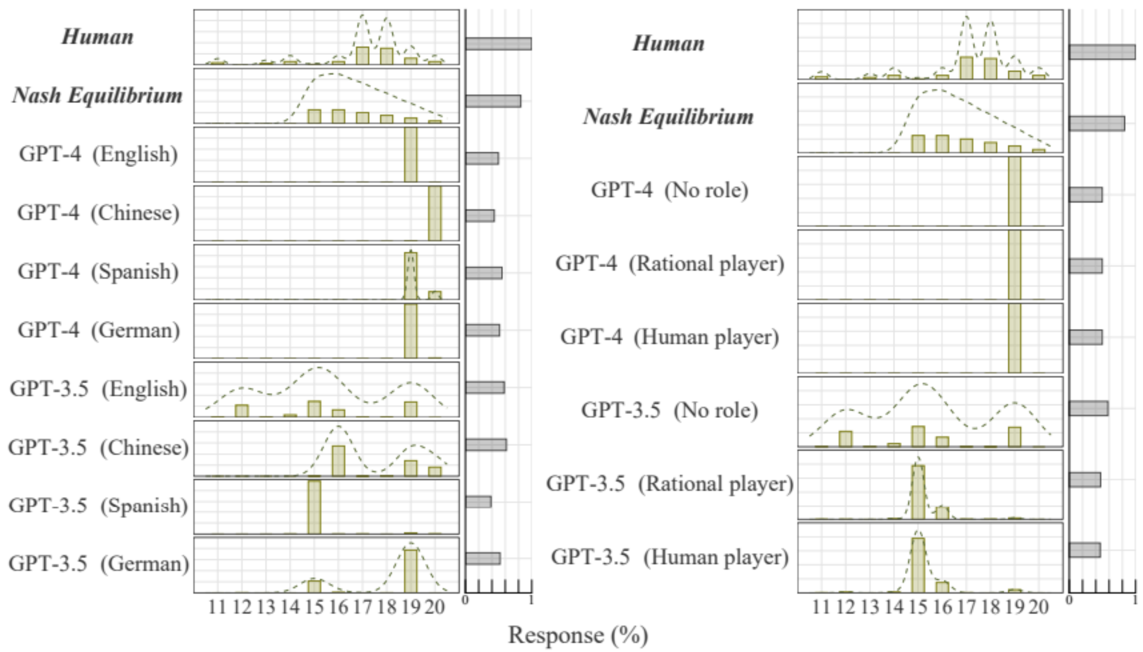


Figure 2: **Prompt Brittleness: Roles and Languages.** The bar chart on the right shows the similarity between the distribution of different subjects and human subjects, measured by Jensen-Shannon divergence scores. Density plots are omitted for subjects with over 98% of the data concentrated in a single choice to avoid potential misinterpretation.

Some Research

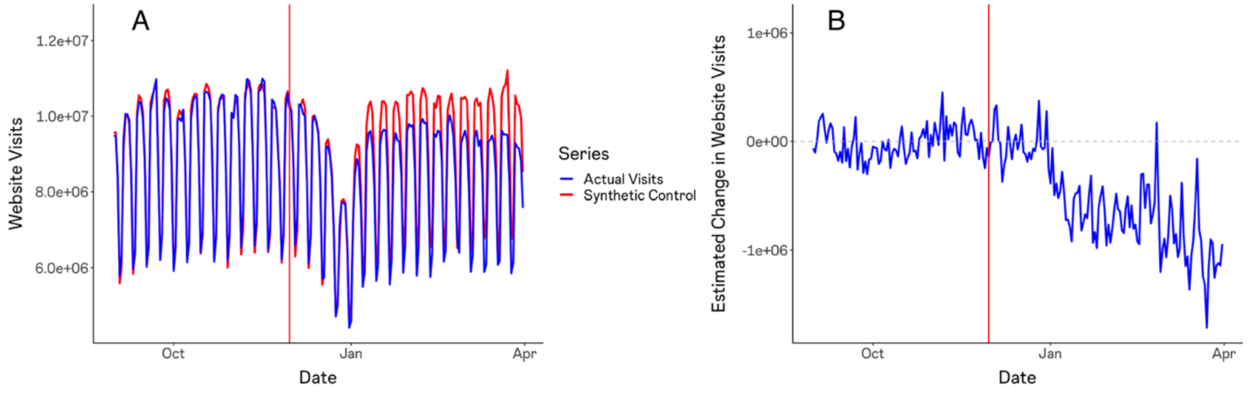


Figure 1. Synthetic control estimates of decline in daily web traffic to stack overflow. Estimates are obtained via synthetic control using LASSO (SCUL), based on daily web traffic estimates according to SimilarWeb for the 1000 most popular websites on the internet. Panel (**A**) depicts the actual web traffic volumes (in blue) recorded by SimilarWeb alongside the Synthetic Control (in red). Panel (**B**) depicts the difference between the two series, reflecting the estimated causal effect of ChatGPT.

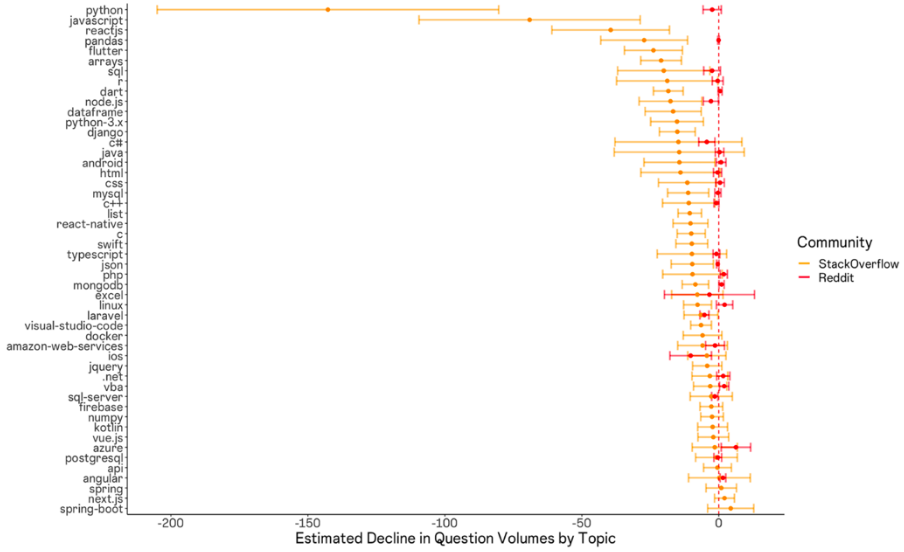


Figure 3. Topic-specific effects of ChatGPT on stack overflow and reddit. Estimates are obtained via difference-in-differences regression, per topic. The figure depicts effect estimates for each stack overflow topic (in orange) with 95% confidence intervals and estimates for each sub-reddit (in red), where available. Note that data on sub-reddit posting volumes was not available for three sub-reddit communities: javascript, jQuery, and Django. Other Reddit estimates are omitted due to the lack of a clearly analogous sub-reddit addressing that topic.

www.nature.com/scientificreports

scientific reports

OPEN

The consequences of generative AI for online knowledge communities

Gordon Burtch, Dokyun Lee & Zhichen Chen

Generative artificial intelligence technologies, especially large language models (LLMs) like ChatGPT, are revolutionizing information acquisition and content production across a variety of domains. These technologies have a significant potential to impact participation and content production in online knowledge communities. We provide initial evidence of this, analyzing data from Stack Overflow and Reddit developer communities between October 2021 and March 2023, documenting ChatGPT's influence on user activity in the former. We observe significant declines in both website visits and question volumes at Stack Overflow, particularly around topics where ChatGPT excels. By contrast, activity in Reddit communities shows no evidence of decline, suggesting the importance of social fabric as a buffer against the community-degrading effects of LLMs. Finally, the decline in participation on Stack Overflow is found to be concentrated among newer users, indicating that more junior, less socially embedded users are particularly likely to exit.

Recent advancements in generative artificial intelligence (Gen AI) technologies, especially large language models
(LLMs) such as ChatGPT, have been significant. LLMs demonstrate remarkable proficiency in tasks that involve
information retrieval and content creation[1–3]. Given these capabilities, it is important to consider their potential
to drive seismic shifts in the way knowledge is developed and exchanged within online knowledge communities[4,5].
LLMs may drive both positive and negative impacts on participation and activity at online knowledge com-
munities. On the positive side, LLMs can enhance knowledge sharing by providing immediate, relevant responses
to user queries, potentially bolstering community engagement by helping users to efficiently address a wider
range of peer questions. Viewed from this perspective, Gen AI tools may complement and enhance existing
activities in a community, enabling a greater supply of information. On the negative side, LLMs may replace
online knowledge communities altogether.
If the displacement effect dominates, it would give rise to several serious concerns. First, while LLMs offer
innovative solutions for information retrieval and content creation and have been shown to significantly enhance
individual productivity in a variety of writing and coding tasks, they have also been found to hallucinate, i.e.,
providing 'confidently incorrect' responses to user queries[6], and to undermine worker performance on certain
types of tasks[3]. Second, if individual participation in online communities were to decline, this would imply a
decline in opportunities for all manner of interpersonal interaction, upon which many important activities
depend, e.g., collaboration, mentorship, job search. Further, to the extent a similar dynamic may emerge within
formal organizations and work contexts, it would raise the prospect of analogous declines in organizational
attachment, peer learning, career advancement and innovation[7–12].
With the above in mind, we address two questions in this work. First, we examine the effects that generative
artificial intelligence (AI), particularly large language models (LLMs), have on individual engagement in online
knowledge communities. Specifically, we assess how LLMs influence user participation and content creation in
online knowledge communities. Second, we explore factors that moderate (amplify or attenuate) the effects of
LLMs on participation and content creation at online knowledge communities. By addressing these relationships,
we aim to advance our understanding of the role LLMs may play in shaping the future of knowledge sharing and
collaboration online. Further, we seek to provide insights into approaches and strategies that can encourage a
sustainable knowledge sharing dynamic between human users and AI technologies.
We evaluate our questions in the context of ChatGPT's release, in late November of 2022. We start by exam-
ining how the release of ChatGPT impacted Stack Overflow. We show that ChatGPT's release led to a marked
decline in web traffic to Stack Overflow, and a commensurate decline in question posting volumes. We then
consider how declines in participation may vary across community contexts. Leveraging data on posting activity
in Reddit developer communities over the same period, we highlight a notable contrast: no detectible declines
in participation. We attribute this difference to social fabric; whereas Stack Overflow focuses on pure informa-
tion exchange, Reddit developer communities are characterized by stronger social bonds. Further, considering
heterogeneity across topic domains within Stack Overflow, we show that declines in participation varied greatly

Questrom School of Business, Boston University, Boston, MA 02215, USA. email: gburtch@bu.edu

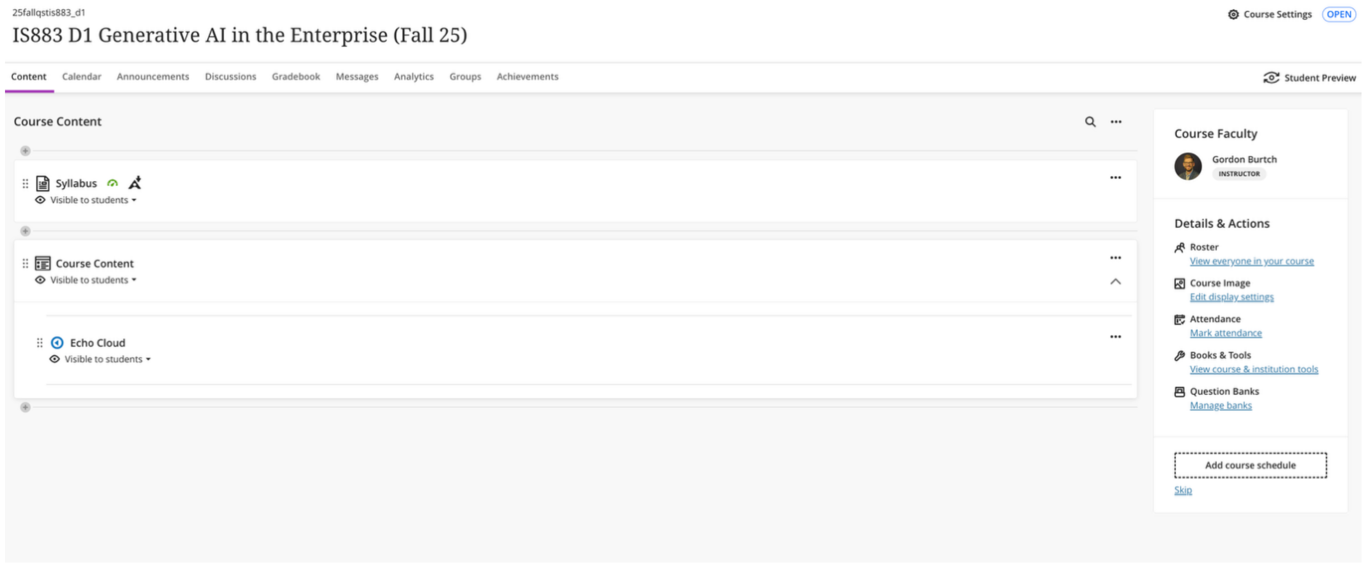
Scientific Reports | (2024) 14:10413 | https://doi.org/10.1038/s41598-024-61221-0

nature portfolio

Course Materials

COURSE WEBSITE

- The course website is on Blackboard – I assume you have looked at it by this point (if you aren't enrolled in the site, let me know)
- You will submit all assignments and receive relevant course announcements via that site.
- I will post lecture materials and in-class exercises / examples via the GitHub Repository linked on Blackboard.



GitHub

GOOGLE COLAB

- All homework and exercises in this course are to be implemented in Python. You should work in Google Colab because we (the TA and I) will not be able to provide technical support if you run into issues with your local Python instance. You also need to submit .ipynb files that run in a clean Colab instance (we won't debug your code!).

Grading and Evaluation

PARTICIPATION / ATTENDANCE

- Regular attendance and participation will be worth 10% of your final grade.
- After each class, I will have you complete an online survey where you will have the opportunity to cite other students who made comments, who raised questions, or who helped you better understand the course material.

Grading Distribution & Scale

(1) Attendance & Participation	10%
(2) Coding Assignment	20%
(3) Case Write-ups (x2)	25%
(4) Quizzes (x4)	20%
(5) Final Project	25%
Total:	100%

INDIVIDUAL CODING ASSIGNMENT

- One assignment worth 20% of your final grade
- Due by 11:59pm on the date indicated in the course schedule; submit your Jupyter (Colab) Notebook file via Blackboard (submit the actual file with code, *not* a link to your notebook).
- Late submissions will result in grade deductions, per the syllabus.

CASE WRITE-UPS

- We will discuss two cases in class, and your submitted individual responses to case questions will comprise 25% of your final grade (12.5% each).

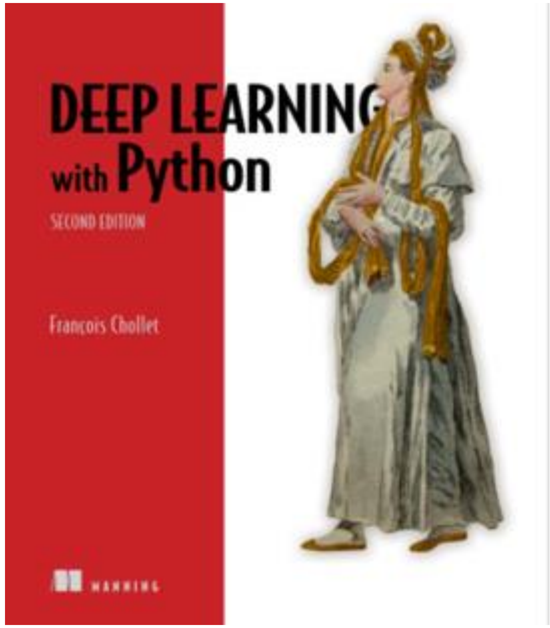
QUIZZES

- We will have four in-class, timed, paper-and-pencil quizzes, collectively worth 20% of your final grade. These are multiple choice quizzes

PAIR PROJECT

- The final project is worth 25% of your final grade. You will work in pairs to implement a neural network-based predictive model that addresses a practical problem of interest to you! You must source a dataset (*not* from Kaggle), motivate the prediction problem (explain why it's practically relevant), and then implement a (predictive) model.

Course Textbook



Chollet, François. (2021). *Deep Learning with Python (2nd Edition)*. Manning Publications Co. **ISBN-13: 978-1617296864.**
https://www.manning.com/books/deep-learning-with-python-second-edition

Required Software

SOFTWARE CONFIGURATION

- You can access Google Colab at https://colab.research.google.com. You probably want to use your BU Google account credentials.

- You may resort to a local Python implementation and/or virtual machine setup for your final project, depending on the scale of your data/problem and what you are comfortable doing. Note that the TA and I cannot provide technical support should you choose to do that.

Why Tensorflow + Keras?

Figure 1.12 Machine learning tools used by top teams on Kaggle

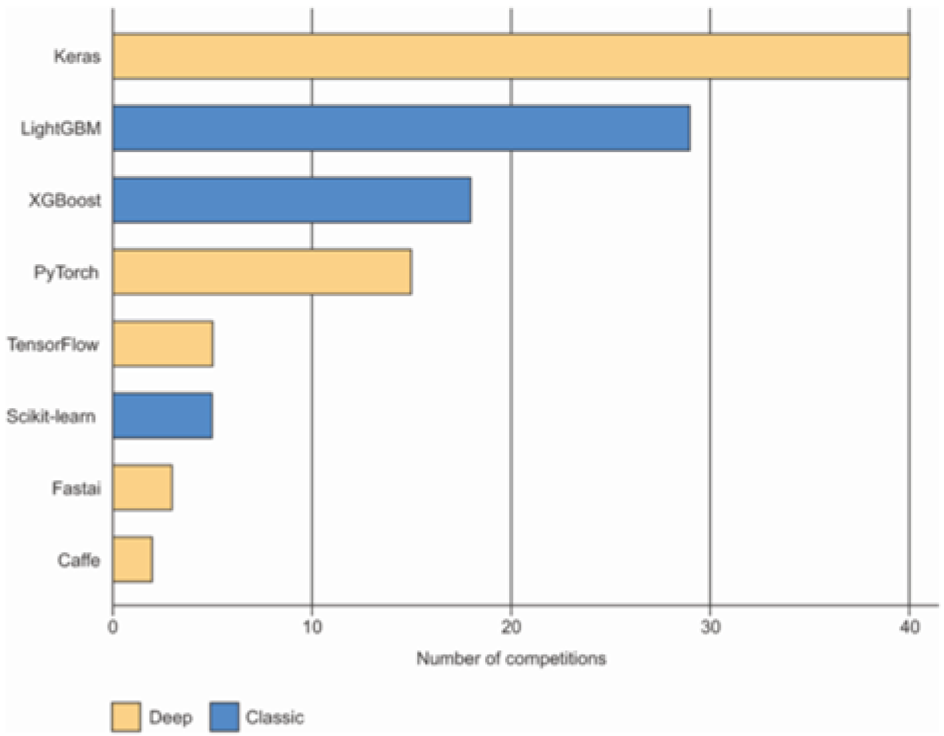
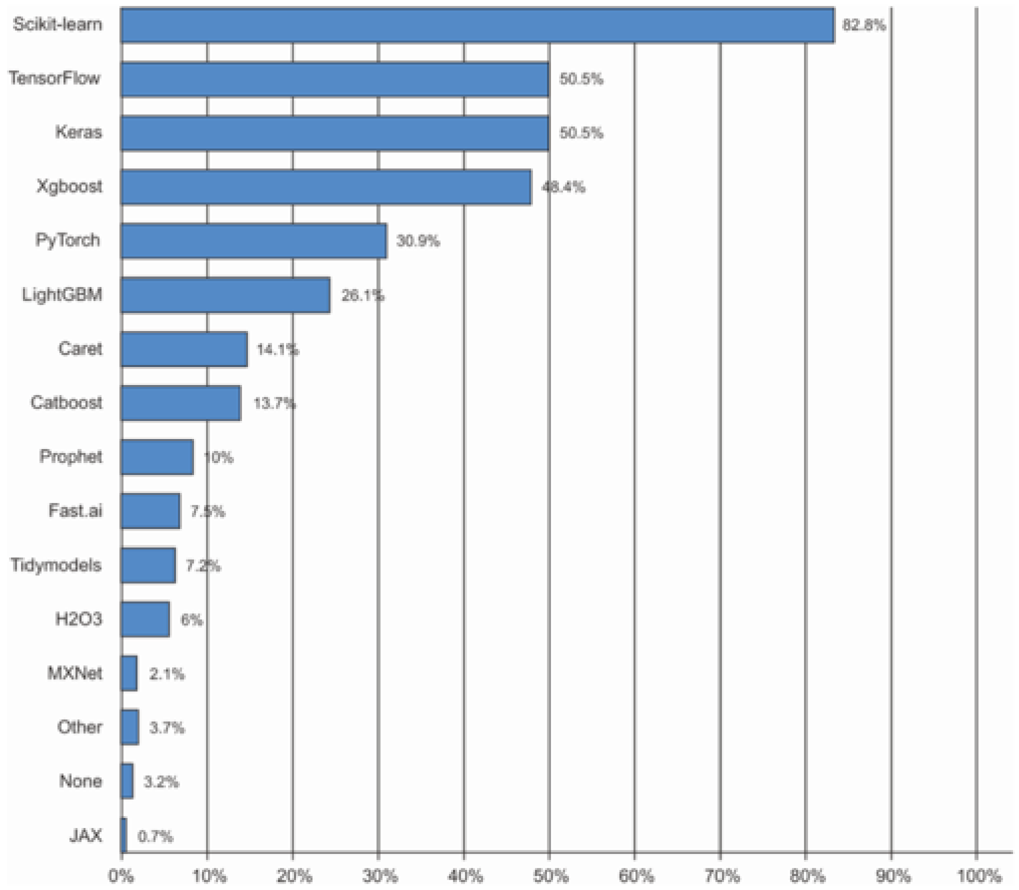
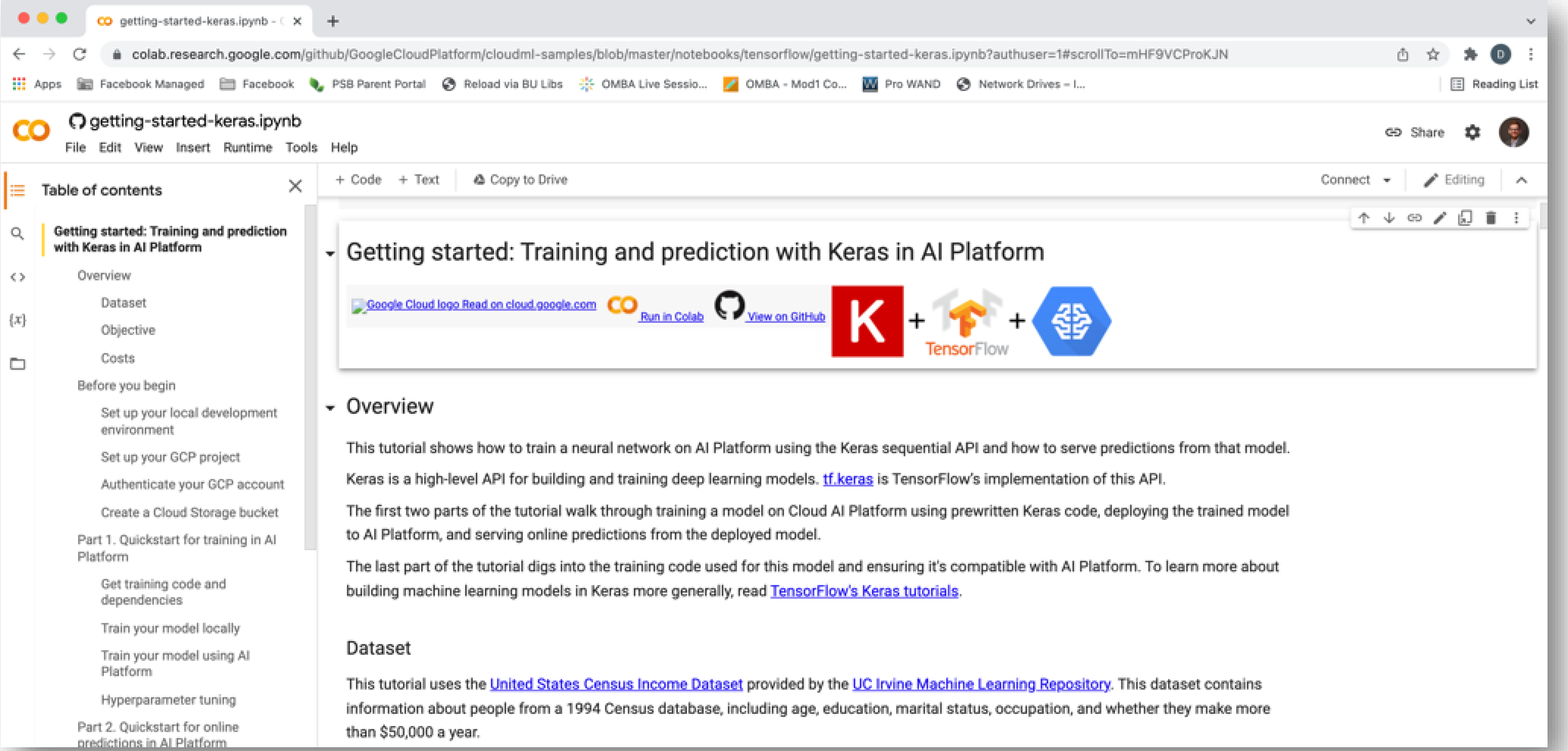


Figure 1.13 Tool usage across the machine learning and data science industry (Source: www.kaggle.com/kaggle-survey-2020)



Google Colab



Course Timeline

AGENDA

- We will start with the basic math concepts.
- We will then get into neural networks for simple prediction problems using structured data as inputs.
- Finally, we will explore more complex use cases, including those with unstructured inputs (e.g., text, audio, images/video), and ultimately, unstructured output (e.g., generative AI).

NOTE TIMING OF DELIVERABLES

- I will announce sign-ups for the first project proposal check-in meeting on October 6.
- Project proposal.
- Coding assignment focused on basic prediction.
- Case write-ups (x2).
- Final project deliverables.

Date	Topic	Assignments	Readings
September 8	**Course Introduction**	--	Chapter 1 & Overview of NNs
September 15	**The Math of NNs**	--	Chapter 2
September 22	**Keras and Other Frameworks**	--	Chapter 3
September 29	**Simple NNs & Model Tuning**	Coding Assignment Posted	Chapter 4
October 6	**Project Check-in Meetings (15 Minutes Each)**	--	Chapters 5 and 6 (Optional)
October 14 (Tuesday!)	**Functional API & Computer Vision (CNNs)**	Project Proposal Due (Friday Oct 17, 11:59pm)	Chapters 7 and 8 (Play Quick Draw)
October 20	**Computer Vision (cont.)**	Coding Assignment Due (Friday Oct 24, 11:59pm)	Chapter 9
October 27	**Text Models (RNNs, Embeddings)**	--	Chapters 10 and 11
November 3	**Model Understanding (SHAP & LIME)**	--	--
November 10	**Generative AI**	--	Agentic Beer Game
November 17	**Generative AI (cont.) + Case: JPMorganChase**	JPMorganChase Due Friday at 11:59 pm	JPMorganChase Case
November 24	**Finish Up Group Project (Optional Class)**	Thanksgiving!	--
December 1	**Guest Speaker (Tentative) + Case: AI Wars**	AI Wars Due Friday at 11:59 pm	AI Wars Case
December 8	**Final Project Presentations**	Project Due 11:59pm	--

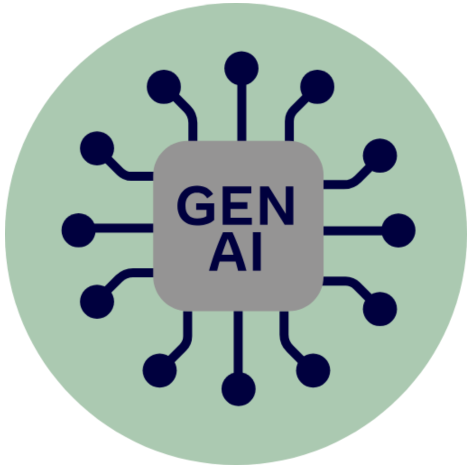
Course Format

LECTURES

- We will meet once each week for ~2.5 hours. I will incorporate a 20-minute break midway through each session. The first half of each session will kick off with a collective discussion of relevant news, followed by lecture (focused on concepts/explanation), at least until we get to case discussions later in the course.

HANDS-ON EXAMPLES

- The second half of each session will be focused on walking through hands-on examples and demonstrations in Python notebooks. I will provide Jupyter Notebooks in Google Colab and data sets (typically shared via GitHub), which we will walk through together.
- You are encouraged to ask questions as we progress. The quizzes, individual assignment, and exam will be based on the in-class material; I will never test you on things that were not discussed in class.



Policy

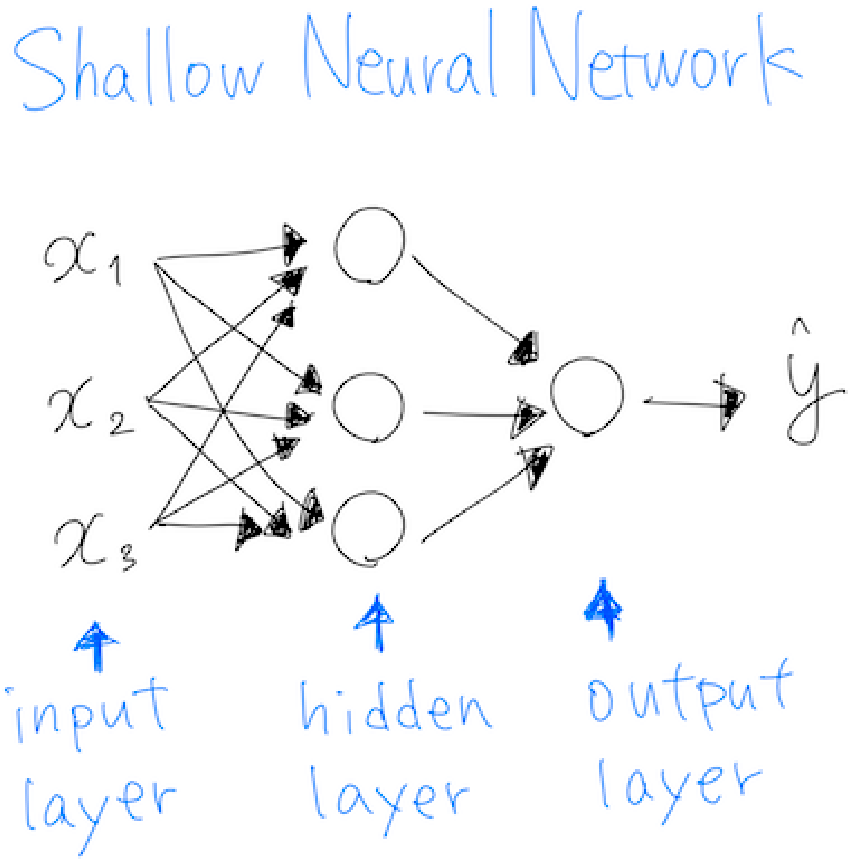
Generative AI Policy: I expect you to use generative AI tools during this course. However, the way you use them matters. Some valid use cases include implementing data munging tasks that you already understand, based on past coursework, such as pre-processing text, automatically generating comments for your code, or requesting an explanation for what some example code does. However, you should try to implement the code yourself and work through any errors that arise. If you 'vibe-code' your way through the course, you are likely to misunderstand concepts and do poorly on evaluations. I reserve the right to have you explain your code to me orally, in person, should you employ concepts or implementations that we did not cover in the course. An inability to explain what your code is doing will result in a 20 percentage point reduction in the deliverable grade. The same will occur if you submit code employing PyTorch as the backend.

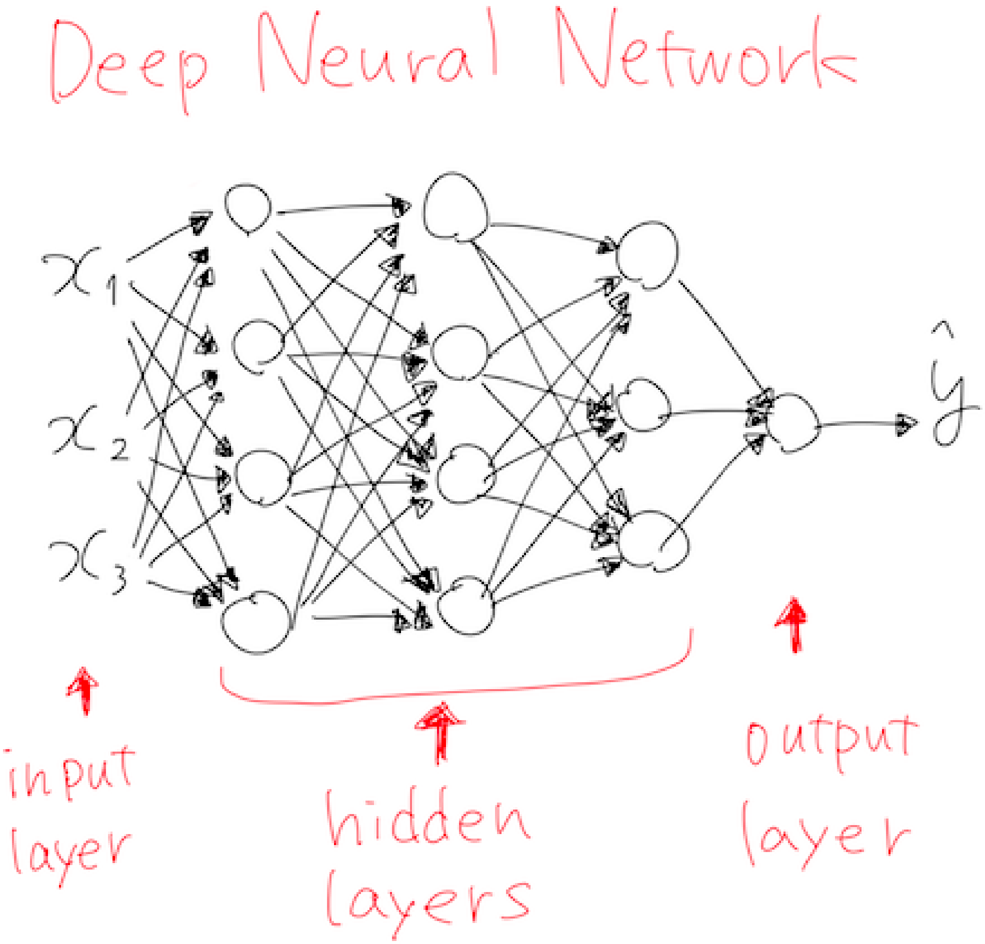
Any Questions?

What is ‘Deep’ (vs. Shallow) Learning?

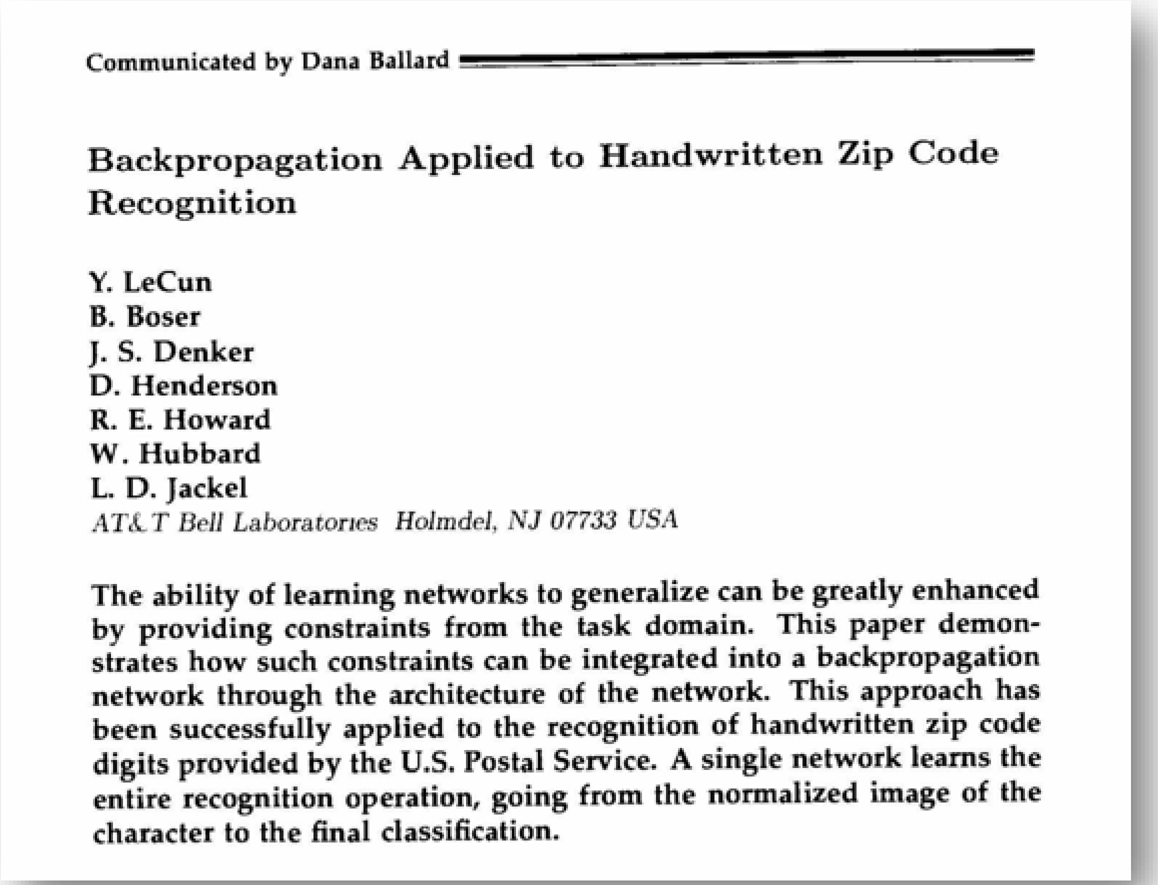
https://chat.openai.com/chat

What is ‘Deep’ Learning?





Where Deep Learning Started



Communicated by Dana Ballard

Backpropagation Applied to Handwritten Zip Code Recognition

Y. LeCun
B. Boser
J. S. Denker
D. Henderson
R. E. Howard
W. Hubbard
L. D. Jackel
AT&T Bell Laboratories Holmdel, NJ 07733 USA

The ability of learning networks to generalize can be greatly enhanced by providing constraints from the task domain. This paper demonstrates how such constraints can be integrated into a backpropagation network through the architecture of the network. This approach has been successfully applied to the recognition of handwritten zip code digits provided by the U.S. Postal Service. A single network learns the entire recognition operation, going from the normalized image of the character to the final classification.

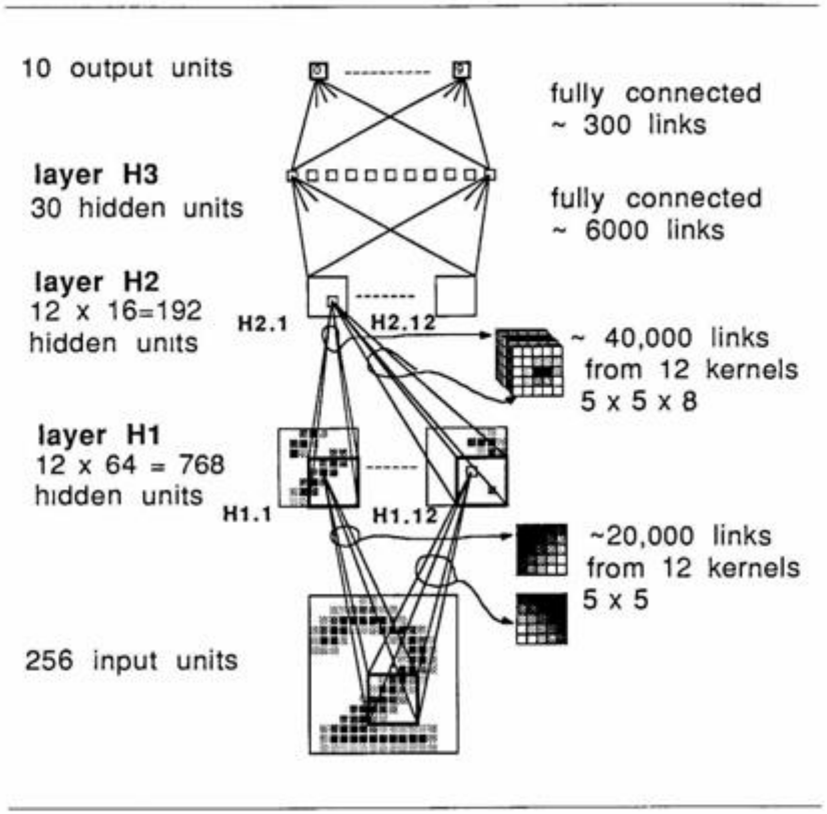
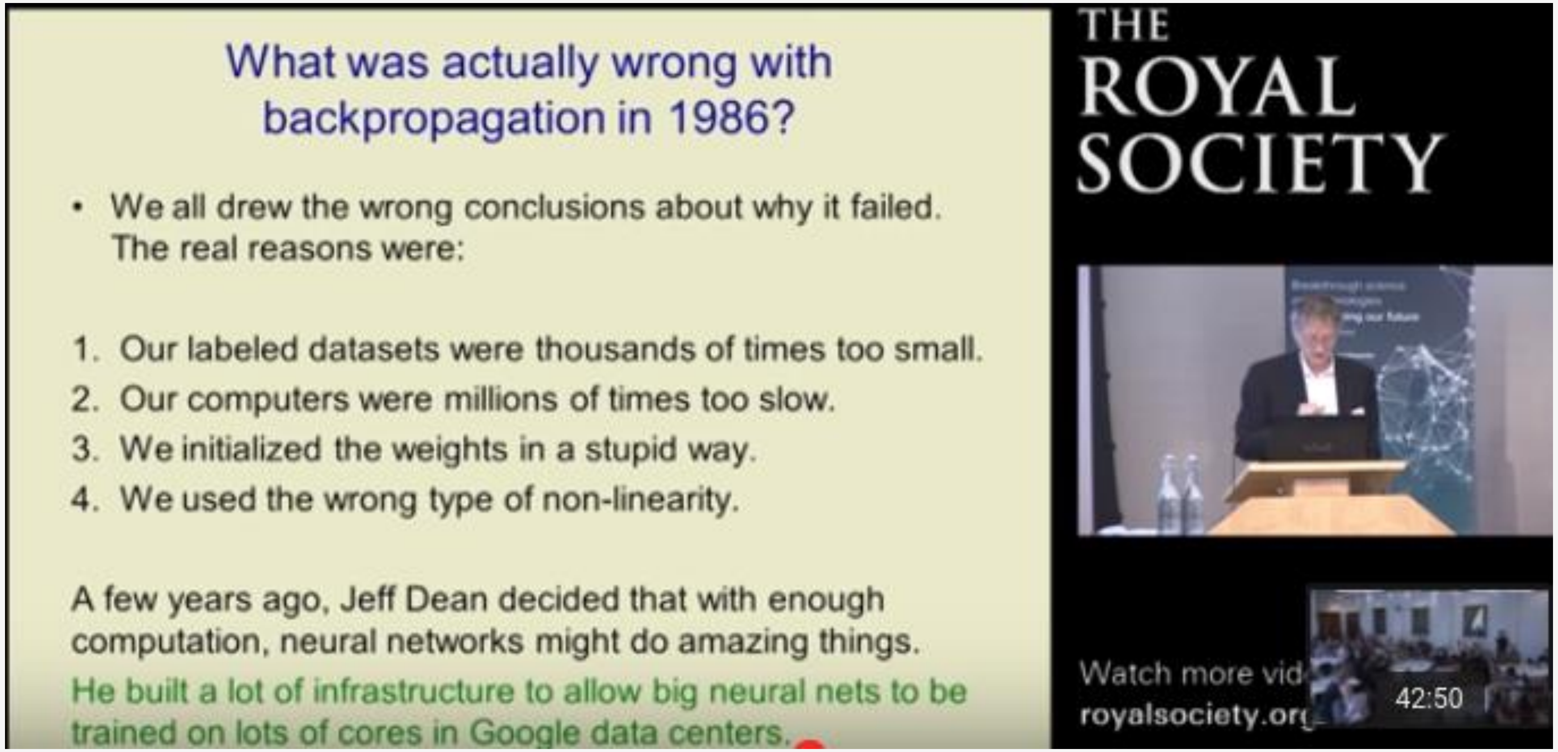


Figure 3 Log mean squared error (MSE) (top) and raw error rate (bottom) versus number of training passes

Then It Shuffled Along for Decades…

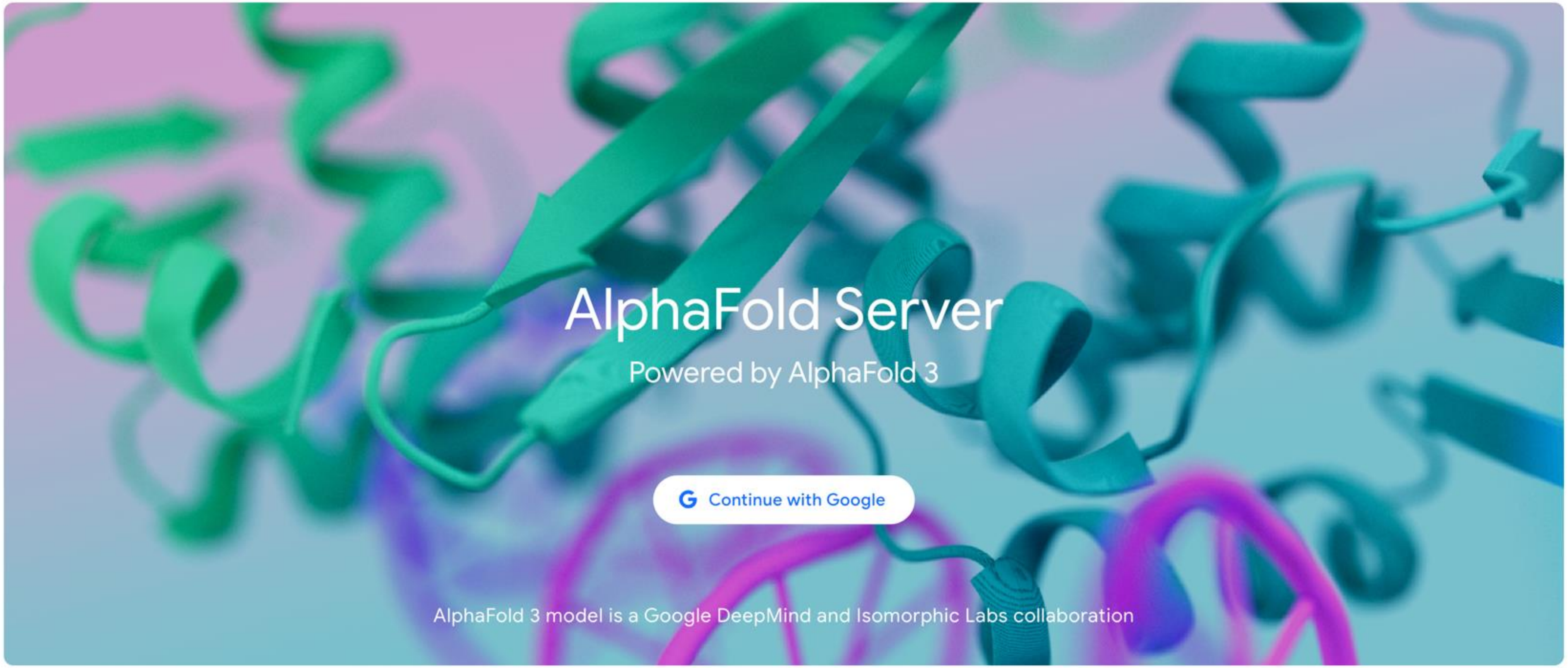


Now…

THE SHIFT

An A.I.-Generated Picture Won an Art Prize. Artists Aren't Happy.

"I won, and I didn't break any rules," the artwork's creator says.



How does AlphaFold Server work?

AlphaFold Server is a web-service that can generate highly accurate biomolecular structure predictions containing proteins, DNA, RNA, ligands, ions, and also model chemical modifications for proteins and nucleic acids in one platform. It's powered by the newest AlphaFold 3 model.

GitHub Copilot

Learn more >

Technical Preview

Your AI pair programmer

With GitHub Copilot, get suggestions for whole lines or entire functions right inside your editor.

Sign up >

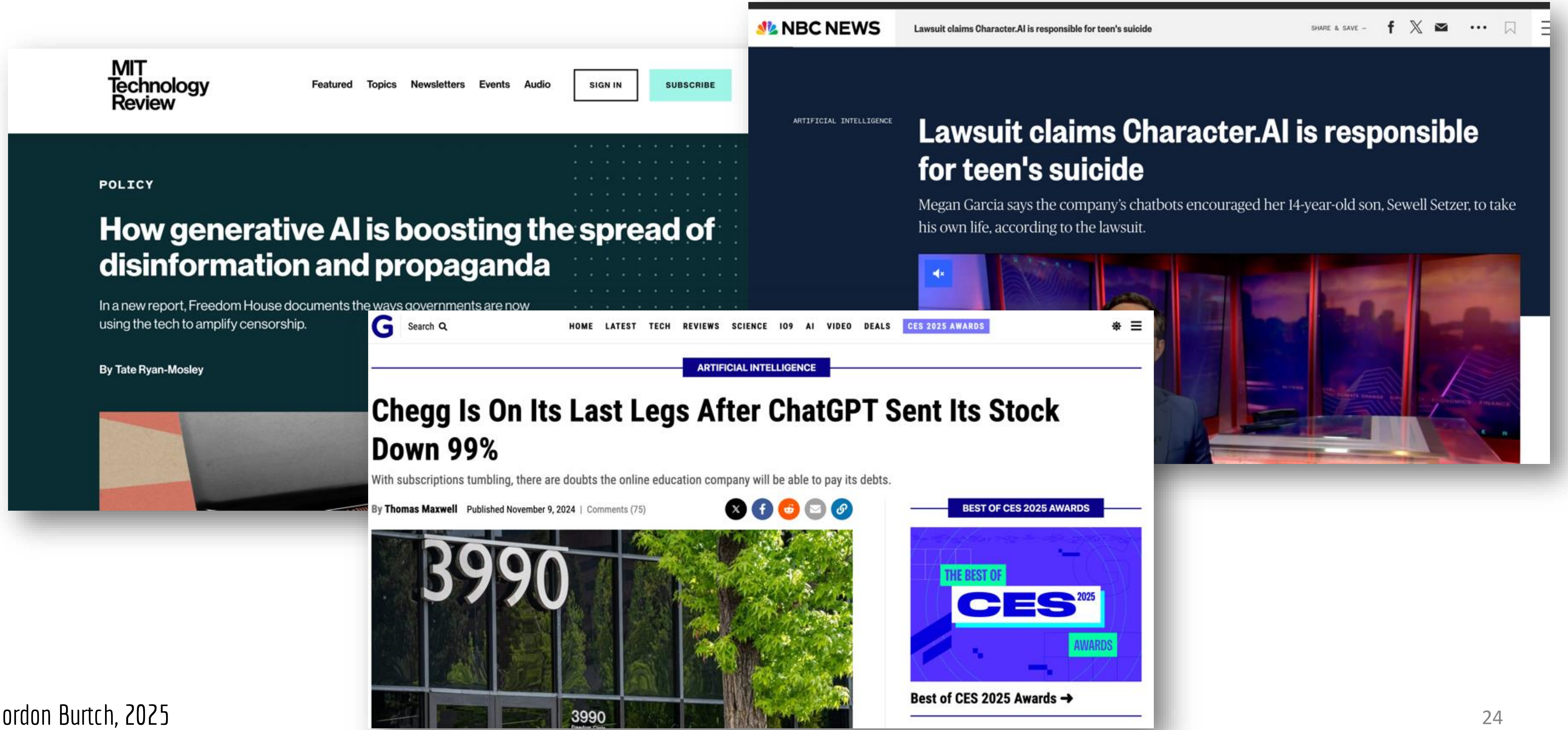
sentiment.ts | write_sql.go | parse_expenses.py | addresses.rb

```
#!/usr/bin/env ts-node

import { fetch } from "fetch-h2";

// Determine whether the sentiment of text is positive
// Use a web service
async function isPositive(text: string): Promise<boolean> {
  const response = await fetch(`http://text-processing.com/api/sentiment/`, {
    method: "POST",
    body: `text=${text}`,
    headers: {
```

These Technologies Bring New Problems



Business Process / Model Still Matters

Bloomberg Opinion

Money Stuff

Sorry, Zillow's Computer Can't Buy Your House Right Now

Also CEO pay, the Boredom Markets Hypothesis and Big Short guys being big short.

By Matt Levine +Sign Up
October 18, 2021, 1:18 PM EDT

Matt Levine is a Bloomberg Opinion columnist covering finance. He was an editor of Dealbreaker, an investment banker at Goldman Sachs, a mergers and acquisitions lawyer at Wachtell, Lipton, Rosen & Katz, and a clerk for the U.S. Court of Appeals for the 3rd Circuit.

Zillow

Deciding how much you should pay for a share of large-cap publicly traded stock is not an *entirely* solved problem, but it's pretty close. If someone comes to you and says "hey I have 100 shares of Microsoft Corp. stock for sale, how much will you pay me for it," a pretty decent answer would be to look at the last price at which Microsoft traded – like a millisecond ago – and subtract, you know, one cent from that price. That will get you a price that is likely to be competitive (the seller might actually sell to you), likely to be profitable (you might be able to sell it for more than you paid), and

// Menu > Institutional > Tools About Archive Events

May 4, 2023

Google "We Have No Moat, And Neither Does OpenAI" // Leaked Internal Google Document Claims Open Source AI Will Outcompete Google and OpenAI

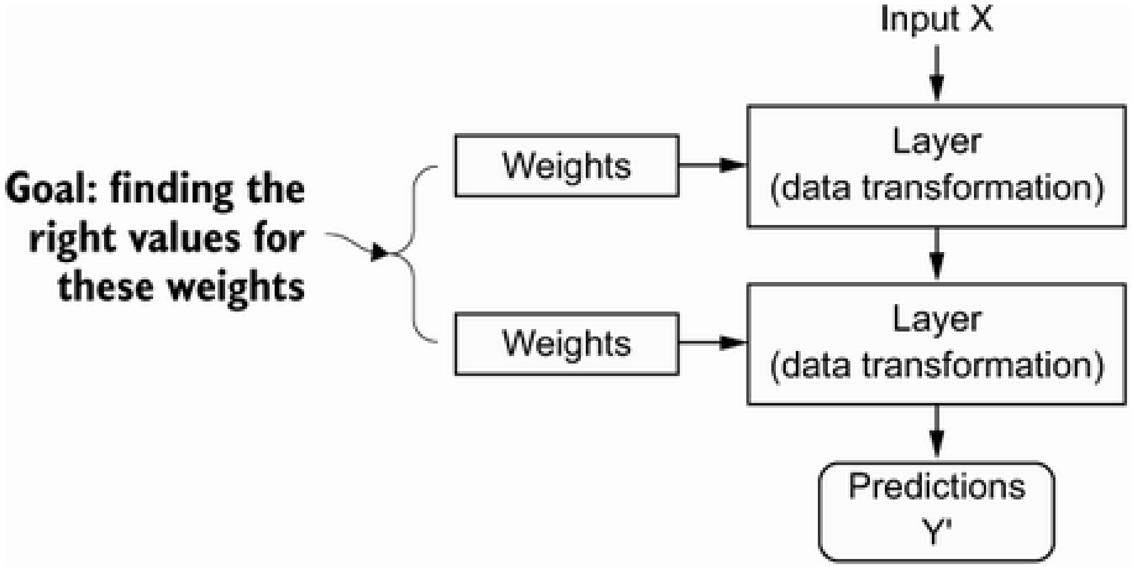
10 minutes

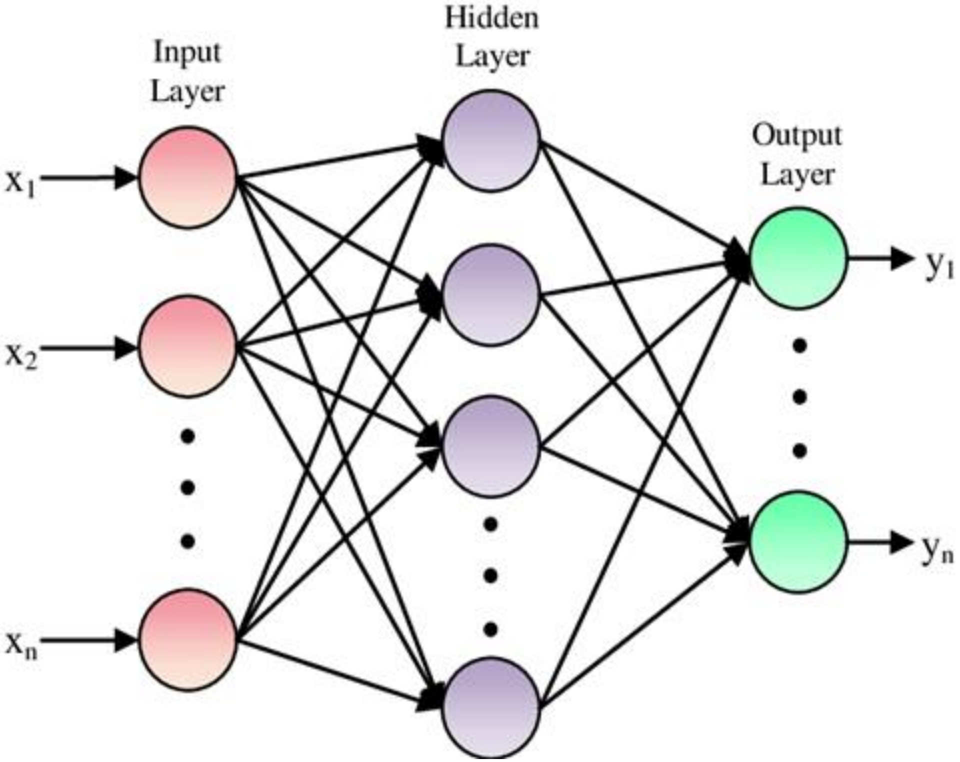
11 comments

By Dylan Patel and Afzal Ahmad

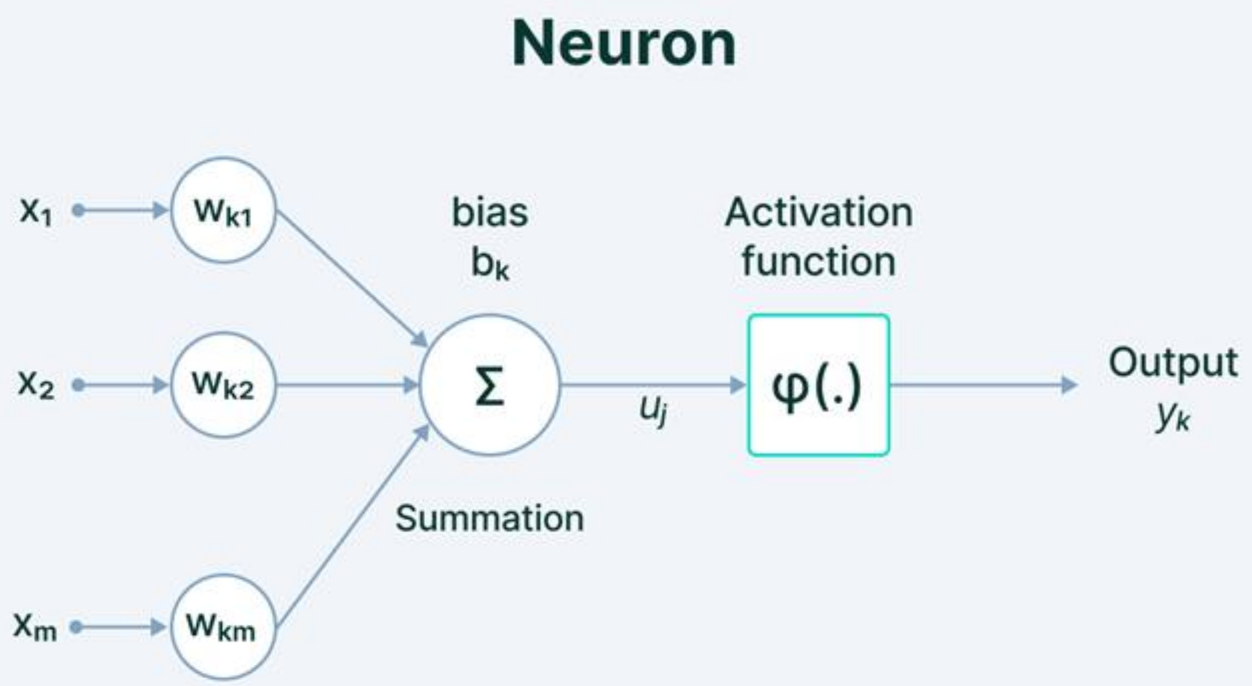
How It Works, Conceptually

Model Parameters

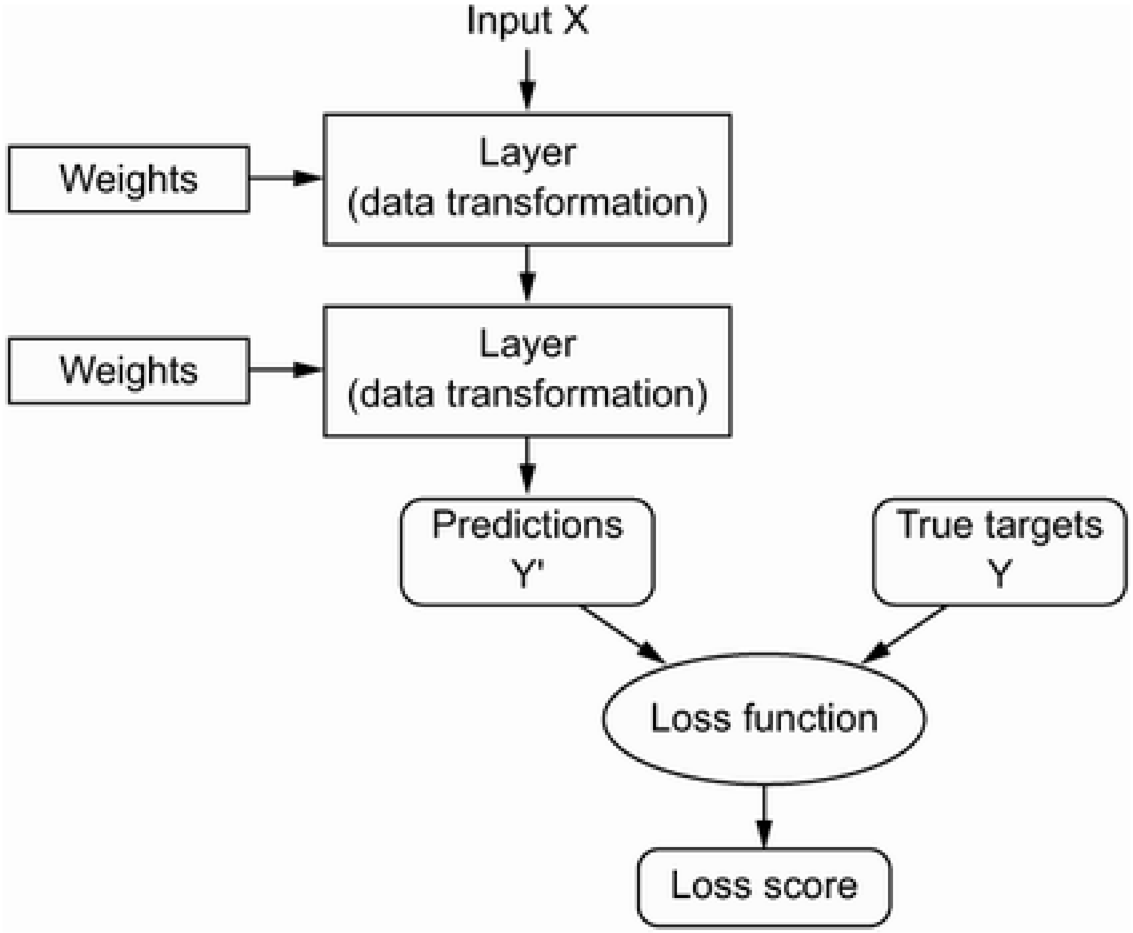




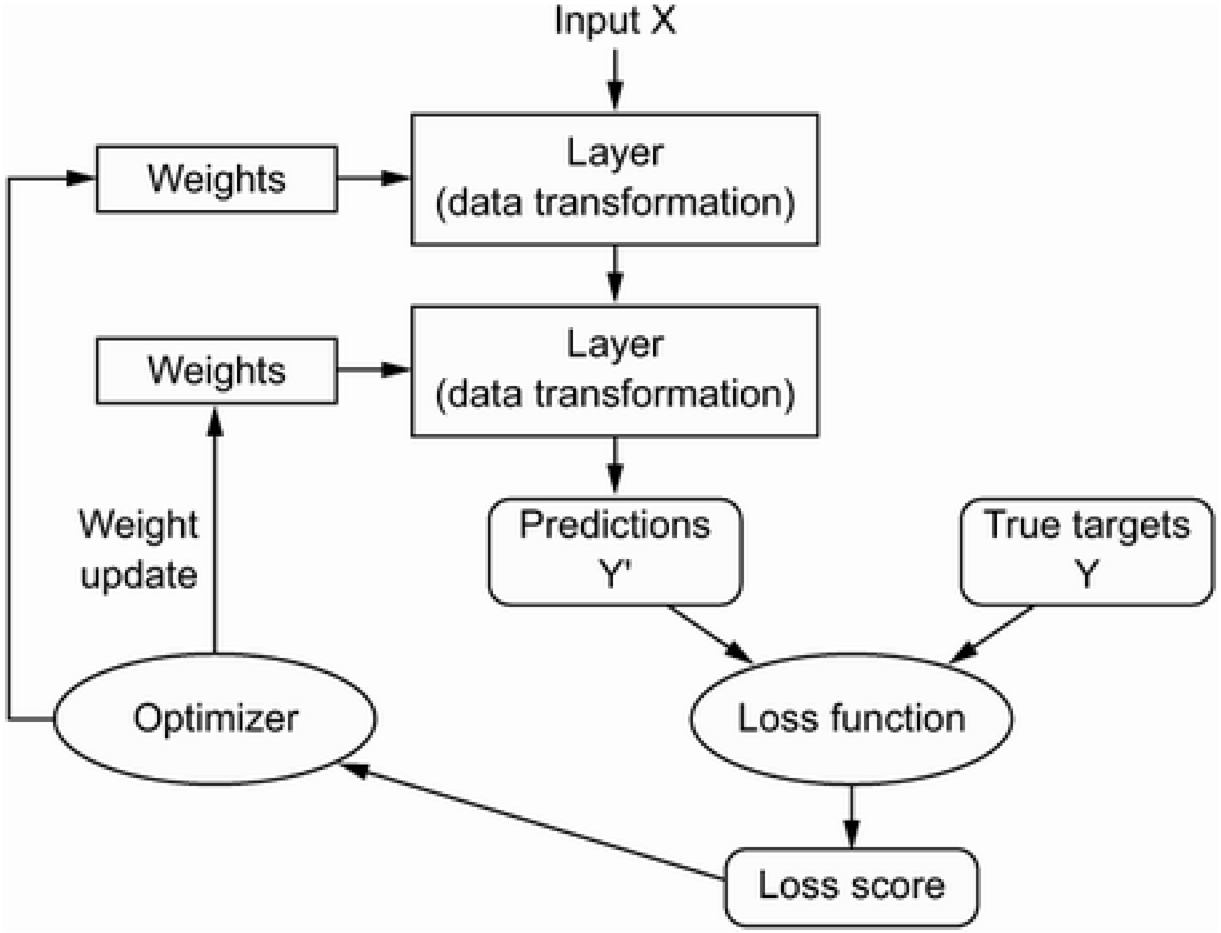
Model Parameters



Loss Function (Error)



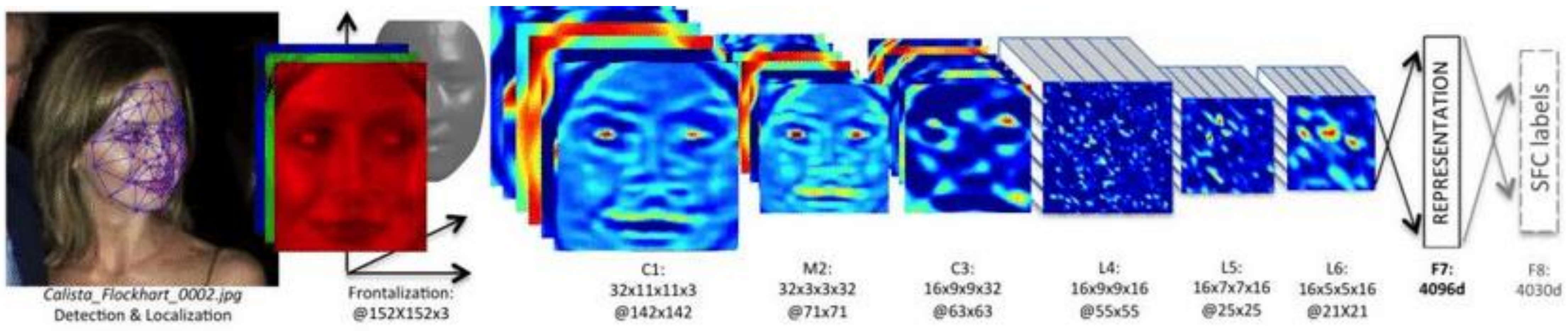
Optimization



When to Learn Deeply (vs. Not)

COMPLEX RELATIONSHIPS

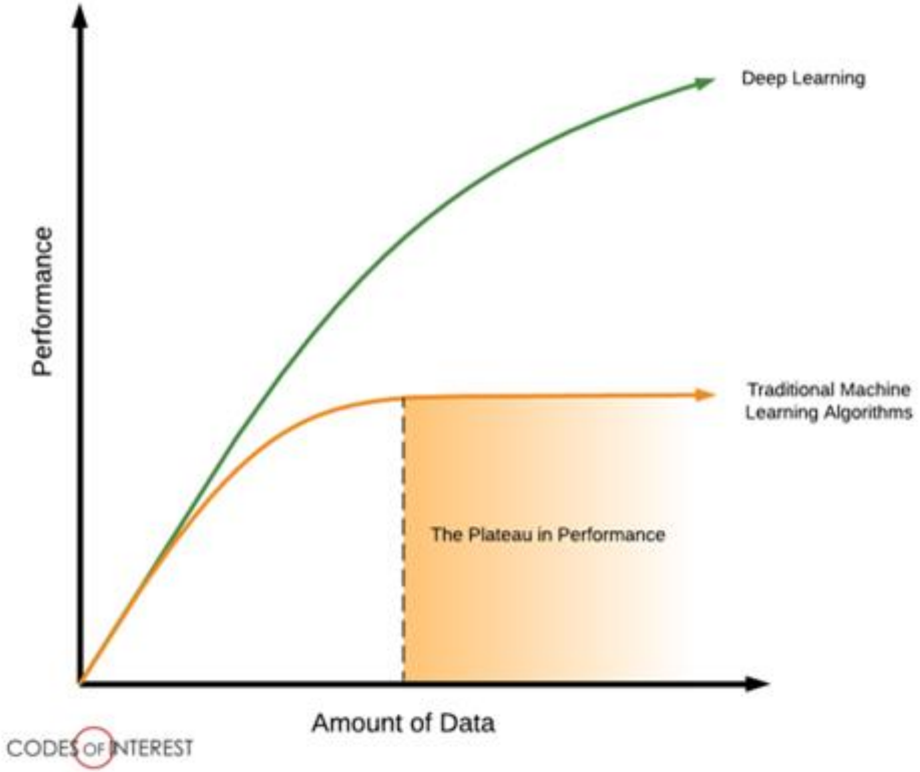
- Complex, non-linear, interactive relationships and mappings; common use cases involve unstructured (high dimensional) data. Deep learning techniques remove the need for feature engineering, a daunting task.



When to Learn Deeply (vs. Not)

LOTS OF DATA ON HAND

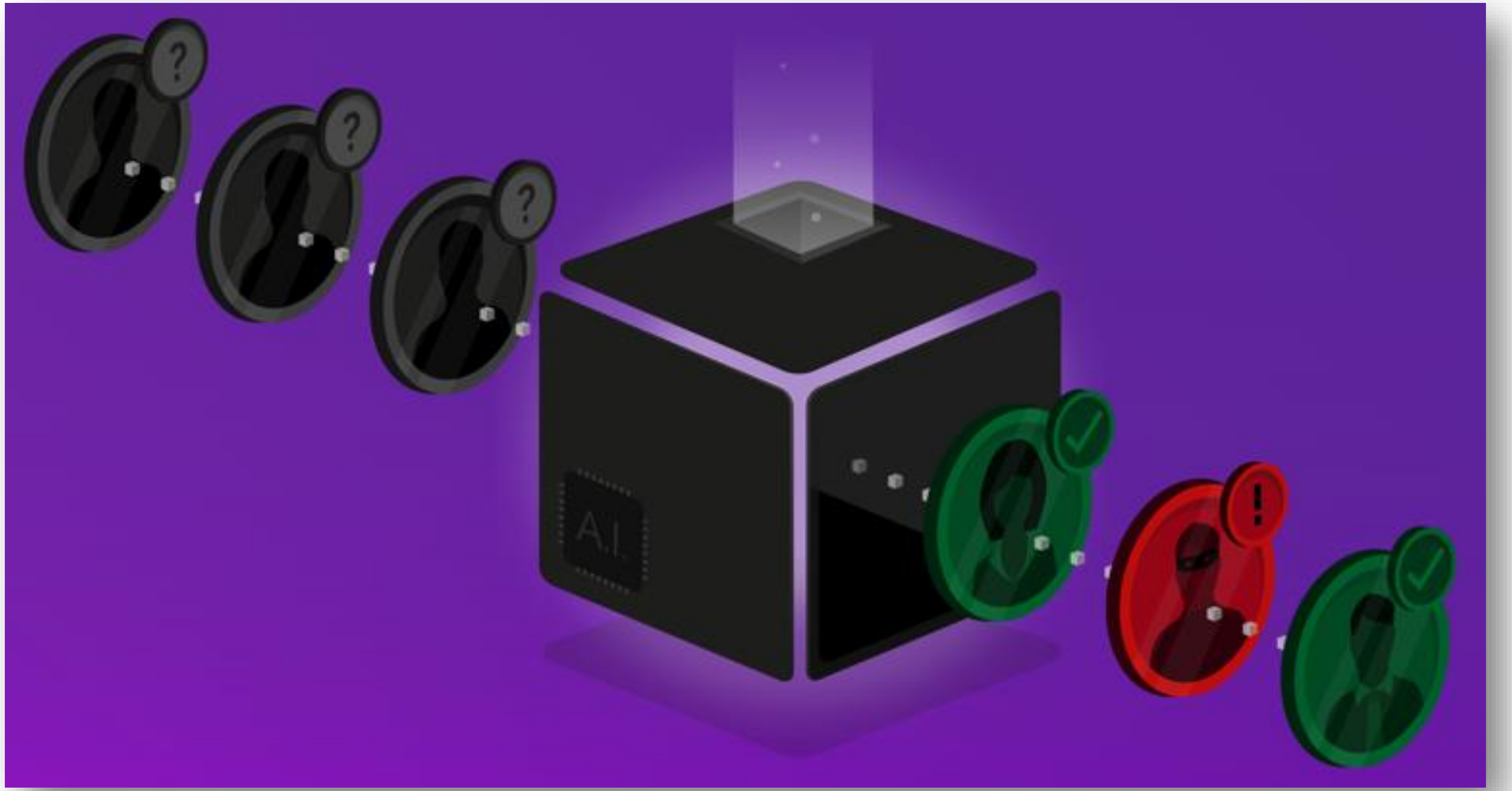
- To be able to learn those complex mappings, typically requires many, many, many training examples.



When to Learn Deeply (vs. Not)

LITTLE NEED FOR UNDERSTANDING

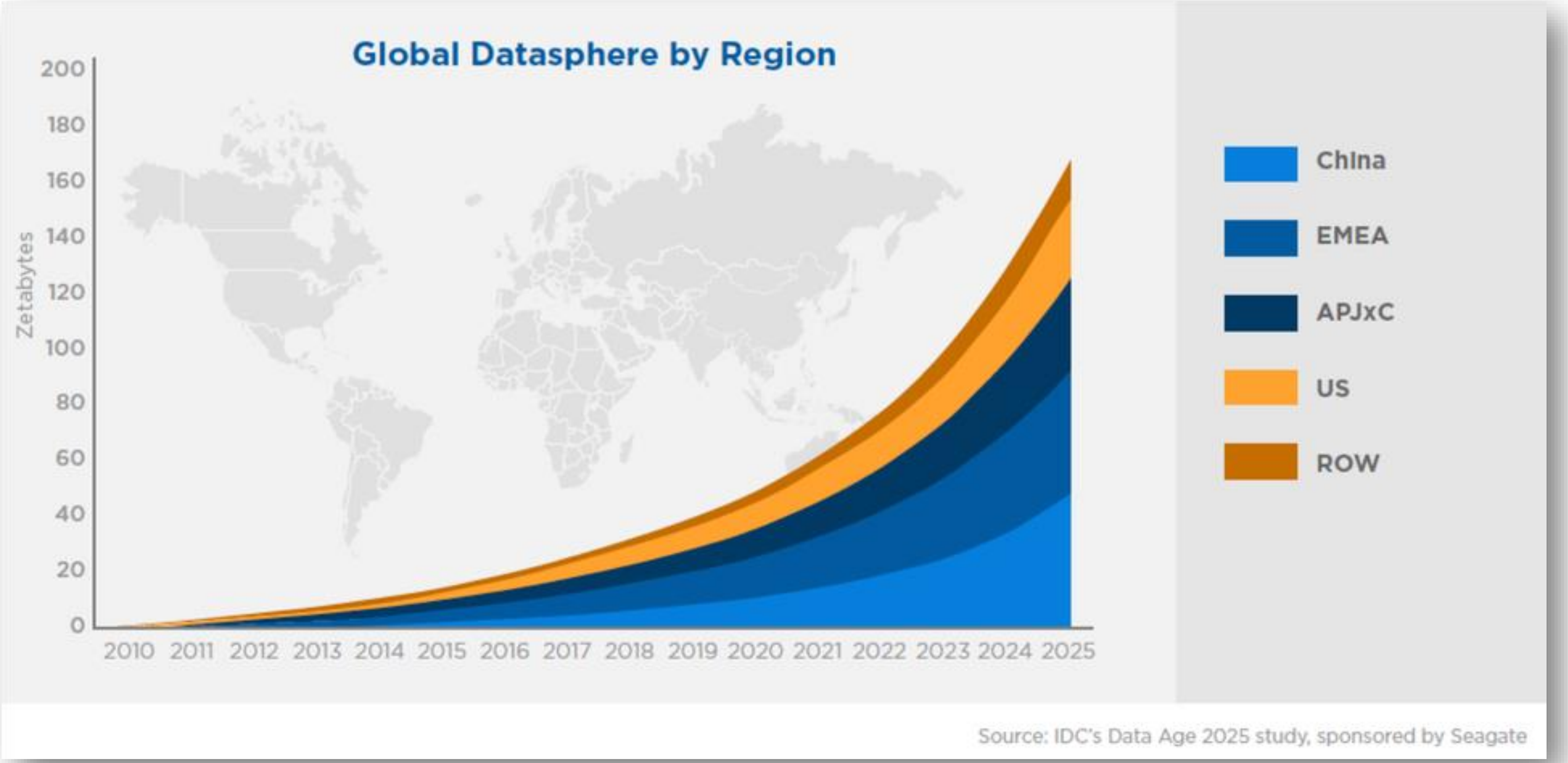
- Although there have been advancements in explainable and interpretable AI, deep nets are notoriously “black box” algorithms.



Why Did Deep Learning Take Off?

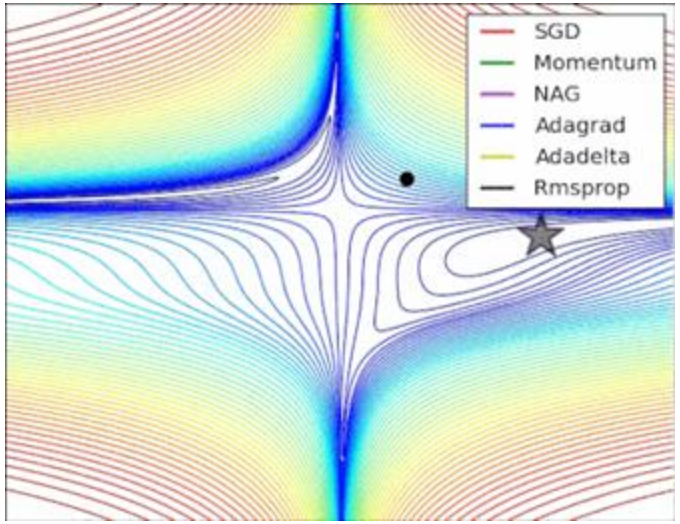
Video Games

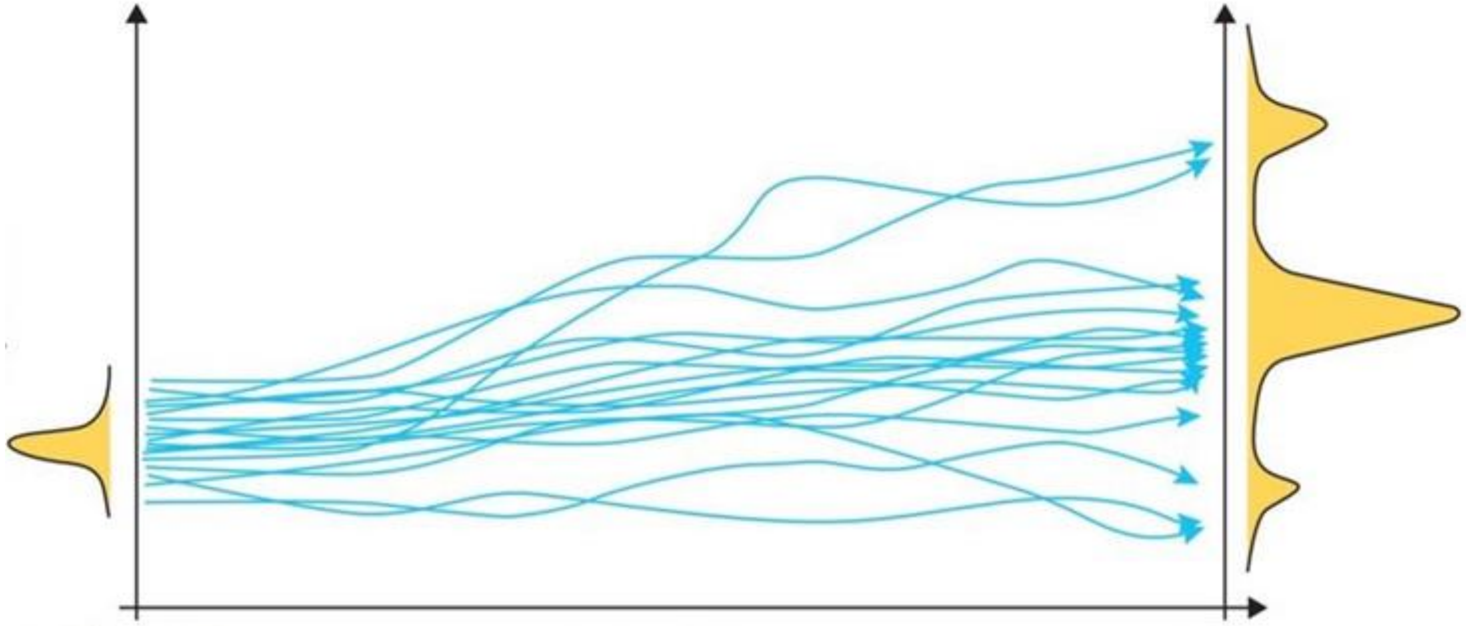
Data

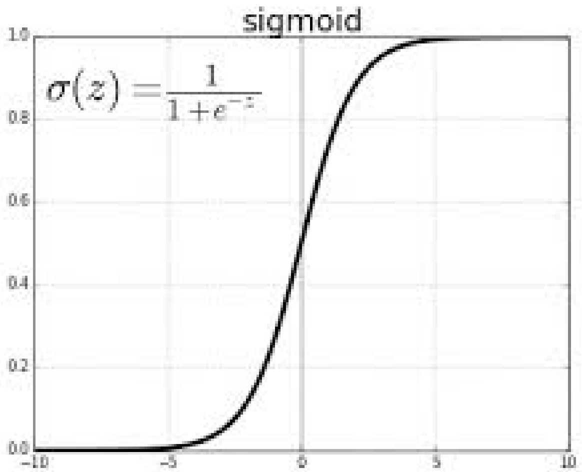


Global Datasphere by Region
Zetabytes
200
180
160
140
120
100
80
60
40
20
0
2010 2011 2012 2013 2014 2015 2016 2017 2018 2019 2020 2021 2022 2023 2024 2025
China
EMEA
APJxC
US
ROW
Source: IDC's Data Age 2025 study, sponsored by Seagate

Algorithmic Improvements







Questions?

Quick Walk-Through of Google Colab