

Data Management Group @Utah

**Data Management Research Center for
Human-Centered, Efficient, and Scalable Systems**

<http://mod.cs.utah.edu/>

Data is the new oil!



Courtesy: <https://www.economist.com/> David Parkins

But oil has to be refined and extracted to be usable

Our job is to develop refinement machinery to extract *information* from data!

Modern data challenges: Data volume

Mick W@tson
@BioMickWatson

Bioinformatics over the years:

- 1990s: doing a BLAST search
- 2000s: analysing 30 microarrays
- 2010s: analysing 6Tb of NGS
- 2020s: creating a cloud the size of Netflix to reanalyse the whole of SRA for one figure

12:57 AM · 2/12/21 · Twitter Web App

117 Retweets 9 Quote Tweets 697 Likes

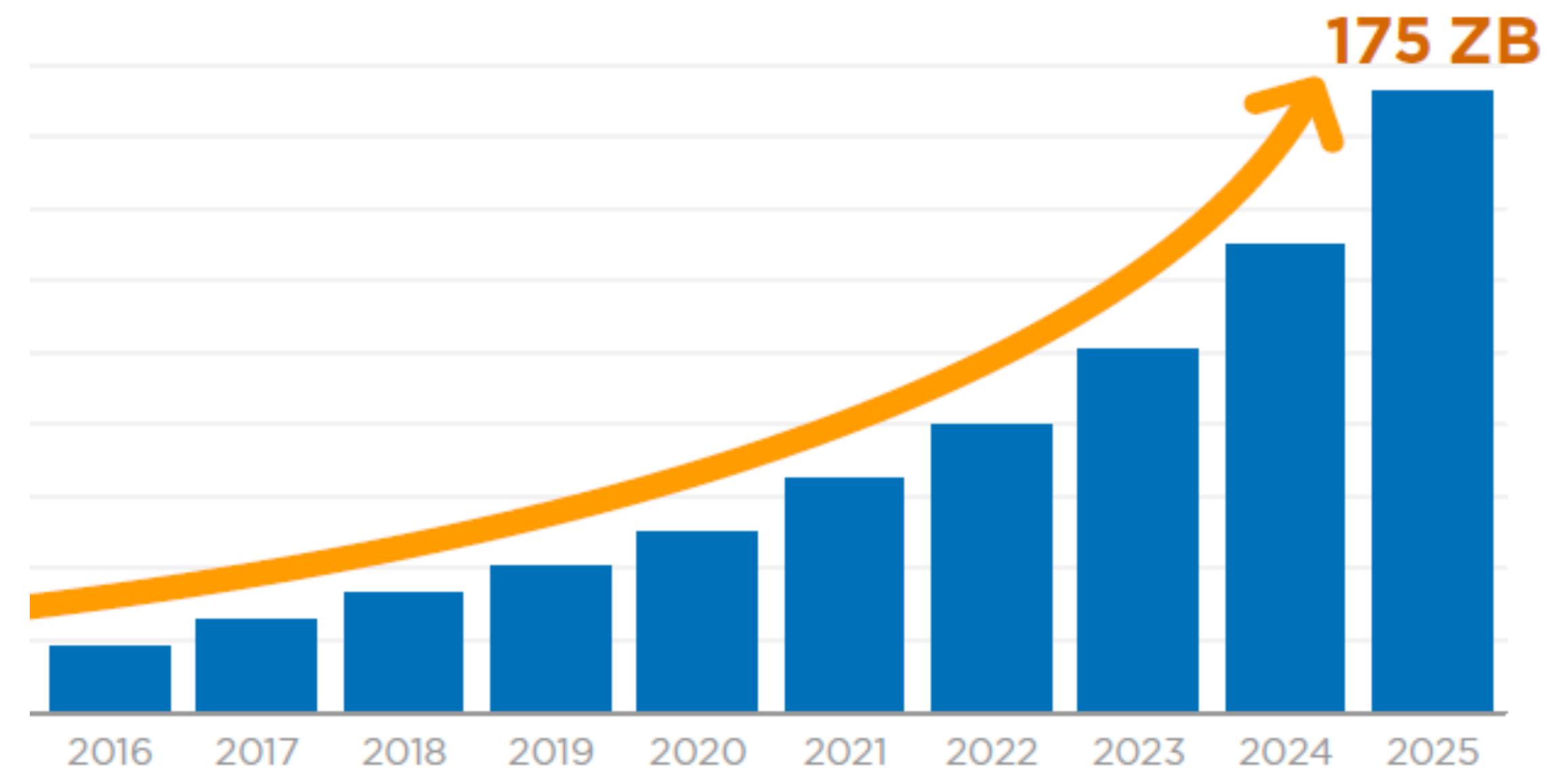
Michael Schatz @mike_schatz · 2h
Replying to @BioMickWatson

This is basically my life right now

1 8



IDC says 175 ZB will be created by 2025 (Image courtesy IDC)



How many 0s in a Zettabyte?

Modern data challenges: Data debugging

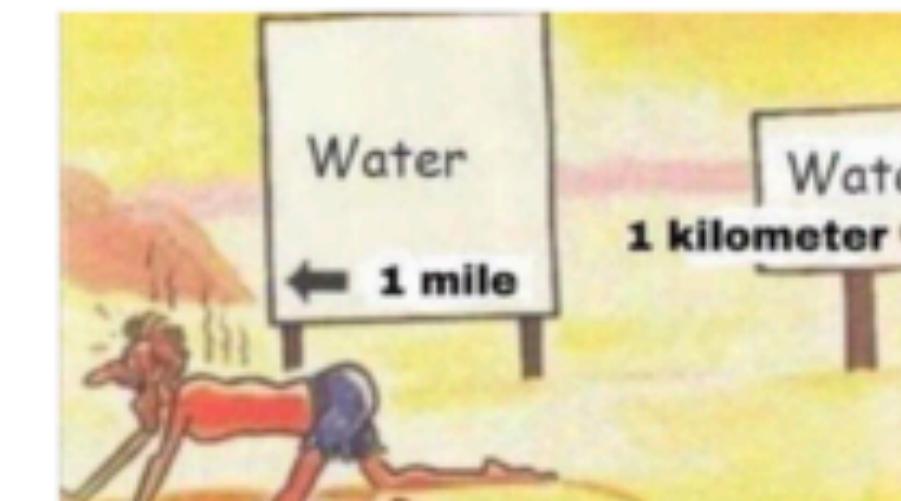
Mismatch between data and systems



90% Accuracy

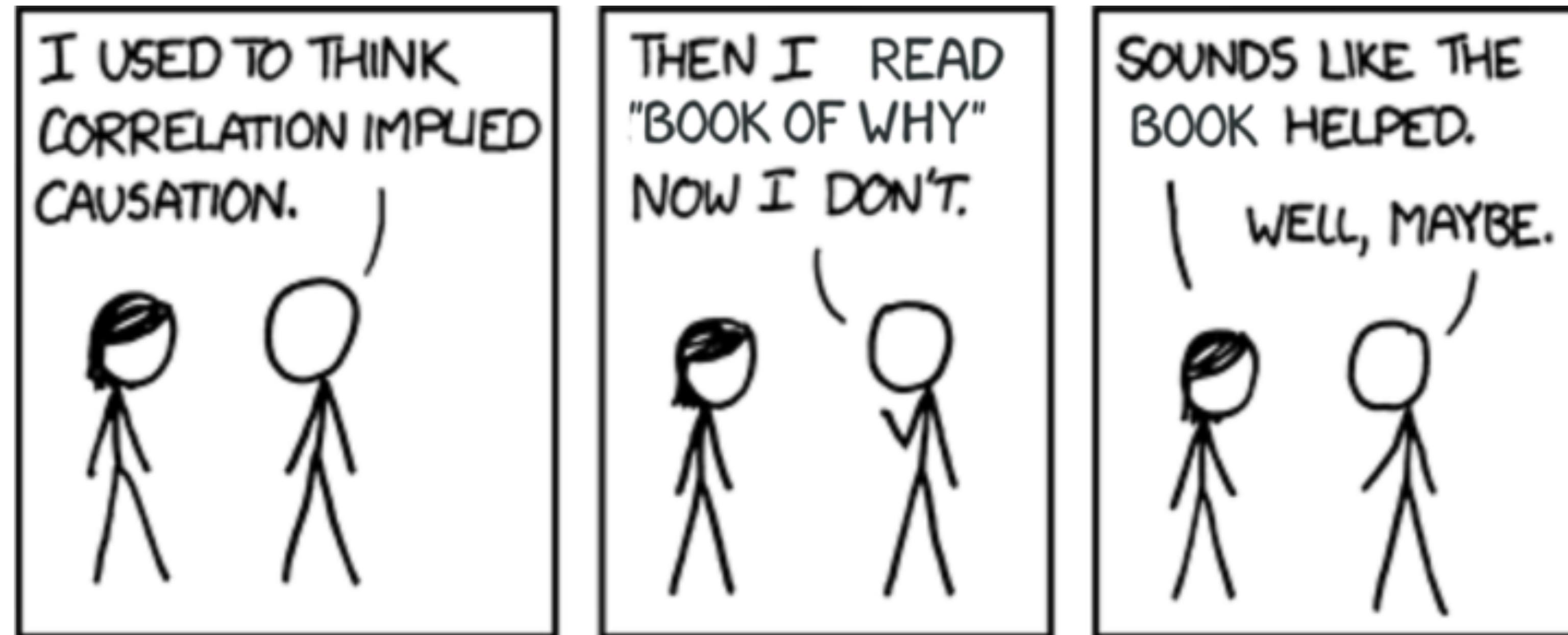


30% Accuracy



Metric vs Imperial
KG vs lbs

Modern data challenges: Data explainability



Faculty members



Prof. Jeff Phillips
Research Interests:
Streaming & Sketching
Spatial data management
Spatial statistics and trajectories
High-dimensional data



Prof. Prashant Pandey
Research Interests:
Data Structures for Big Data
Graph processing
Computational biology
Deep learning on graphs



Prof. El Kindi Rezig
Research Interests:
Data exploration and debugging
Data quality and cleaning
Data visualization
Query languages and UI



Prof. Anna Fariha
Research Interests:
Data system usability
Data summarization
Explainable AI
Data fairness

Current projects

Democratizing data-driven systems

— Anna Fariha

Data systems on modern hardware

— Prashant Pandey

Spatial exposome data: CEDaR
— Jeff Phillips

Efficient & scalable graph processing

— Prashant Pandey

Data debugging, exploration, & information extraction
— El Kindi Rezig

Data structures for Scalable Computing

— Jeff Phillips/Prashant Pandey

Large-Scale indexing & analysis of genomic data
— Prashant Pandey

Persistent data summaries

— Jeff Phillips/Prashant Pandey

Data sketching
— Jeff Phillips

Students



Hasan Pourmahmood



Alex Tokita



Manoj Marneni



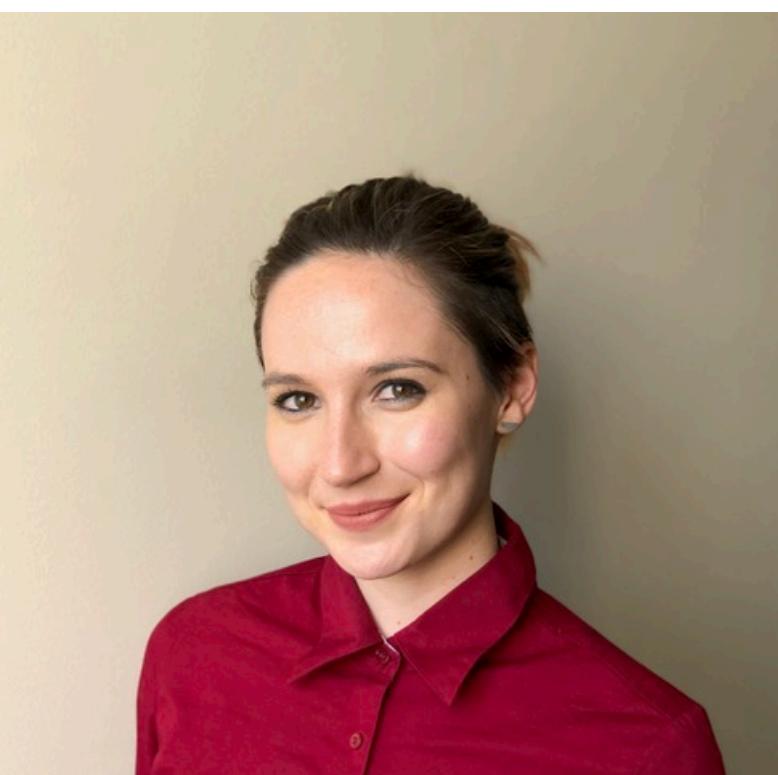
Hunter McCoy



Medha Kalkur



Benwei Shi



Anna Bell



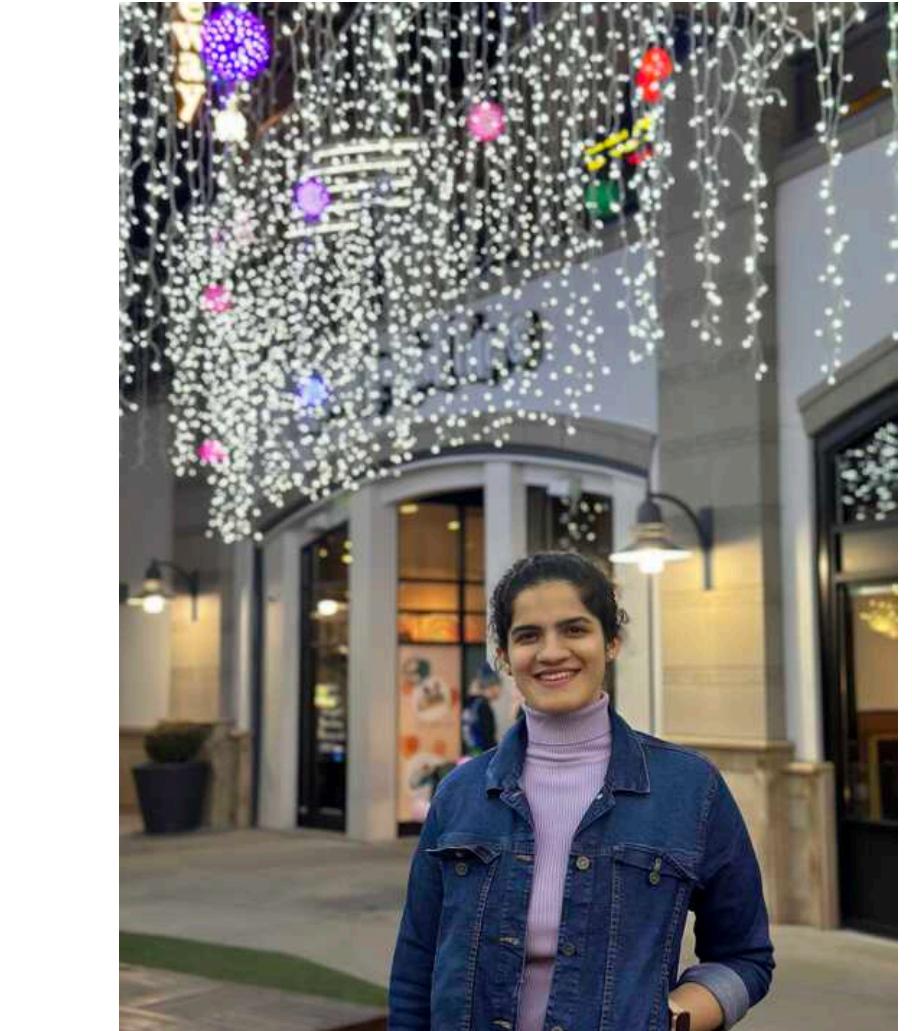
Susmitha Raja



Pranjali Patil



Tao Yang



And more....

Our alumni:

- Dong Xie – Professor at Penn State
- Zhouyue Zhao – Professor at University of Buffalo
- Sunipa Dev – Research Scientist at Google Research
- Yan Zheng – Research Scientist at Visa Research
- Debjyoti Paul – Research Scientist at Meta
- Yanqing Peng – Research Scientist at Meta
- Mina Ghashmi – Applied Scientist at Amazon
- You – Whatever you want to be...