

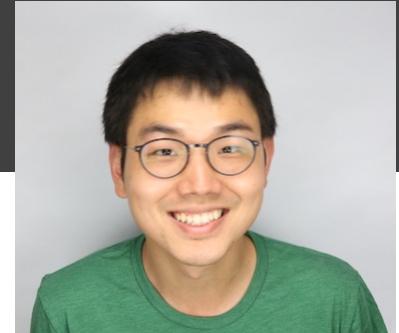
DSC 102

Systems for Scalable Analytics

Fall 2023

Haojian Jin

About Myself



Haojian Jin (<http://haojianj.in/>)

Asst. Prof @ UCSD-HDSI

Data Smith Lab:

We study the security and privacy of data systems by researching the people who design, implement, and use these systems.

Ph.D. from CMU Human-Computer Interaction Institute

Ph.D. Thesis: Modular Privacy Flow

Before Ph.D.: worked at Yahoo Research, ran a startup

My Current Research

*We study the **security** and **privacy** of data **systems** by researching the **people** who **design**, **implement**, and **use** these systems.*

HCI, Software Engineering, Data Science, Mobile Computing, AI.

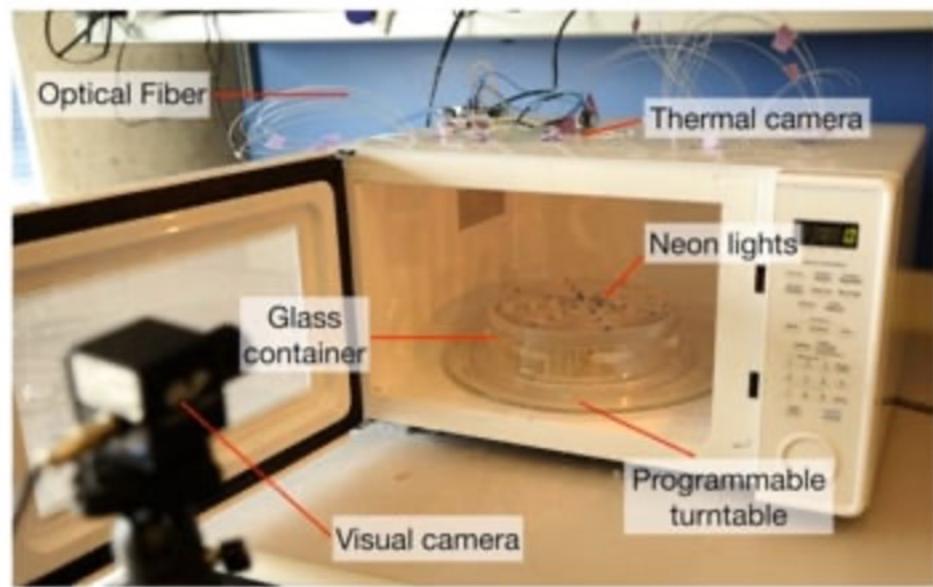
Build cool | fun | sexy systems!

Working Code Trumps All Hype!

Software Defined Cooking (SDC) using a microwave oven

Haojian Jin
Jingxian Wang
Swarun Kumar
Jason Hong

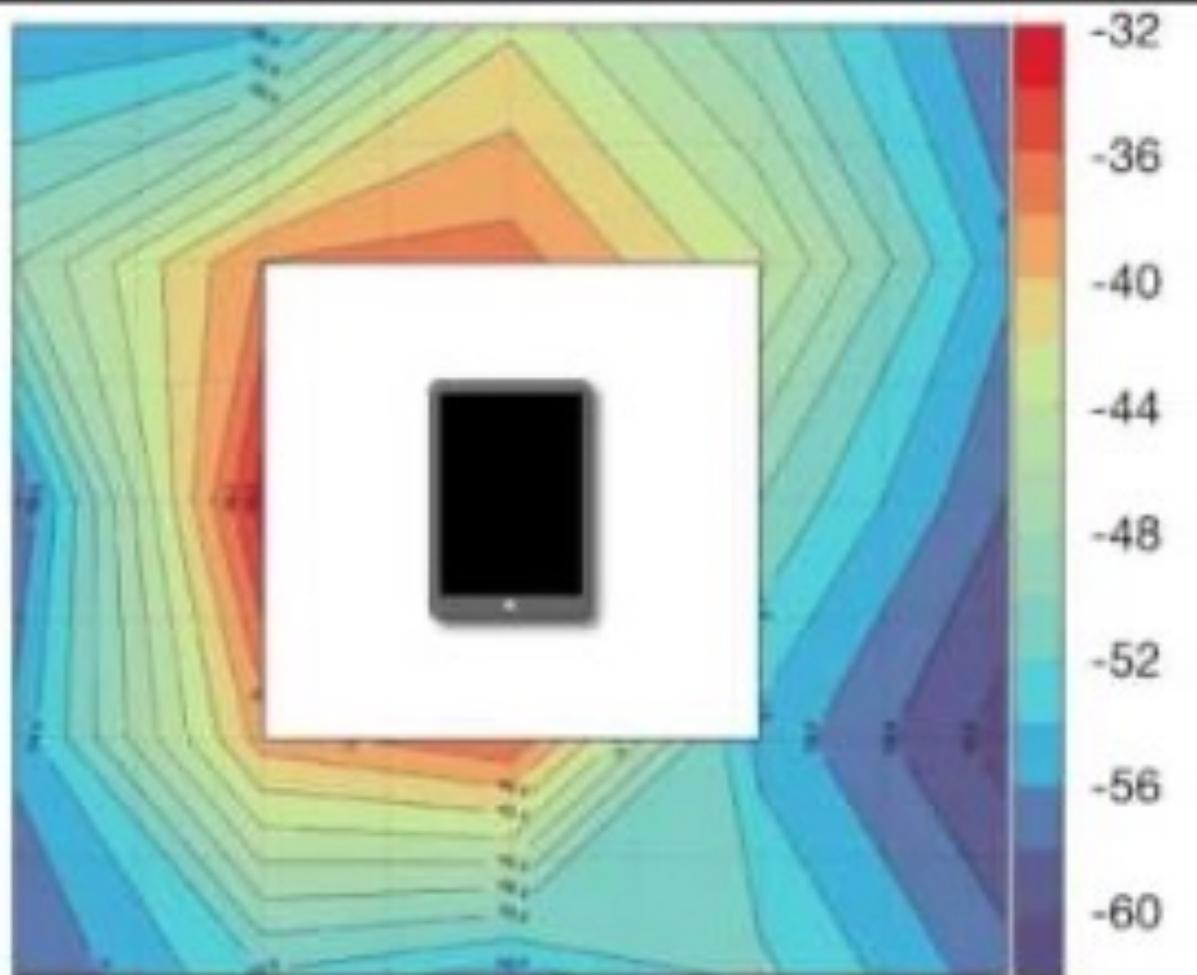
**Carnegie
Mellon
University**



Working Code Trumps All Hype!



Working Code Trumps All Hype!



And mapped the
asymmetric RSSI
distribution around
the iPad.

Working Code Trumps All Hype!

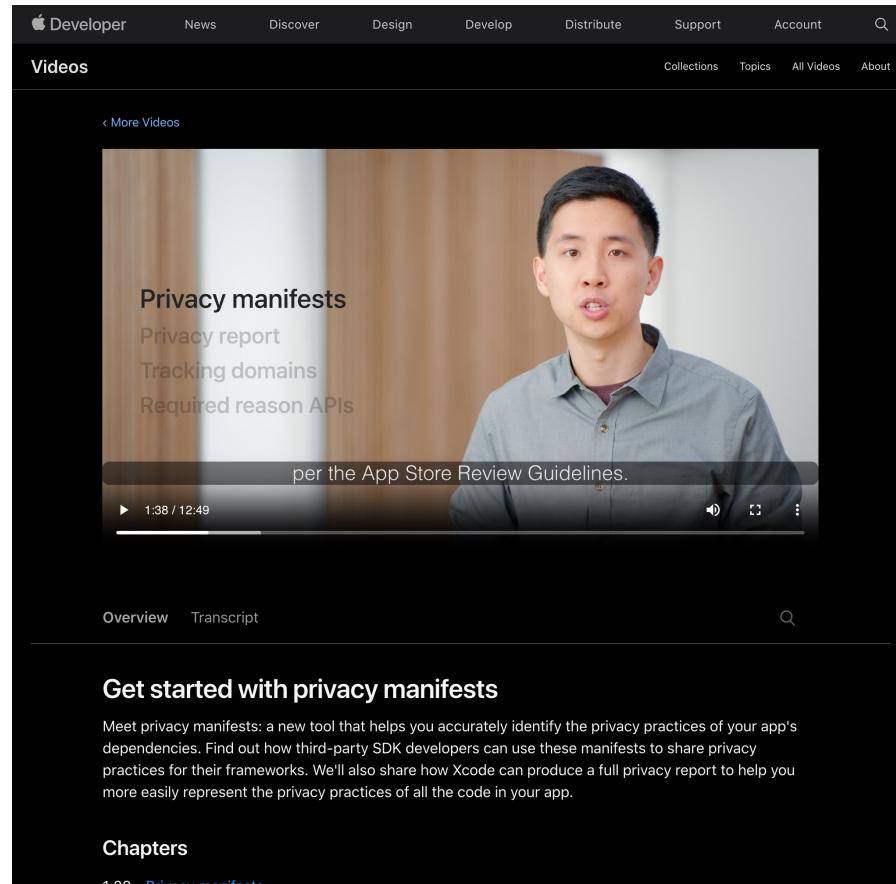
**“Uber” Would Like to Use
Your Location.**

Uber picks you up exactly where you are.
To start riding, choose “Allow” so the
app can find your location.

Don't Allow

OK

Working Code Trumps All Hype!



The screenshot shows a video player interface from the Apple Developer website. At the top, there's a navigation bar with links for Apple Developer, News, Discover, Design, Develop, Distribute, Support, Account, and a search icon. Below that is a sub-navigation bar for 'Videos' with links for Collections, Topics, All Videos, and About. The main content area features a video player with a play button, a progress bar showing 1:38 / 12:49, and a video frame showing a man speaking. On the left side of the video frame, there's a sidebar with text: 'Privacy manifests', 'Privacy report', 'Tracking domains', and 'Required reason APIs'. Below the video frame, there's a caption: 'per the App Store Review Guidelines.' At the bottom of the video player, there are 'Overview' and 'Transcript' buttons and a search icon.

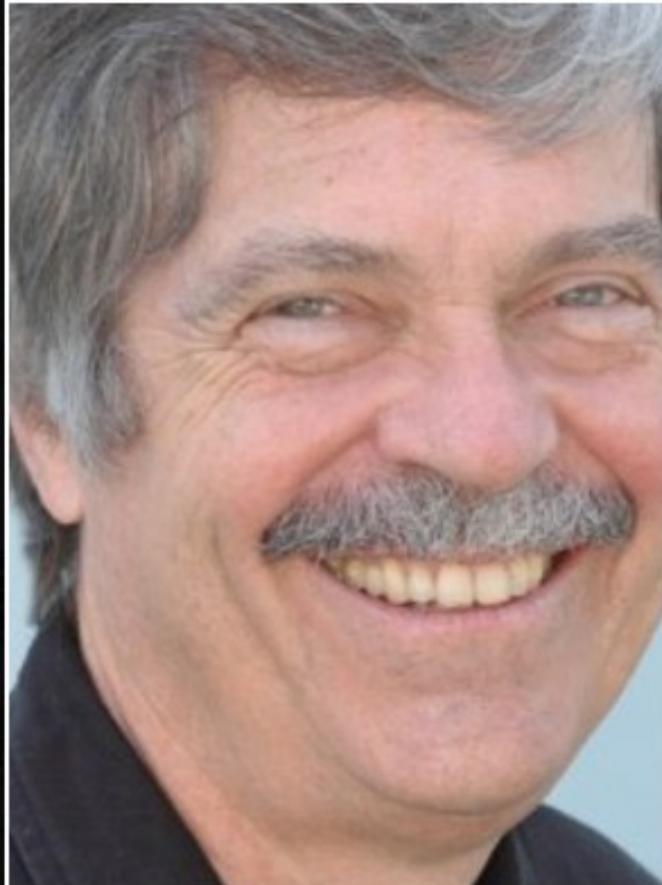
Get started with privacy manifests

Meet privacy manifests: a new tool that helps you accurately identify the privacy practices of your app's dependencies. Find out how third-party SDK developers can use these manifests to share privacy practices for their frameworks. We'll also share how Xcode can produce a full privacy report to help you more easily represent the privacy practices of all the code in your app.

Chapters

1:28 – Privacy manifests

- ✓ Categorized Purpose string (2017 -> 2022)
- ✓ Declared manifests (2020->2024)
- ✓ Operator-based API (??)

A close-up portrait of Alan Kay, an elderly man with grey hair and a prominent grey mustache, smiling warmly at the camera.

The best way to predict the future is
to invent it.

— *Alan Kay* —

What is this course about? Why take it?

the 21st century is the age of ...???

1. Netflix's “spot-on” recommendations

NETFLIX ORIGINAL **STRANGER THINGS**

95% Match 2017 2 Seasons 4K Ultra HD 5.1

When a young boy vanishes, a small town uncovers a mystery involving secret experiments, terrifying supernatural forces and one strange little girl.

Winona Ryder, David Harbour, Matthew Modine
TV Shows, TV Sci-Fi & Fantasy, Teen TV Shows



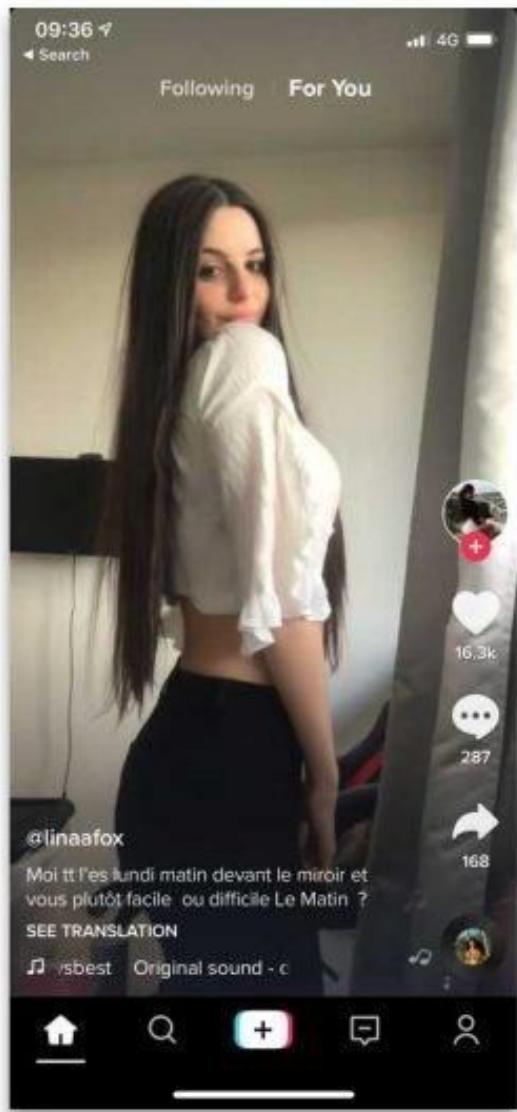
Popular on Netflix



Recently Watched



1. Tiktok



TV Schedule



FALL TV GRID 2022

New series are listed in **RED**

SUN	8:00	8:30	9:00	9:30	10:00	10:30
ABC	7 pm AFV CELEBRITY JEOPARDY		CELEBRITY WHEEL OF FORTUNE		THE ROOKIE	
CBS	7 pm 60 MIN THE EQUALIZER		EAST NEW YORK		NCIS: LOS ANGELES	
FOX	To be announced					
NBC	SUNDAY NIGHT FOOTBALL					
CW	FAMILY LAW (acquired)		CORONER (acquired)			
MON	8:00	8:30	9:00	9:30	10:00	10:30
ABC	BACHELOR IN PARADISE				THE GOOD DOCTOR	
CBS	THE NEIGHBORHOOD	BOB HEARTS ABISHOLA	NCIS		NCIS: HAWAII	
FOX	To be announced					
NBC	THE VOICE				QUANTUM LEAP	
CW	ALL AMERICAN	ALL AMERICAN: HOMECOMING				
TUE	8:00	8:30	9:00	9:30	10:00	10:30
ABC	BACHELOR IN PARADISE				THE ROOKIE: FEDS	
CBS	FBI		FBI: INTERNATIONAL		FBI: MOST WANTED	
FOX	To be announced					
NBC	THE VOICE	LA BREA			NEW AMSTERDAM	
CW	THE WINCHESTERS	PROFESSIONALS (acquired)				
WED	8:00	8:30	9:00	9:30	10:00	10:30
ABC	THE CONNERS	THE GOLDBERGS	ABBOTT ELEMENTARY	HOME ECONOMICS	BIG SKY	
CBS	SURVIVOR		THE AMAZING RACE		THE REAL LOVE BOAT	
FOX	To be announced					
NBC	CHICAGO MED	CHICAGO FIRE			CHICAGO P.D.	
CW	DC'S STARGIRL	KUNG FU				
THU	8:00	8:30	9:00	9:30	10:00	10:30
ABC	STATION 19		GREY'S ANATOMY		ALASKA	
CBS	YOUNG SHELDON	GHOSTS	SO HELP ME TODD		CSI: VEGAS	
FOX	To be announced					
NBC	LAW & ORDER	LAW & ORDER: SVU			LAW & ORDER: ORGANIZED CRIME	
CW	WALKER	WALKER INDEPENDENCE				
FRI	8:00	8:30	9:00	9:30	10:00	10:30
ABC	SHARK TANK		20/20			
CBS	S.W.A.T.		FIRE COUNTRY		BLUE BLOODS	
FOX	FRIDAY NIGHT SMACKDOWN (presumably!)					
NBC	COLLEGE BOWL (until November) LOPEZ VS LOPEZ YOUNG ROCK	DATELINE NBC				
CW	PENN & TELLER: FOOL US	WHOSE LINE IS IT ANYWAY? X2				
SAT	8:00	8:30	9:00	9:30	10:00	10:30
ABC	COLLEGE FOOTBALL					
CBS	Drama encores		Drama encores		48 HOURS	
FOX	To be announced					
NBC	Drama encores		DATELINE WEEKEND MYSTERY		SNL VINTAGE	11:30 SNL
CW	MAGIC WITH THE STARS	WORLD'S FUNNIEST ANIMALS x2				

How does Netflix know that?

Large datasets + Machine learning!



Log all user behavior (views, clicks, pauses, searches, etc.)
Recommender systems apply ML to TBs of data from all users and movies to deliver a tailored experience

2. Structured data with search results

Google X 🔍 ⚙️

All Images Books News Videos More Tools

About 13,000,000 results (0.52 seconds)

 **Alan Turing**
Mathematician ⋮

Overview Education Books Videos

     More images

[https://en.wikipedia.org › wiki › Alan_Turing](https://en.wikipedia.org/wiki/Alan_Turing) ⋮

Alan Turing - Wikipedia

Alan Mathison Turing OBE FRS (/'tjøərɪŋ/; 23 June 1912 – 7 June 1954) was an English mathematician, computer scientist, logician, cryptanalyst, philosopher, ...

Partner(s): [Joan Clarke](#); (engaged in 194... Known for: [Cryptanalysis of the Enigm...](#)

Awards: Smith's Prize (1936) Resting place: Ashes scattered in gard...

[The Enigma](#) · [Alan Turing law](#) · [Legacy of Alan Turing](#) · [Alan Turing Year](#)

About

Alan Mathison Turing OBE FRS was an English mathematician, computer scientist, logician, cryptanalyst, philosopher, and theoretical biologist.

[Wikipedia](#)

Born: June 23, 1912, [Maida Vale, London, United Kingdom](#)

Died: June 7, 1954, [Wilmslow, United Kingdom](#)

Academic advisor: [Alonzo Church](#)

Education: Princeton University (1936–1938), [MORE](#)

Influenced by: [Alonzo Church](#), [Kurt Gödel](#), [Ludwig Wittgenstein](#), [Max Newman](#)

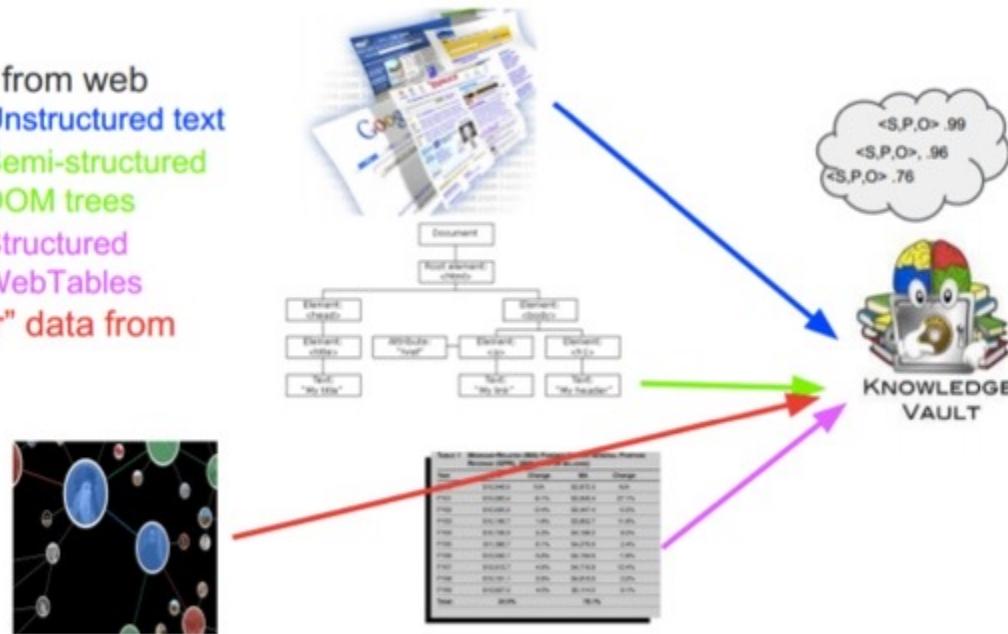
Notable students: [Robin Gandy](#), [Beatrice Worsley](#)

How does Google know that?

Large datasets + Machine learning!

Knowledge Vault* fuses all these signals together

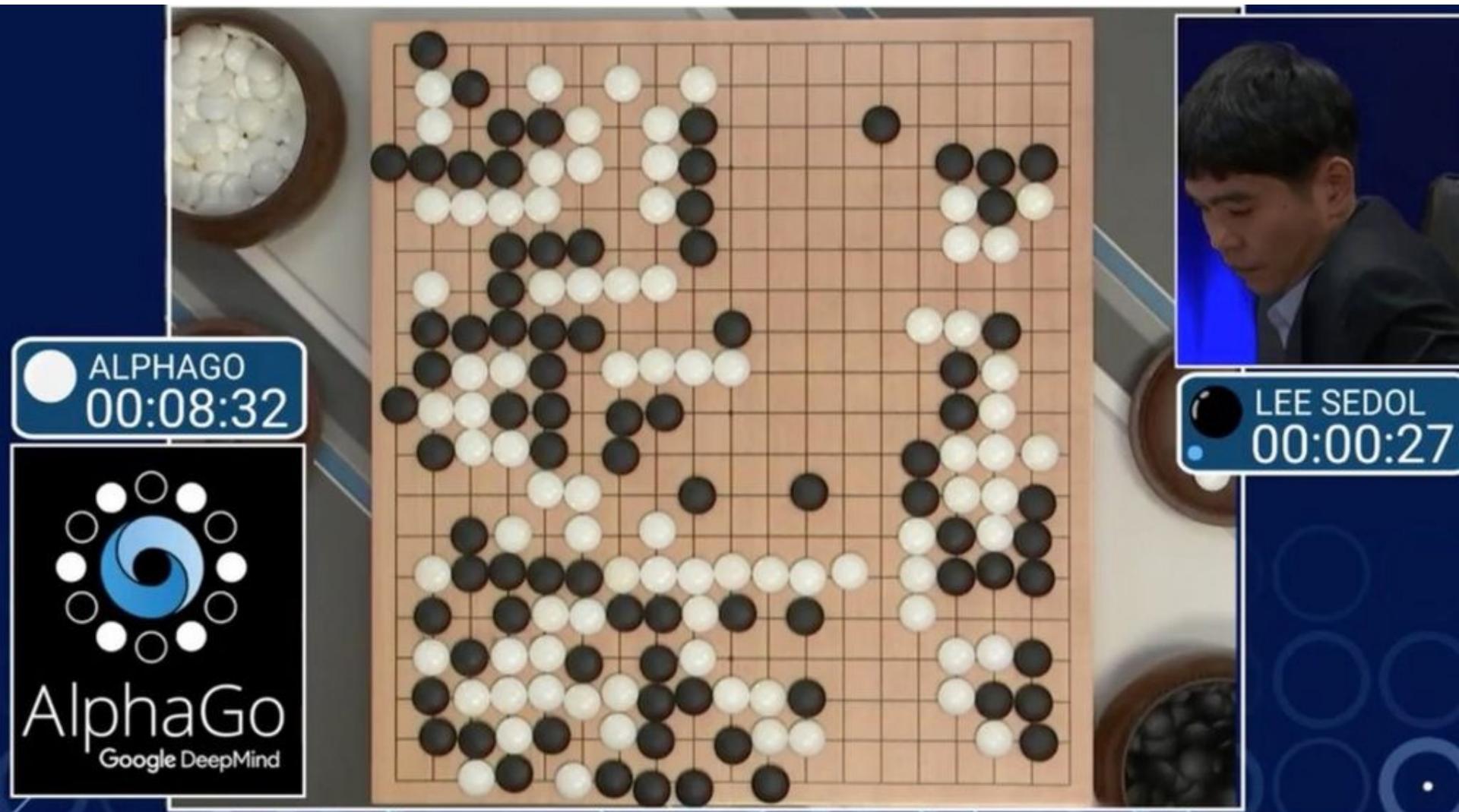
- Data from web
 - Unstructured text
 - Semi-structured DOM trees
 - Structured WebTables
 - "Prior" data from FB



* Details in a paper submitted to WWW'14 (Dong et al)

Knowledge Base Construction (KBC) process extracts tabular/relational data from large amounts of text data

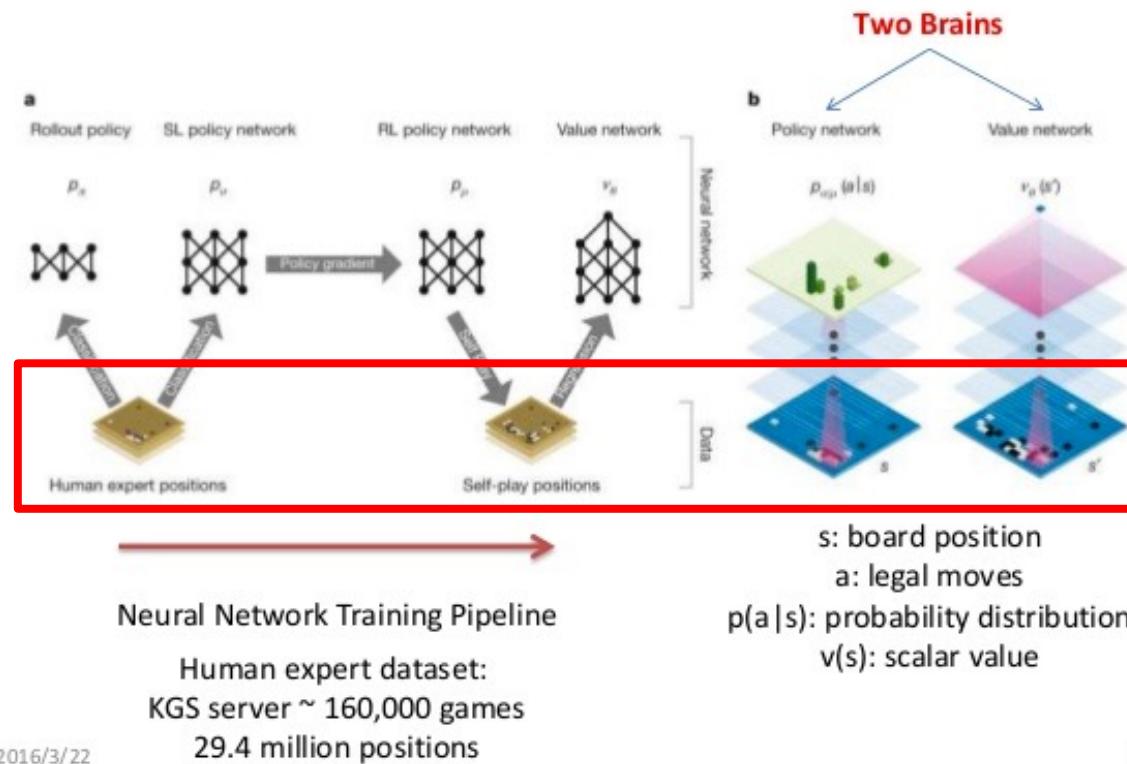
3. AlphaGo defeats human champion!



How did AlphaGo achieve that?

Breakthrough powered by deep learning!

Architecture of AlphaGo



Deep CNNs to visually process board status in plays

Innumerable “enterprise” applications

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

se



307 comments, 167 called-out

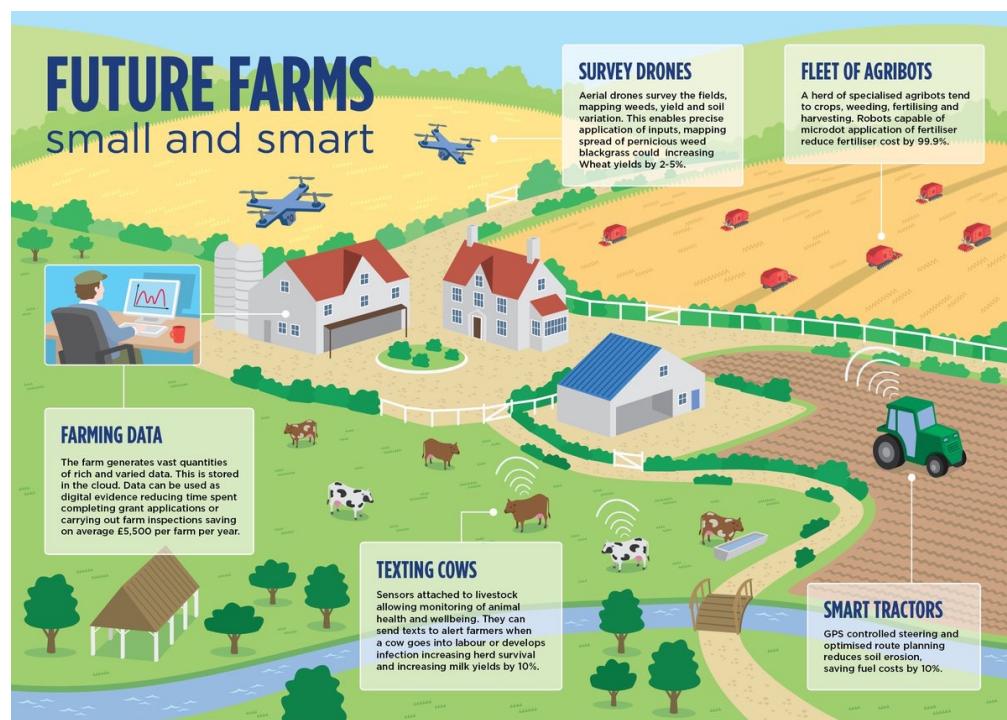
+ Comment Now + Follow Comments

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

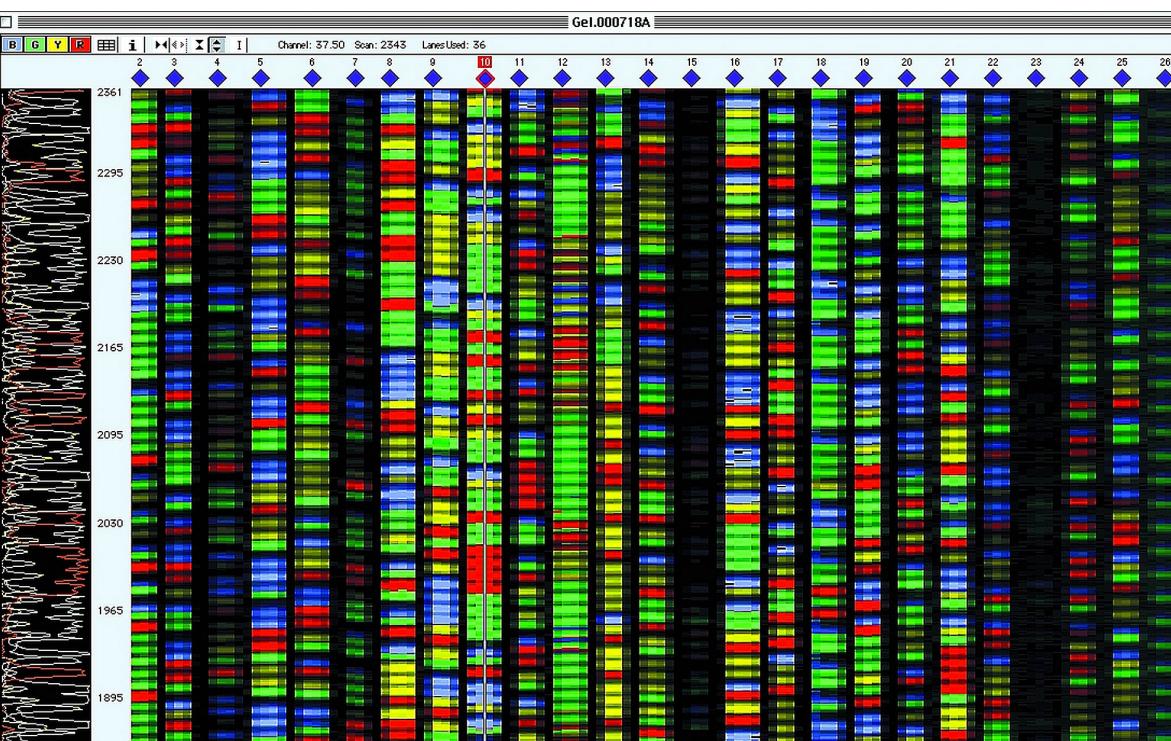
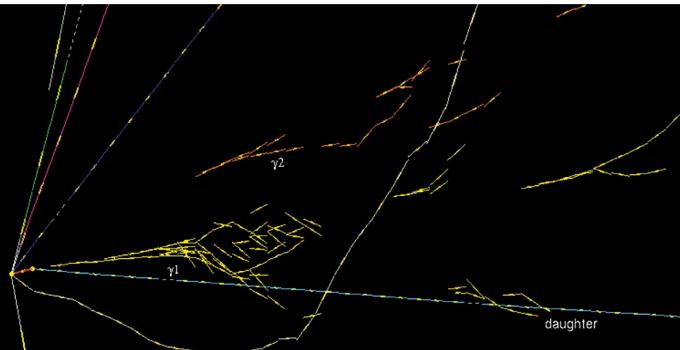
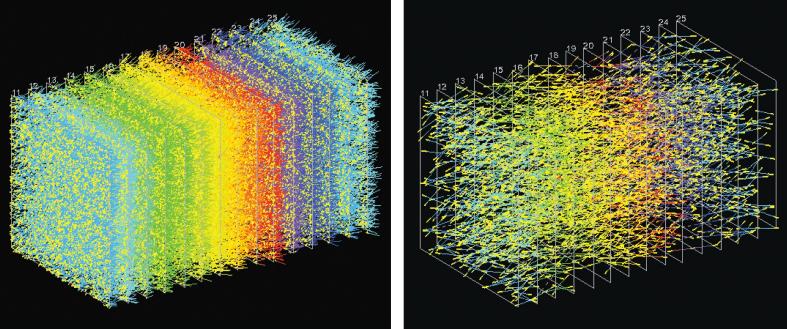
Charles Duhigg outlines in the New York Times how Target tries to hook parents-to-be at that crucial moment before they turn into



Target has got you in its aim

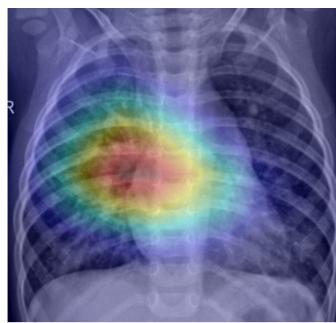


“Domain sciences” and healthcare tech
are also becoming data+ML intensive

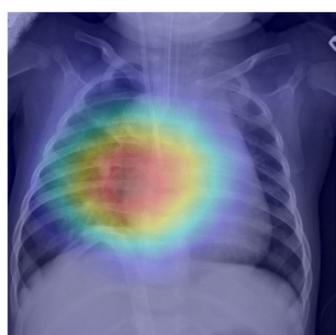


This is Data Release 16.

Data Surveys Instruments



(a)



(b)

Software systems for data analytics and ML over large and complex datasets are now critical for digital applications in many domains

The Age of “Big Data”/“Data Science”

The New York Times

SundayReview | NEWS ANALYSIS

The Age **Forbes** / Entrepreneurs

By STEVE LOHR F

MAR 25, 2015 @ 7:33 PM 4,407 VIEWS

Email

Share

Tweet

Save

Drowning In Big Data - Finding Insight In A Digital DATA Josh Steimle, CON **Data Scientist: The Sexiest Job of the 21st Century**

by Thomas H. Davenport and D.J. Patil

FROM THE OCTOBER 2012 ISSUE

[SUMMARY](#) [SAVE](#) [SHARE](#) [COMMENT](#) [TEXT SIZE](#) [PRINT](#) [\\$8.95](#)



**Harvard
Business
Review**

For roughly a decade, information about Big Data. The IDC industry will experience by 2018. What this

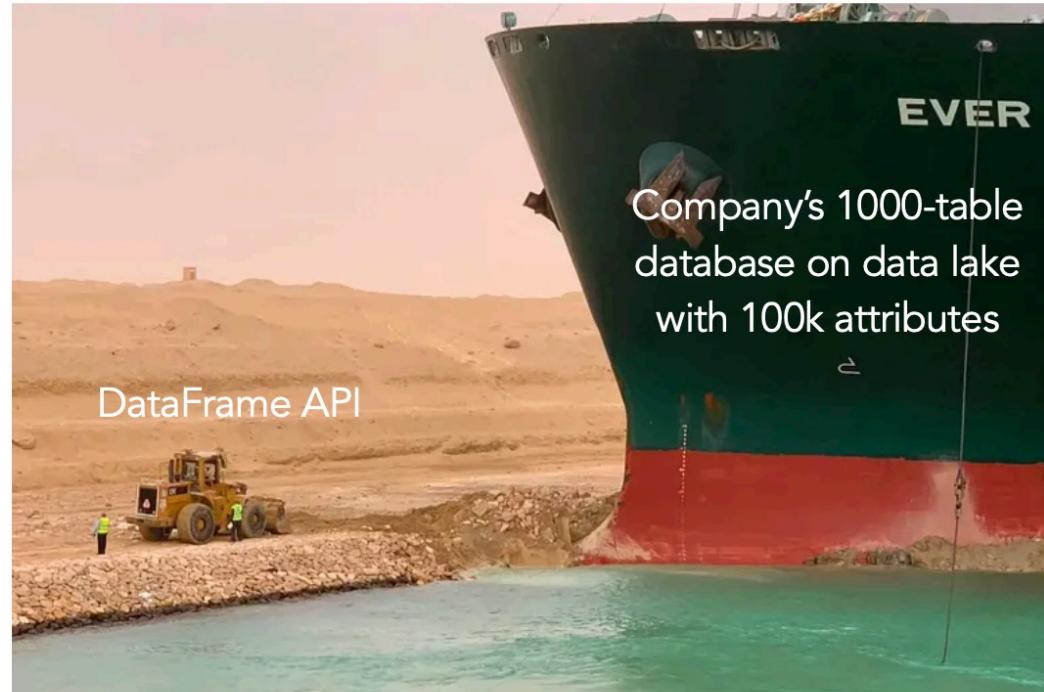
When Jonathan Goldman arrived for work in June 2006 at LinkedIn, the business networking site, the place still felt like a start-up. The company had just under 8 million accounts, and the number was growing quickly as existing members invited their friends and colleagues to join. But users weren't seeking out connections with the people who were already on the site at the rate executives had expected. Something was apparently missing in the social experience. As one LinkedIn manager put it, "It was like arriving at a conference reception and realizing you don't know anyone. So you just stand in the corner sipping your drink—

*Data data everywhere,
All the wallets did shrink!*

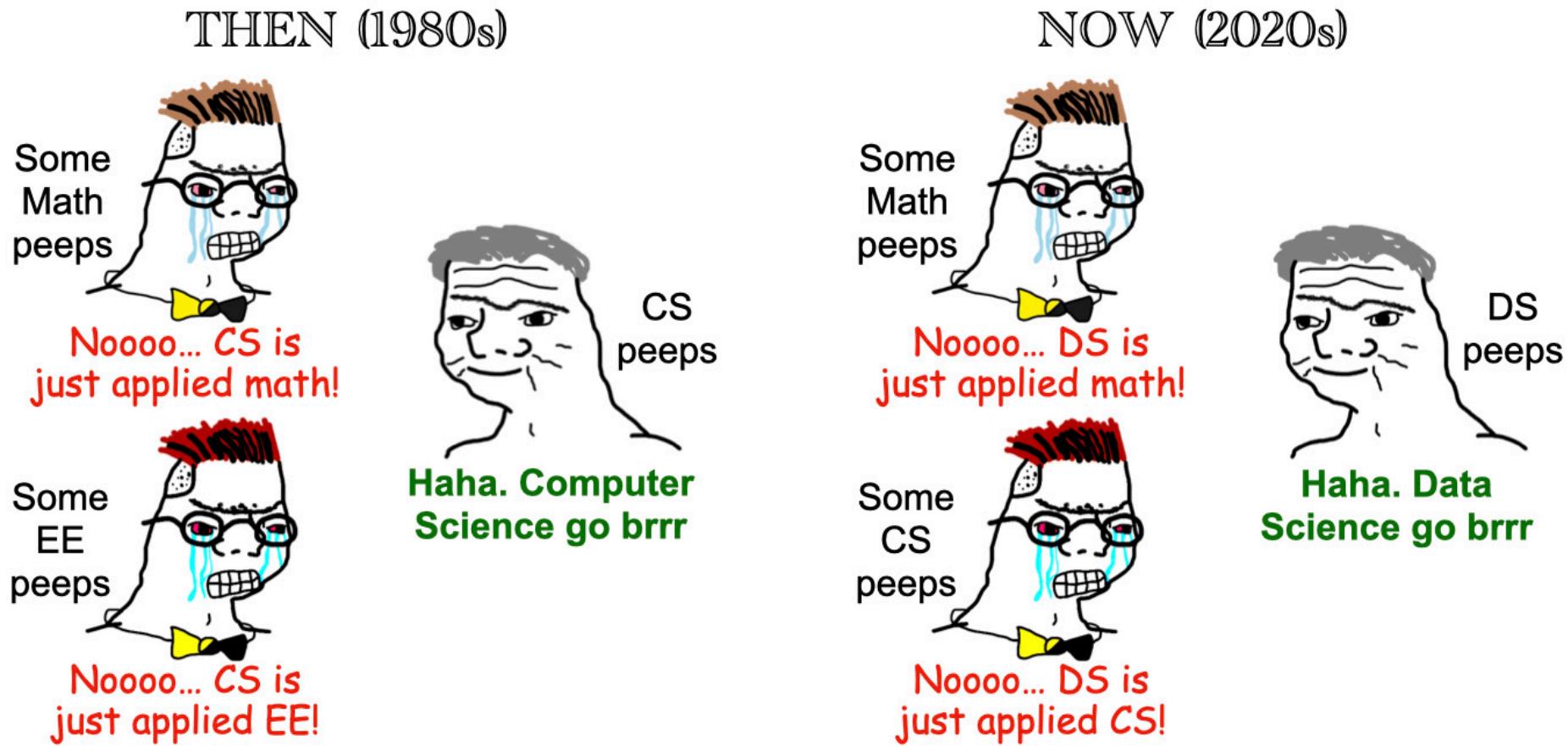
*Data data everywhere,
Nor any moment to think?*

DSC 204a Scalable Data Systems

- Haojian Jin



Meme from Previous DSC 102



Vision

Data science professionals ought to be familiarized with data systems from a user's standpoint, as opposed to the conventional approach of a system implementer.

15-213/15-513/14-513 Introduction to Computer Systems (ICS)

Fall 2023

- 15-213 Pittsburgh: Tue, Thu 12:30 PM–01:50 PM, GHC 4401, [Brian Railing](#) and [Phillip Gibbons](#)
- 14-513 Pittsburgh: Tue, Thu 12:30 PM–01:50 PM, CIC 1202, [David Varodayan](#)

12 units

The ICS course provides a programmer's view of how computer systems execute programs, store information, and communicate. It enables students to become more effective programmers, especially in dealing with issues of performance, portability and robustness. It also serves as a foundation for courses on compilers, networks, operating systems, and computer architecture, where a deeper understanding of systems-level issues is required. Topics covered include: machine-level code and its generation by optimizing compilers, performance evaluation and optimization, computer arithmetic, memory organization and management, networking technology and protocols, and supporting concurrent computation.

[Course Syllabus](#)

Prerequisites: 15-122

DSC 102 will get you thinking about the **fundamentals of systems for scalable analytics**

1. “**Systems**”: What resources does a computer have? How to store and efficiently compute over large data? What is cloud?
2. “**Scalability**”: How to scale and parallelize data-intensive computations?
3. **For “Analytics”:**
 1. **Source**: Data acquisition & preparation for ML
 2. **Build**: Model selection & deep learning systems
 3. **Deploying** ML models
4. Hands-on experience with scalable analytics tools

The Lifecycle of ML-based Analytics

