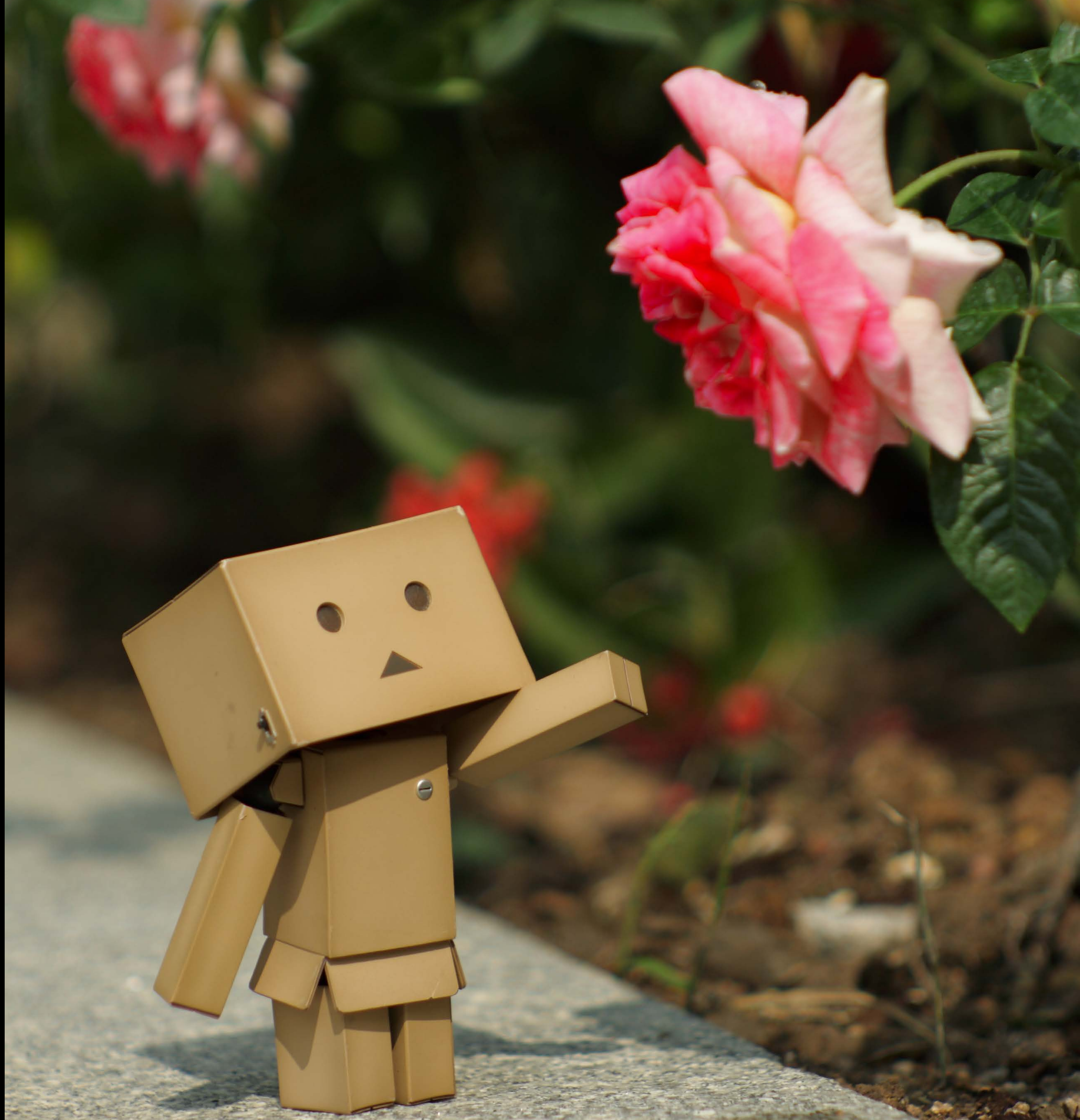
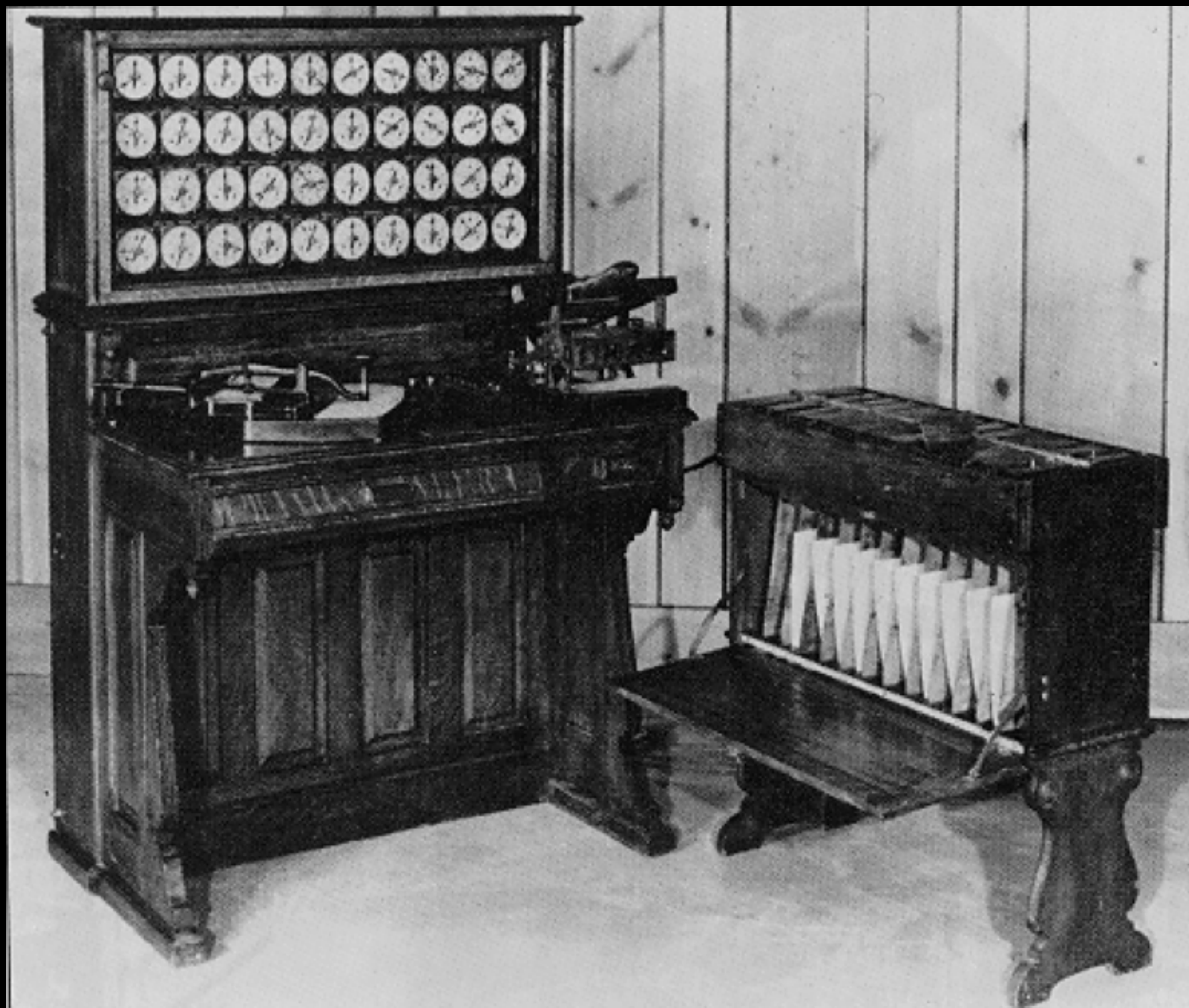


Artificial Intelligence in 2017

Hilary Mason
Fast Forward Labs

@hmason
hilary@fastforwardlabs.com





AI?

Machine
Learning

Data Science

Analytics

“Big Data”

A few key terms:

- Artificial Intelligence
- Machine Learning
- Deep Learning
- Neural Networks
- Cognitive
- Data Science
- Big Data

I BELIEVE OUR ADVENTURE
THROUGH TIME HAS TAKEN



A MOST
SERIOUS TURN

AI products are products that would not be possible without the use of data.

(it's all the same.)

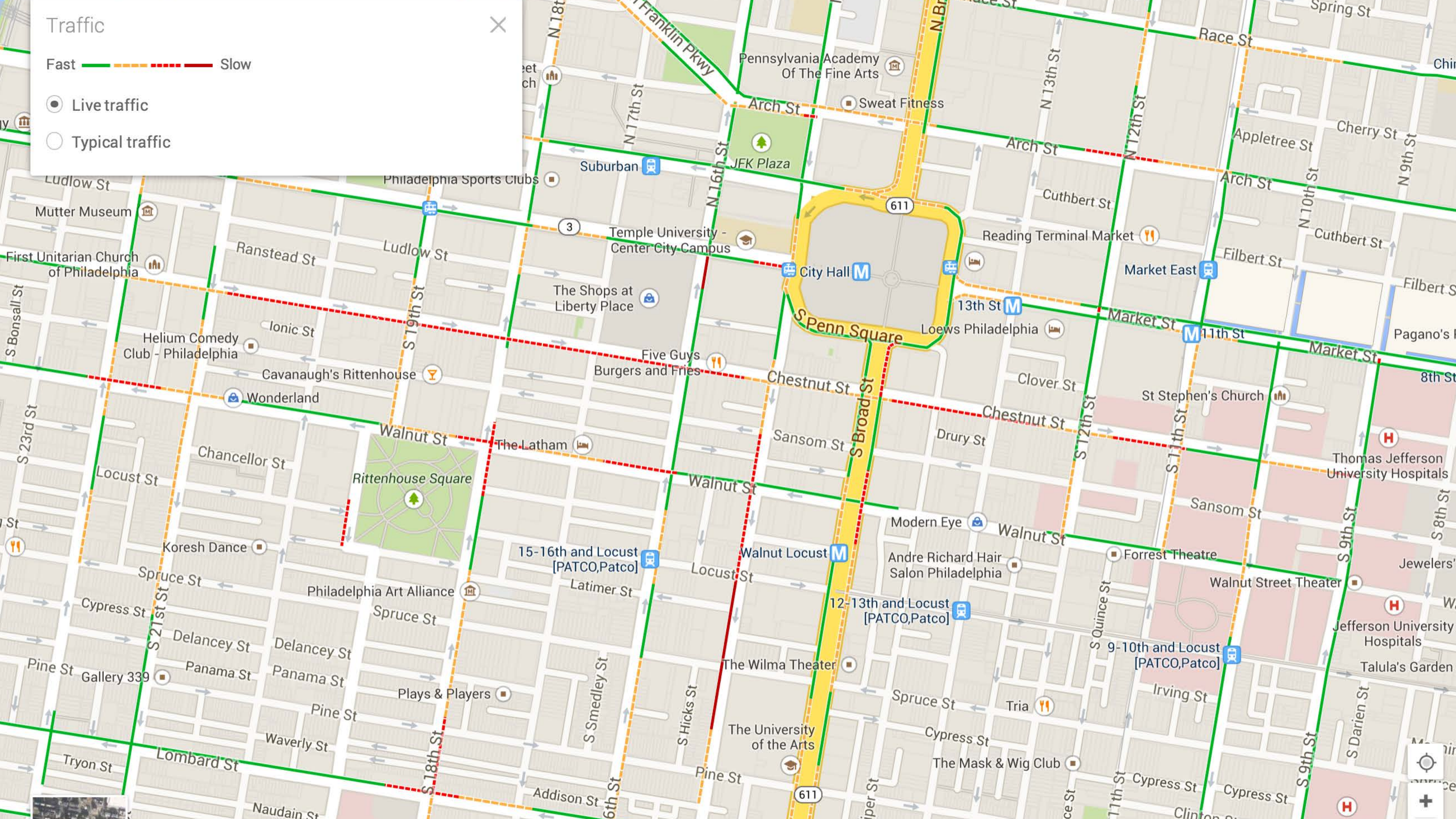
”AI is whatever computers
can’t do today.”

Traffic

Fast    Slow

 Live traffic

 Typical traffic



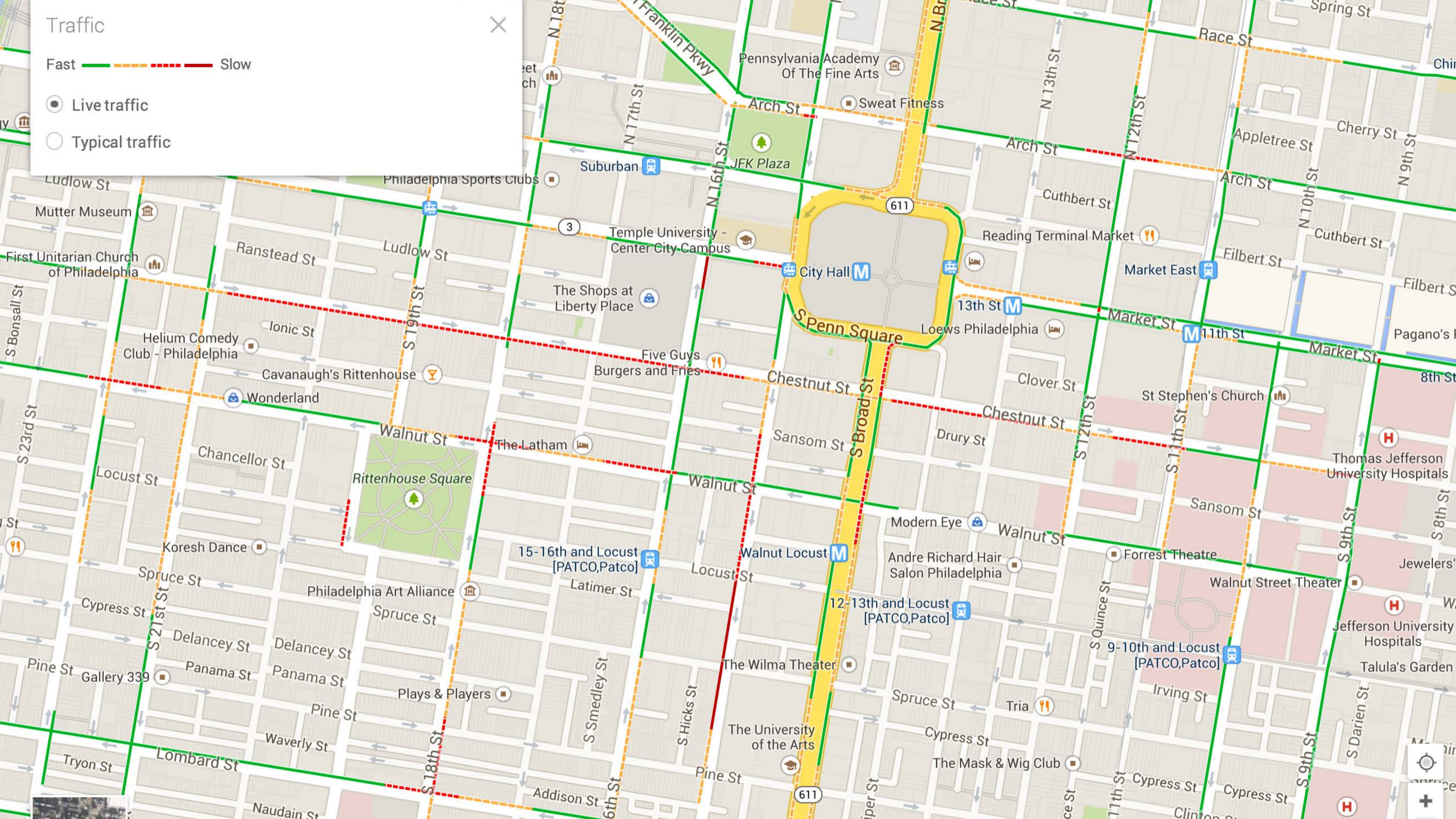
(it's boring)

Traffic

Fast    Slow

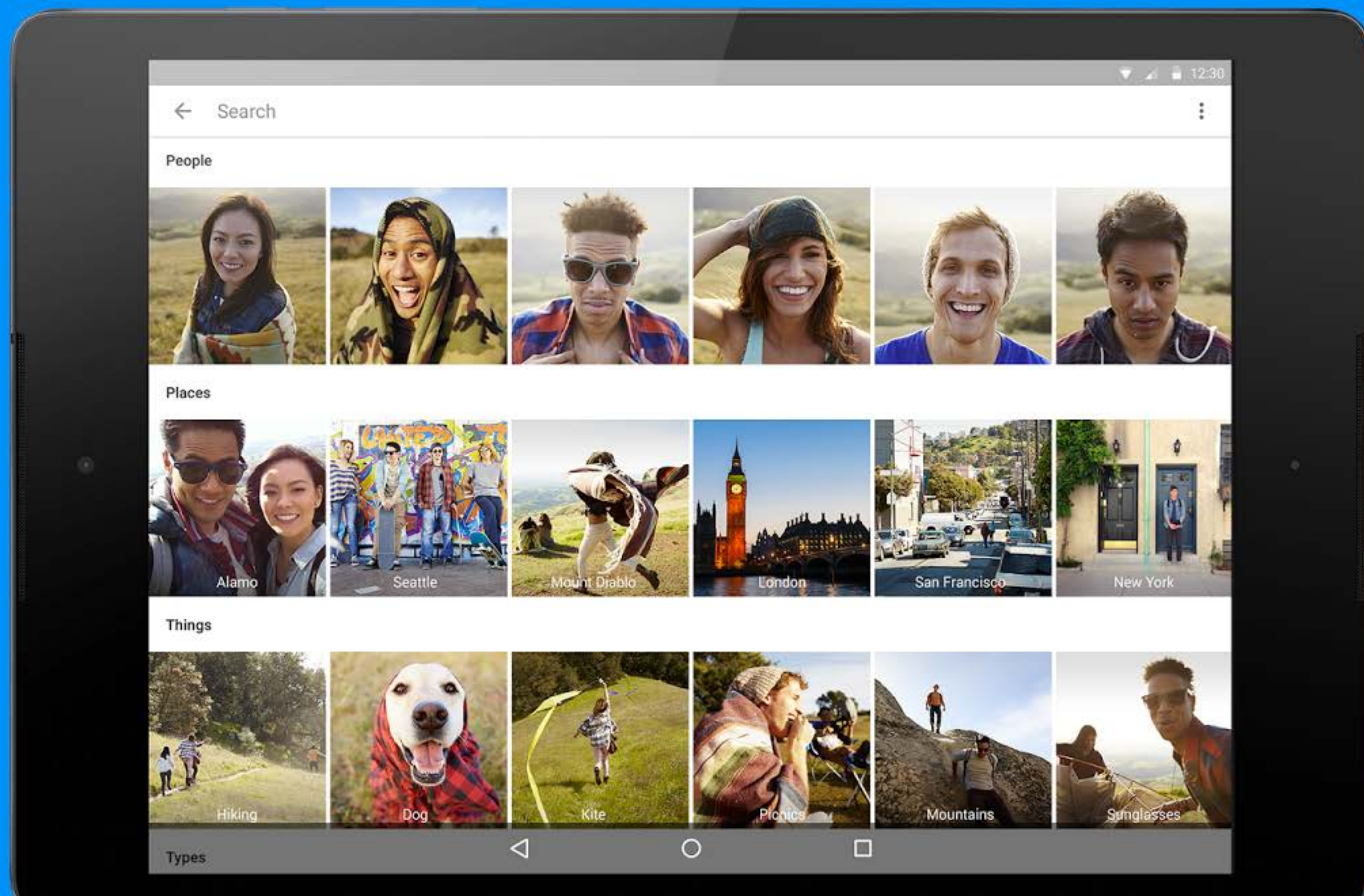
 Live traffic

 Typical traffic



What makes an **intelligent** product?

Automatically organized by what matters



Customers Who Bought This Item Also Bought



K9Cakery Silicone Cake Pan, 7-Inch by 10-Inch, Small
★★★★★ 61
\$14.99 Prime



Puppy Cake Pumpkin Cake Mix and Frosting (Wheat-free)
★★★★★ 59
\$6.99 Prime



Multipet Plush 5.5-Inch Musical Birthday Cake Dog Toy
★★★★★ 332
\$5.99

Maple Bacon Flavor



Puppy Scoops Ice Cream Mix: Maple Bacon
★★★★★ 45
\$6.99 Prime



Puppy Cake Carob Cake Mix and Frosting (9 oz)
★★★★★ 78
\$6.99 Prime



Puppy Cake Red Velvet Cake Mix 9oz
★★★★★ 50
\$6.99



Puppy Party - Cone Hats, (8 ct)
★★★★★ 7
\$6.04



(this is not a solved problem)

AT&T 4G 8:59 96%

RADAR NOTIFY

HEAVY
MED
LIGHT

10 min 20 min 30 min 40 min 50 min

NOW → RAIN, 56°

NEXT HOUR → RAIN

FULL DAY

⊙ Carrabelle, FL

AT&T 82%

14:33
Wednesday, September 26

Dark Sky
Light rain starting soon! ~ New York, NY

→ slide to view

AT&T 4G 9:00 96%

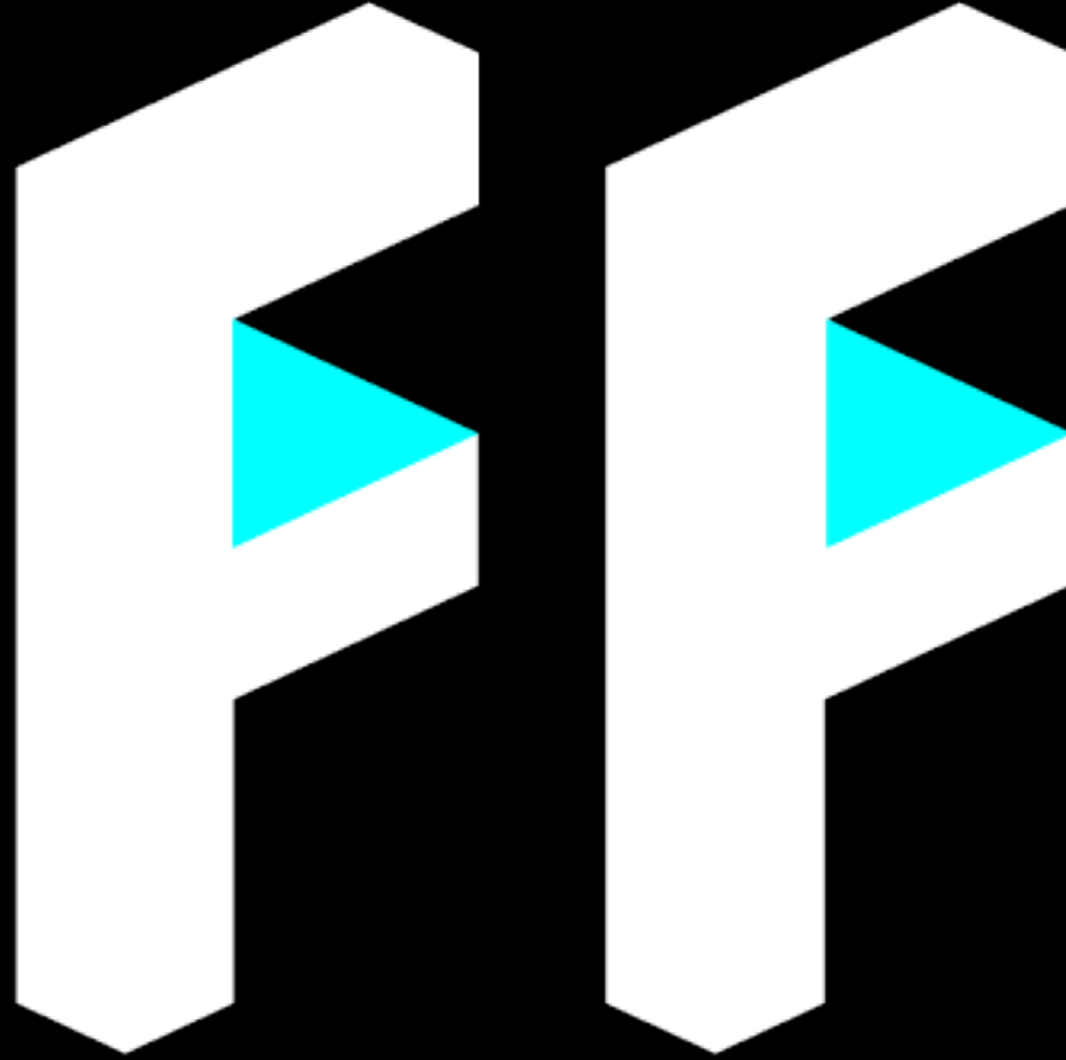
USA

Malone, Marianna, Sneads, Quincy, Tallahassee, Monticello, Madi, Alpha, Blountstown, Sopchoppy, Wewahitchka, Mexico Beach, Apalachicola, Perry

8:59

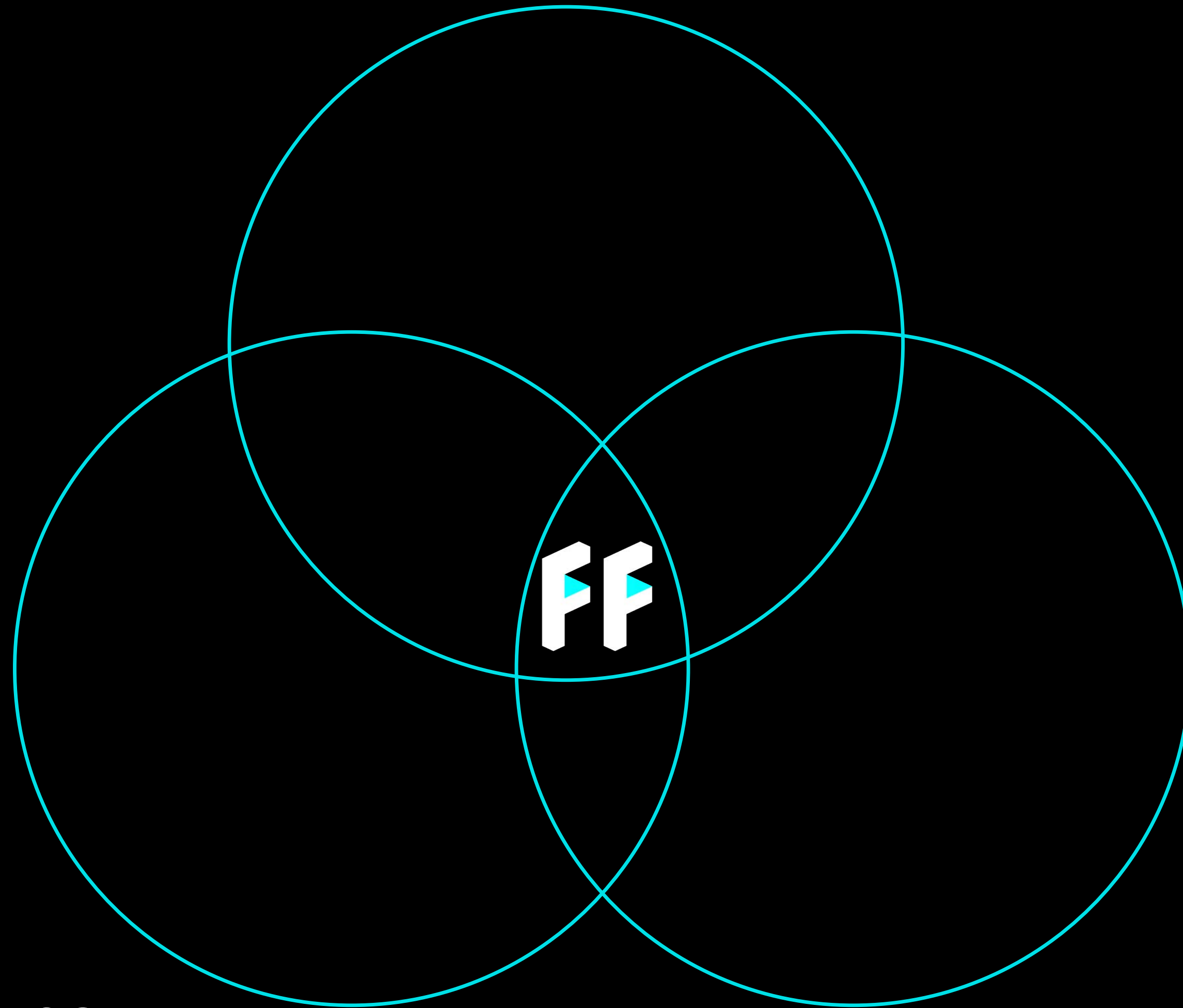
min -20 min -10 min 10 min 20 min 30 min

BACK



FAST FORWARD
LABS

Enterprise



Startups

Academic Research

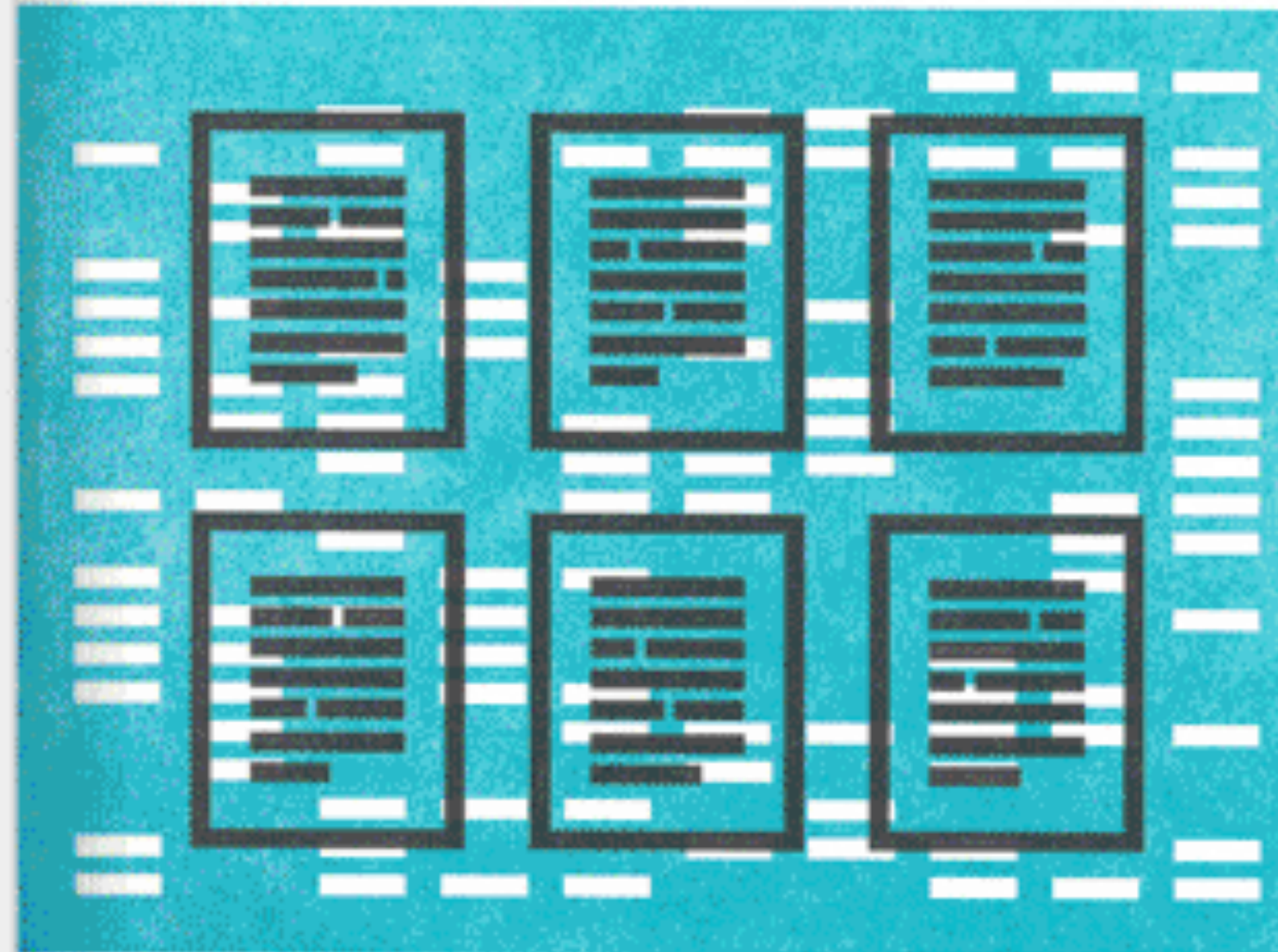


STARK

Fast Forward Labs

FF

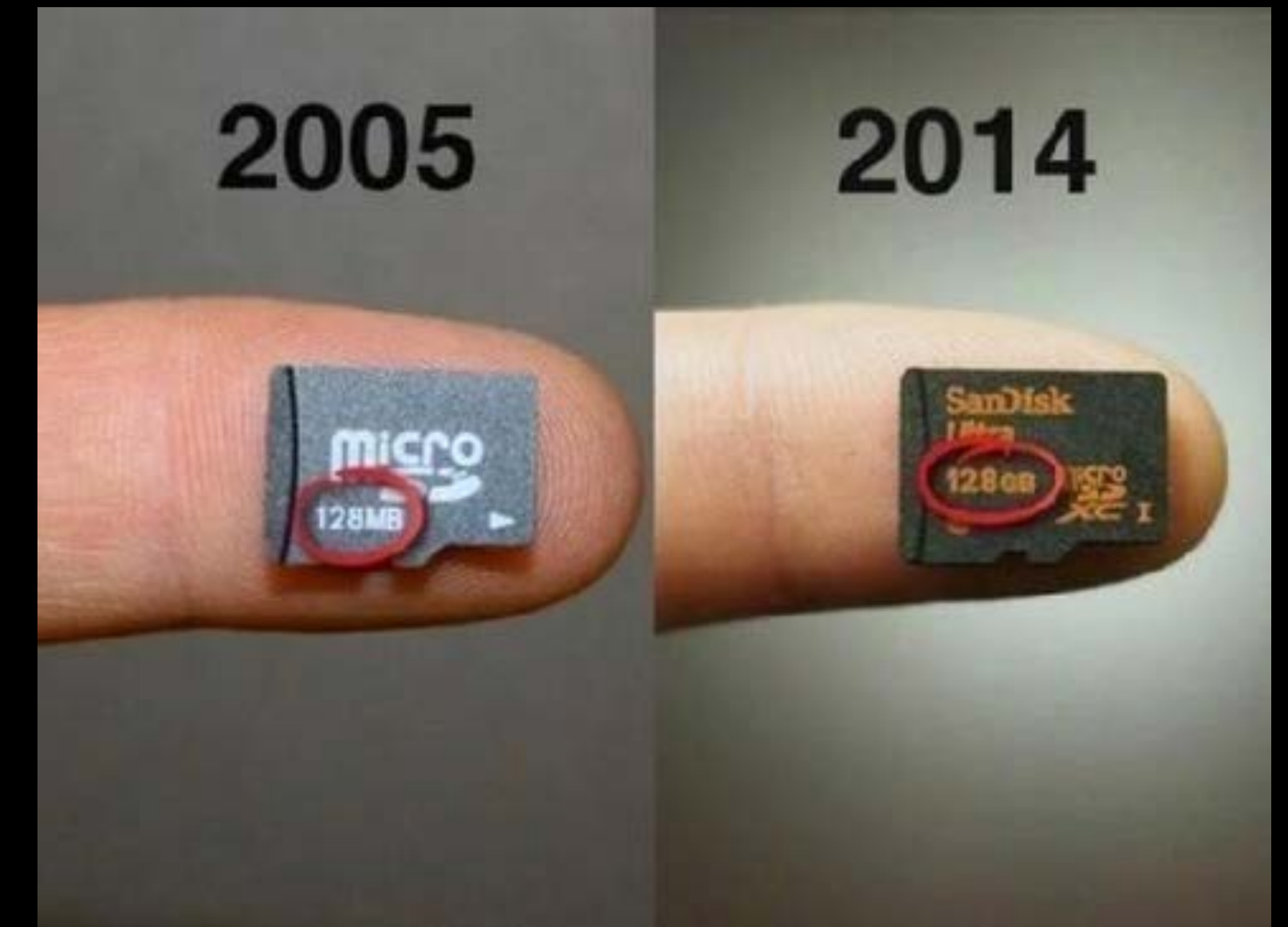
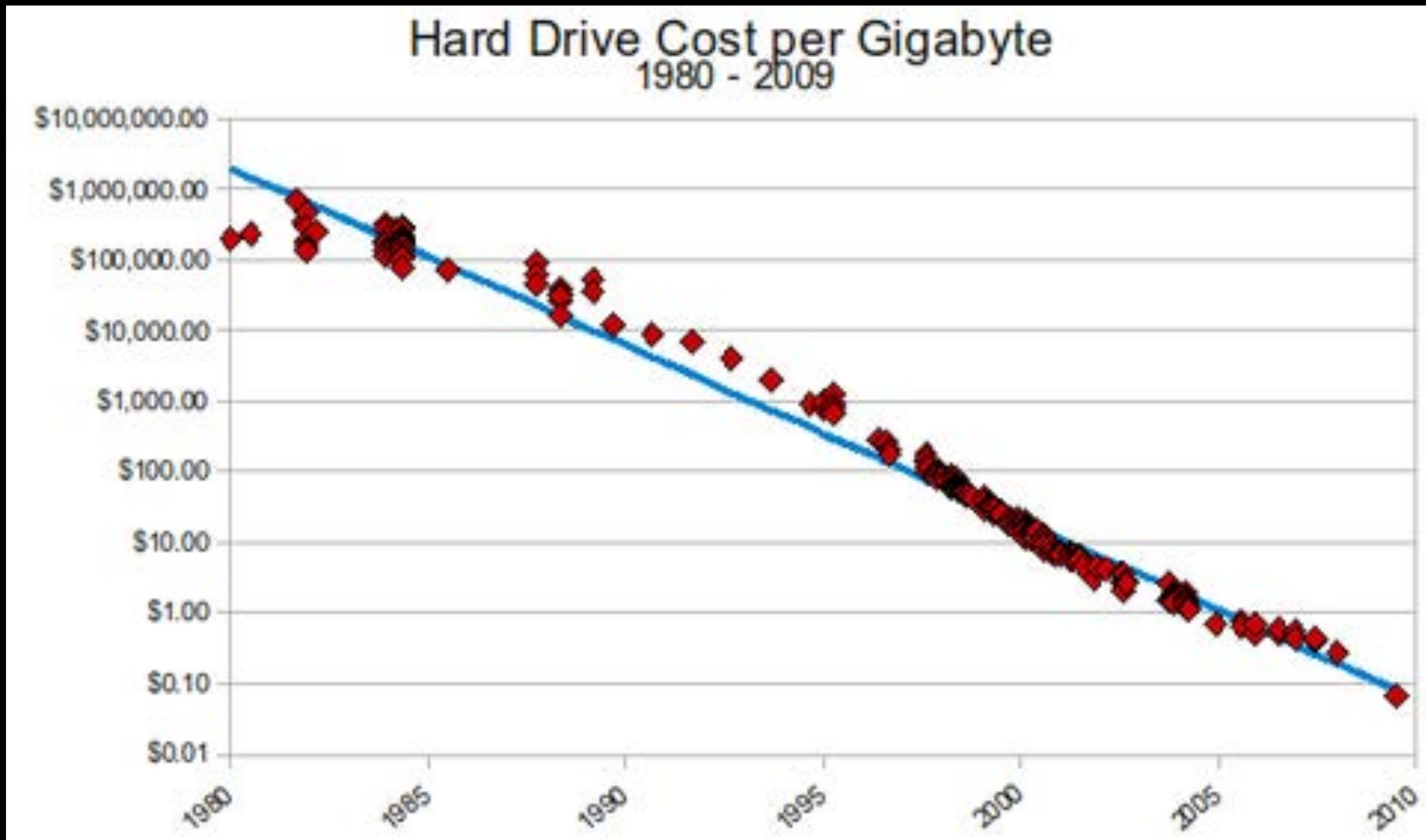
Natural Language Generation



1) a research
breakthrough





2) a change in economics



3) a capability becomes a commodity

Apache > Hadoop >



Search with Apache Solr Search

Last Published: 07/07/2015 17:22:04

Top Wiki


▼ About

- Welcome
- What Is Apache Hado...
- Getting Started ...
- Download Hadoop
- Who Uses Hadoop?...
- News

- Releases
- Mailing Lists
- Issue Tracking
- Who We Are?
- Who Uses Hadoop?
- Buy Stuff
- Sponsorship
- Thanks
- Privacy Policy
- Bylaws
- License

► Documentation

► Related Projects



Welcome to Apache™ Hadoop®!

[PDF](#)

What Is Apache Hadoop?

The Apache™ Hadoop® project develops open-source software for reliable, scalable, distributed computing.

The Apache Hadoop software library is a framework that allows for the distributed processing of large data sets across clusters of computers using simple programming models. It is designed to scale up from single servers to thousands of machines, each offering local computation and storage. Rather than rely on hardware to deliver high-availability, the library itself is designed to detect and handle failures at the application layer, so delivering a highly-available service on top of a cluster of computers, each of which may be prone to failures.

The project includes these modules:

- **Hadoop Common:** The common utilities that support the other Hadoop modules.
- **Hadoop Distributed File System (HDFS™):** A distributed file system that provides high-throughput access to application data.
- **Hadoop YARN:** A framework for job scheduling and cluster resource management.
- **Hadoop MapReduce:** A YARN-based system for parallel processing of large data sets.

Other Hadoop-related projects at Apache include:

- **Ambari™:** A web-based tool for provisioning, managing, and monitoring Apache Hadoop clusters which includes support for Hadoop HDFS, Hadoop MapReduce, Hive, HCatalog, HBase, ZooKeeper, Oozie, Pig and Sqoop. Ambari also provides a dashboard for viewing cluster health such as heatmaps and ability to view MapReduce, Pig and Hive applications visually alongwith features to diagnose their performance characteristics in a user-friendly manner.
- **Avro™:** A data serialization system.
- **Cassandra™:** A scalable multi-master database with no single points of failure.
- **Chukwa™:** A data collection system for managing large distributed systems.
- **HBase™:** A scalable, distributed database that supports structured data storage for large tables.



WIKIPEDIA
The Free Encyclopedia

- Main page
- Contents
- Featured content
- Current events
- Random article
- Donate to Wikipedia
- Wikipedia store

Interaction

- Help
- About Wikipedia
- Community portal
- Recent changes
- Contact page

Tools

- What links here
- Related changes
- Upload file
- Special pages
- Permanent link
- Page information
- Wikidata item
- Cite this page

Print/export

- Create a book
- Download as PDF
- Printable version

Languages

- Deutsch
- فارسی
- Français
- हिन्दी
- Latviešu
- 日本語
- Polski
- Русский
- Simple English
- 中文

Article [Talk](#)

[Read](#) [Edit](#) [View history](#)

Data science

From Wikipedia, the free encyclopedia

 This article's **tone or style may not reflect the encyclopedic tone used on Wikipedia**. See Wikipedia's [guide to writing better articles](#) for suggestions. *(February 2014)*

Data Science is the extraction of [knowledge](#) from large volumes of [data](#) that are structured or unstructured,^{[1][2]} which is a continuation of the field [data mining](#) and [predictive analytics](#), also known as [knowledge discovery and data mining](#) (KDD). "Unstructured data" can include emails, videos, photos, social media, and other user-generated content. Data science often requires sorting through a great amount of information and writing algorithms to extract insights from this data.

Contents [\[hide\]](#)

- 1 Overview
- 2 History
- 3 Domain specific interests
- 4 Criticism
- 5 Research areas
 - 5.1 Security Data Science
 - 5.2 Clinical data science
 - 5.3 Genomic data science
 - 5.4 Agriculture
- 6 Further reading
- 7 References

4) new data is available

Overview [\[edit\]](#)

Data science employs techniques and theories drawn from many fields within the broad areas of [mathematics](#), [statistics](#), [information theory](#), and [computer science](#), including [signal processing](#), [probability models](#), [machine learning](#), [statistical learning](#), [data mining](#), [database](#), [data engineering](#), [pattern recognition and learning](#), [visualization](#), [predictive analytics](#), [uncertainty modeling](#), [data warehousing](#), [data compression](#), [computer programming](#), and [high performance computing](#).

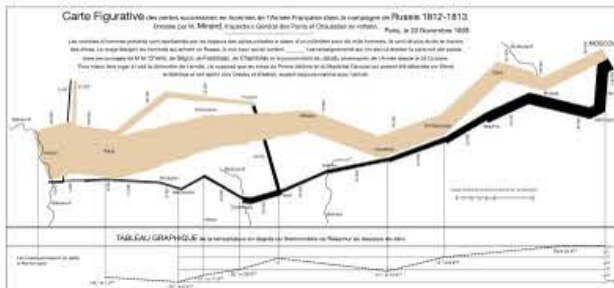
Methods that scale to [Big Data](#) are of particular interest in data science, although the discipline is not generally considered to be restricted to such data. The development of [machine learning](#), a branch of [artificial intelligence](#) used to uncover patterns in data from which predictive models can be developed, has enhanced the growth and importance of data science.

The field of data science utilizes data preparation, statistics, and machine learning to investigate problems in various domains such as agriculture, marketing optimization, fraud detection, setting public policy, etc.

Data scientists use the ability to find and interpret rich data sources; manage large amounts of data despite hardware, software, and bandwidth constraints; merge data sources; ensure consistency of datasets; create visualizations to aid in understanding data; build mathematical models

Part of a series on [Statistics](#)

Data visualization



Major dimensions [\[hide\]](#)

- Exploratory data analysis • Information design
- Interactive data visualization
- Descriptive statistics • Inferential statistics
- Statistical graphics • Plot
- Data analysis • Infographic
- Data science**

Thought leaders [\[hide\]](#)

- John W. Tukey • Edward Tufte

Information graphic types [\[hide\]](#)

- Line chart • Bar chart
- Histogram • Scatterplot
- Boxplot • Pareto chart
- Pie chart • Area chart
- Control chart • Run chart
- Stem-and-leaf display • Cartogram
- Small multiple • Sparkline
- Table

Related Topics [\[hide\]](#)

- Data • Information
- Big data • Database
- Chartjunk • Visual perception
- Regression analysis • Statistical model
- Misleading graph

V • T • E

Specific technologies to watch...



Fast Forward Labs



Natural Language Generation



A Report on the Possibilities



TRENDING NOW
iPad Pro review

f t v r
20
NEW ARTICLES

PREVIOUS STORY
The FCC has changed the definition of broadband

NEXT STORY
The FCC just closed its biggest wireless spectrum auction in history

ENTERTAINMENT TECH WEB REPORT

34
COMMENTS

AP's 'robot journalists' are writing their own stories now

By **Ross Miller** on January 29, 2015 11:55 am [Email](#) [@ohnorosco](#)



THE LATEST HEADLINES

Watch Japan's first commercial jetliner take

REAL ESTATE LISTINGS GENERATOR

Generate listings for a **2** bedroom **1** bathroom apartment in **Upper East Side** with the following amenities: **Laundry in Building**

Storage Available **Roof Deck** **Gym** **Balcony** **Doorman** **Pets Allowed** **Terrace**

Run for **20** seconds.

Generate

TOP SCORED LISTINGS

from the **1350** listings generated

Additional central **laundry** room, cold storage, pet grooming room and bike storage are also available. This luxuriously-appointed **two** bedroom **one** bathroom apartment in the sleek Milan condominium in Midtown East exudes contemporary elegance with double exposures (north and west). Additionally, the building has **storage** units available, bike room, luggage room, and new **laundry** room in the basement, plus a beautiful garden/patio area in the back. Situated in the coveted Sutton area of **Ues**, The Milan condominium tower built in 2005 with its 2-story Jerusalem stone, Bubinga & Jatoba wood lobby and exquisite residences is striking in design

LISTING SCORE

9

Unique **two** Bedroom and **one** Bath with Chef's Eat-in-Kitchen. Adjacent to the kitchen is a full size **laundry** room, and bedroom with full bath currently configured as a gym. Butler's pantry features miles of **storage** and presently has a **washer** (Washer /Dryers are permitted). This 18th floor apartment, with eastern and southern exposures, has direct **Ues** River views.

LISTING SCORE

9

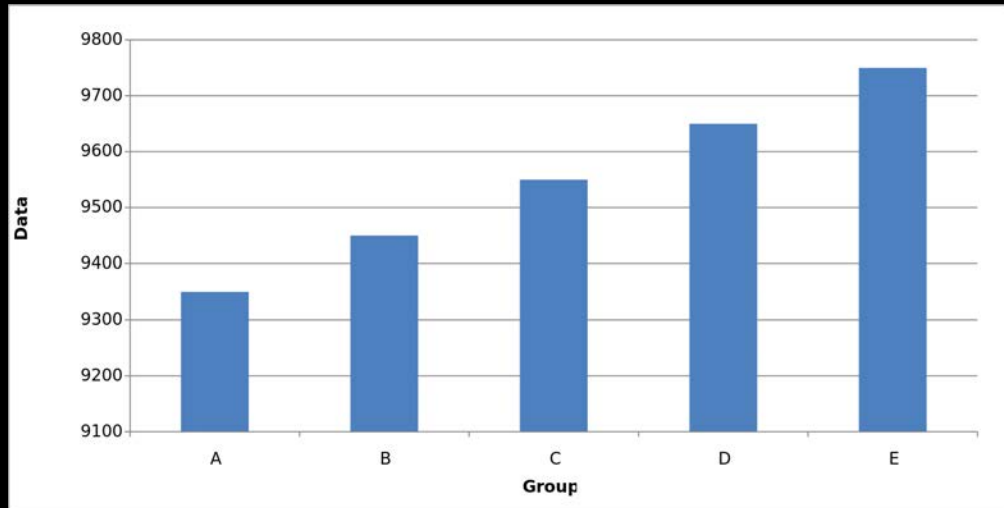
Desirable **Two** Bedroom Prewar Coop. There is basement **storage** available, and while there is central **laundry**, too, residents are allowed to install a Washer/Dryer in their apartment with board approval. Stately 2,021 SF three bedroom, **one** and one half bath double corner residence at 301 East 50th Street. Located in **ues** and close to fine restaurants, MOMA, shopping and all public transportation; this much sought after location has numerous amenities.

LISTING SCORE

8

The real impact will be in making complex data simple.

11374,73.25165072760954	11422,65.03628419255165
11375,76.02168800410799	11423,74.05132710911768
11377,81.22613070574234	11937,72.28167427701676
11378,76.68755893219648	11426,44.68227272727274
11379,72.43703076239288	11427,71.65128225806453
11385,81.37809552925839	11428,69.63122756410257
7302,82.04176470588239	11429,89.3414912280702
11411,65.26092268715267	11432,83.111941675198
11412,70.22615286661932	11433,59.18077011494254
11413,52.38190997811611	11434,74.45432391081869
11414,66.1278529660406	11435,102.4839473798388
11415,97.26002490653225	11436,62.50399652043694
11417,53.88985294117648	11968,83.41529411764708
11418,70.07631771729758	11975,49.78700698757765
11419,49.67555350978138	
11420,94.04366646521625	
11421,62.72676963129751	

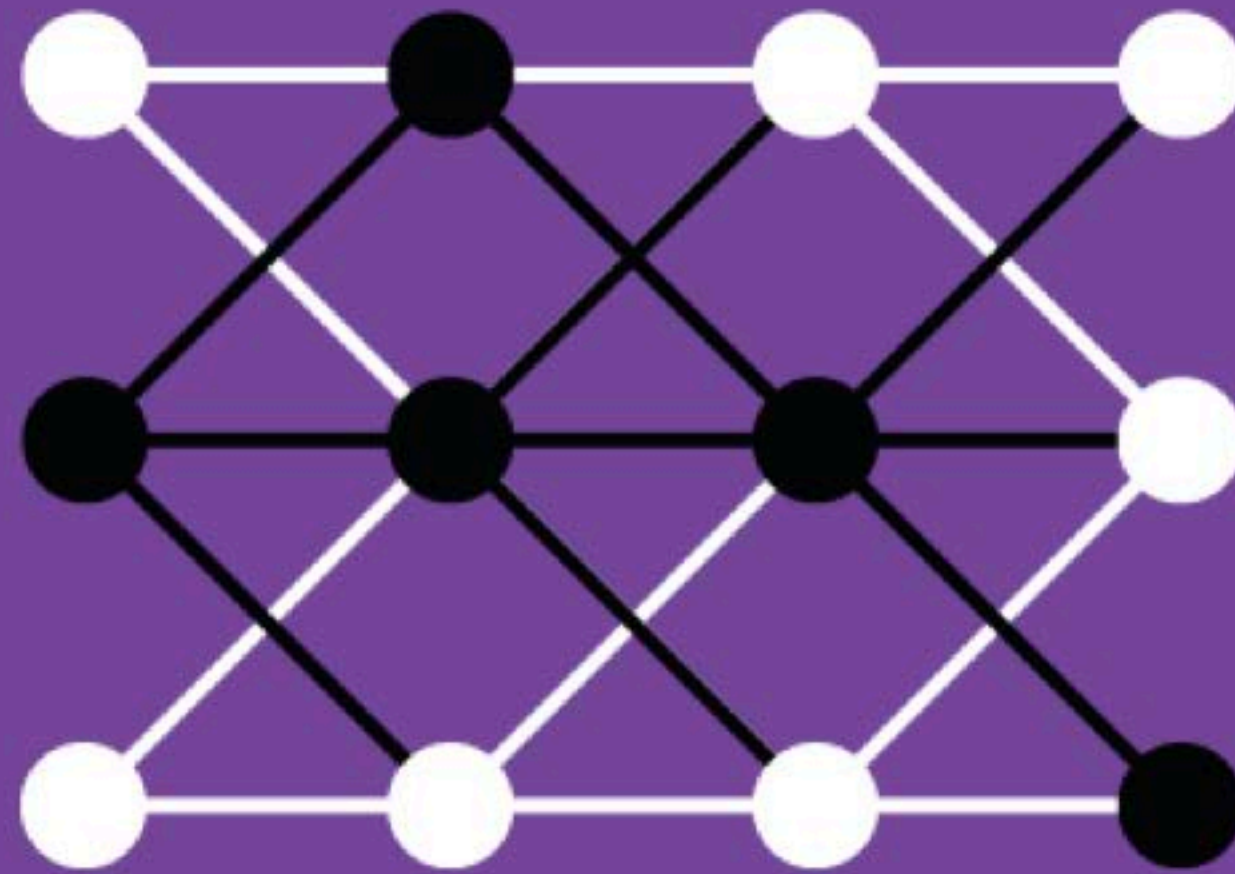


There's been an increase in sales!

Fast Forward Labs

FF

Deep Learning: Image Analysis





tourismireland

loves taking photos of

natural depressions



shops



fortifications



rail fences



generated from **tourismireland's Instagram**

Now we can run analytics on rich
media data.



Mark Harris [Follow](#)

Mostly British, mostly tech journalist reporting on the US from mostly rainy Seattle

Jun 6 · 16 min read

Say Goodbye To Your Highly Skilled Job. It's Now a "Human Intelligence Task."

Digital crowdworkers don't only do menial tasks like data entry. They're smart, capable, and hungrier than any algorithm. And they work for cheap.



♥ 216 💬 6



[Next story](#)
[iMcLaren: A Luxury Wearable for ...](#)

Harry K. sits at his desk in Vancouver, Canada, scanning sepia-tinted swirls, loops and blobs on his computer screen. Every second or so, he jabs at his mouse and adds a fluorescent dot to the image. After a minute, a new image pops up in front of him.

Harry is tagging images of cells removed from breast cancers. It's a painstaking job but not a difficult one, he says: "It's like playing Etch A Sketch or a video game where you color in certain dots."

Harry found the gig on Crowdfunder, a crowdworking platform. Usually that cell-tagging task would be the job of pathologists, who typically start their careers with annual salaries of around \$200,000—an hourly wage of about \$80. Harry, on the other hand, earns just four cents for annotating a batch of five images, which takes him between two to eight minutes. His hourly wage is about 60 cents.

Granted, Harry can't perform most of the tasks in a pathologist's repertoire. But in 2016—11 years after the launch of the ur-platform, Amazon Mechanical Turk—crowdworking (sometimes also called crowdsourcing) is eating into increasingly high-skilled jobs. The engineers who are developing this model of labor have a bold ambition to atomize entire careers into micro-tasks that almost anyone, anywhere in the world, can carry out online. They're banking on the idea that any technology that can make a complex process 100 times cheaper, as in Harry's case, will spread like wildfire.



Top Highlights

But in 2016 -- 11 years after the launch of the ur-platform, Amazon Mechanical Turk -- crowdworking (sometimes also called crowdsourcing) is eating into increasingly high-skilled jobs.

Score: 28 · Scroll: 14%

The engineers who are developing this model of labor have a bold ambition to atomize entire careers into micro-tasks that almost anyone, anywhere in the world, can carry out online.

Score: 30 · Scroll: 15%

They're banking on the idea that any technology that can make a complex process 100 times cheaper, as in Harry's case, will spread like wildfire.

Score: 28 · Scroll: 15%

But as the tech conversation has fixated on how artificial intelligence will affect the job market, crowdwork

Harry K. sits at his desk in Vancouver, Canada, scanning sepia-tinted swirls, loops and blobs on his computer screen. Every second or so, he jabs at his mouse and adds a fluorescent dot to the image. After a minute, a new image pops up in front of him.

Harry is tagging images of cells removed from breast cancers. It's a painstaking job but not a difficult one, he says: "It's like playing Etch A Sketch or a video game where you color in certain dots."

Harry found the gig on Crowdfunder, a crowdworking platform. Usually that cell-tagging task would be the job of pathologists, who typically start their careers with annual salaries of around \$200,000—an hourly wage of about \$80. Harry, on the other hand, earns just four cents for annotating a batch of five images, which takes him between two to eight minutes. His hourly wage is about 60 cents.

Granted, Harry can't perform most of the tasks in a pathologist's repertoire. But in 2016—11 years after the launch of the ur-platform, Amazon Mechanical Turk—crowdworking (sometimes also called crowdsourcing) is eating into increasingly high-skilled jobs. The engineers who are developing this model of labor have a bold ambition to atomize entire careers into micro-tasks that almost anyone, anywhere in the world, can carry out online. They're banking on the idea that any technology that can make a complex process 100 times cheaper, as in Harry's case, will spread like wildfire.

Perhaps it's inevitable that in a few years, software will swallow up these jobs, too. But as the tech conversation has fixated on how artificial intelligence will affect the job market, crowdwork has quietly grown in

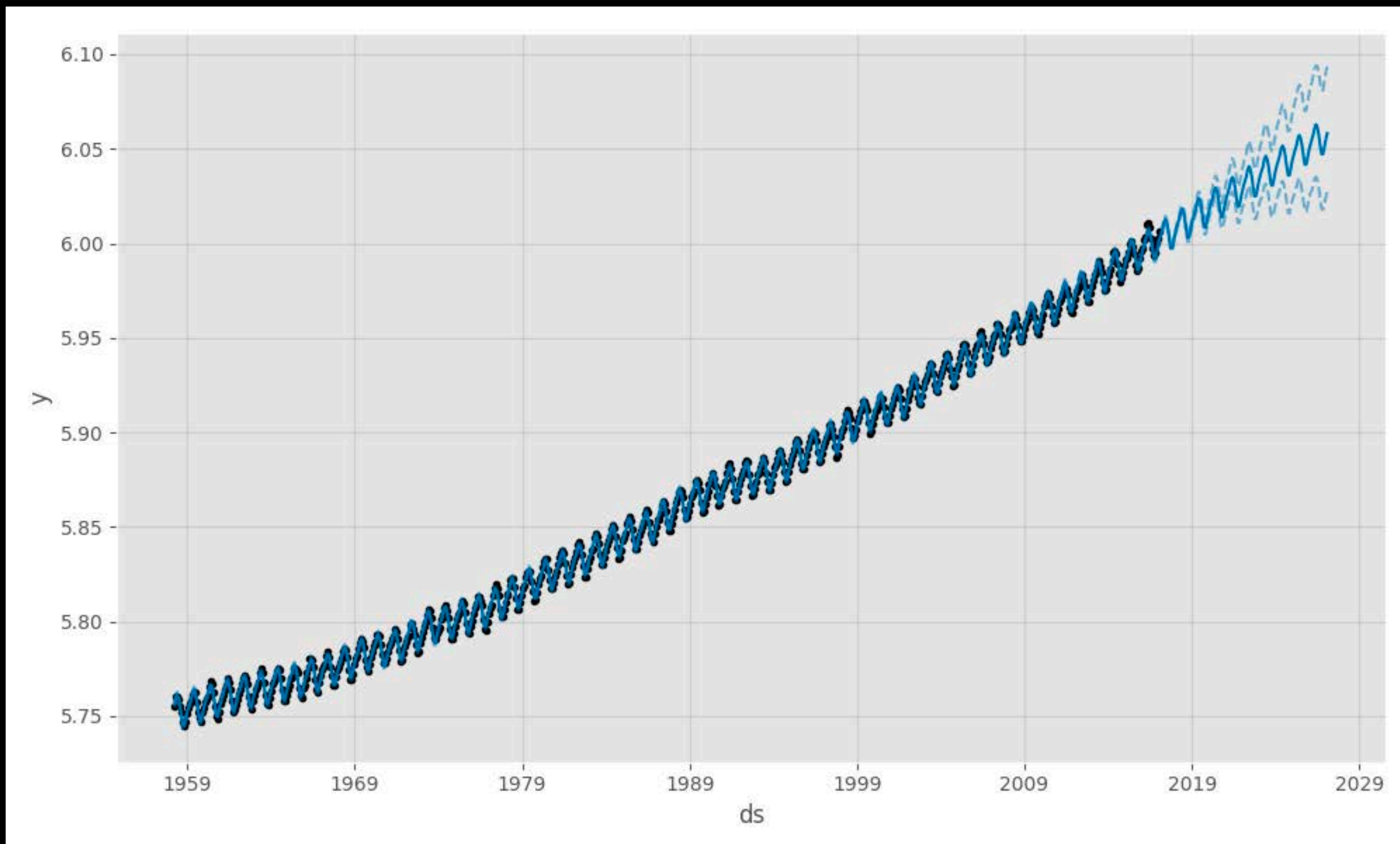
Moving beyond counting words to
computable representations of
concepts.

Fast Forward Labs



Probabilistic Programming





atmospheric co2

Loan Applications

New Rate: 25.0% Set Rate

Rate: 20.0% Approve Loan

Exp. Profit	% Repaid	Upside	Downside
ID: 1395617 \$30,000 at 18%			
10k	78%	19k	25k
ID: 1399387 \$30,000 at 25%			
10k	58%	29k	27k
ID: 859802 \$20,000 at 20%			
9k	78%	15k	17k
ID: 4105394 \$20,000 at 20%			
7k	71%	15k	17k
ID: 1125390 \$10,000 at 20%			
4k	75%	7k	8k
ID: 1456017 \$10,000 at 20%			
3k	69%	7k	9k
ID: 1153671 \$5,000 at 20%			
2k	71%	4k	4k

Loan 859802

Applicant Information

The applicant has an annual income of **\$110,000**. Their monthly debt payments are **17%** of their monthly income, and they have used **33%** of their available credit. They intend to use the loan for **debt consolidation**.

Repayment Probabilities

Expected Profit	% Repaid	Upside	Downside
\$9,000	78%	\$14,600	\$16,600
\$9,600	69%	\$19,100	\$17,900
+\$700	-9%	+\$4,500	+\$1,300



Probabilistic Real Estate

Where can I buy a property for

\$1,000,000

FILTER LIST BY BOROUGH

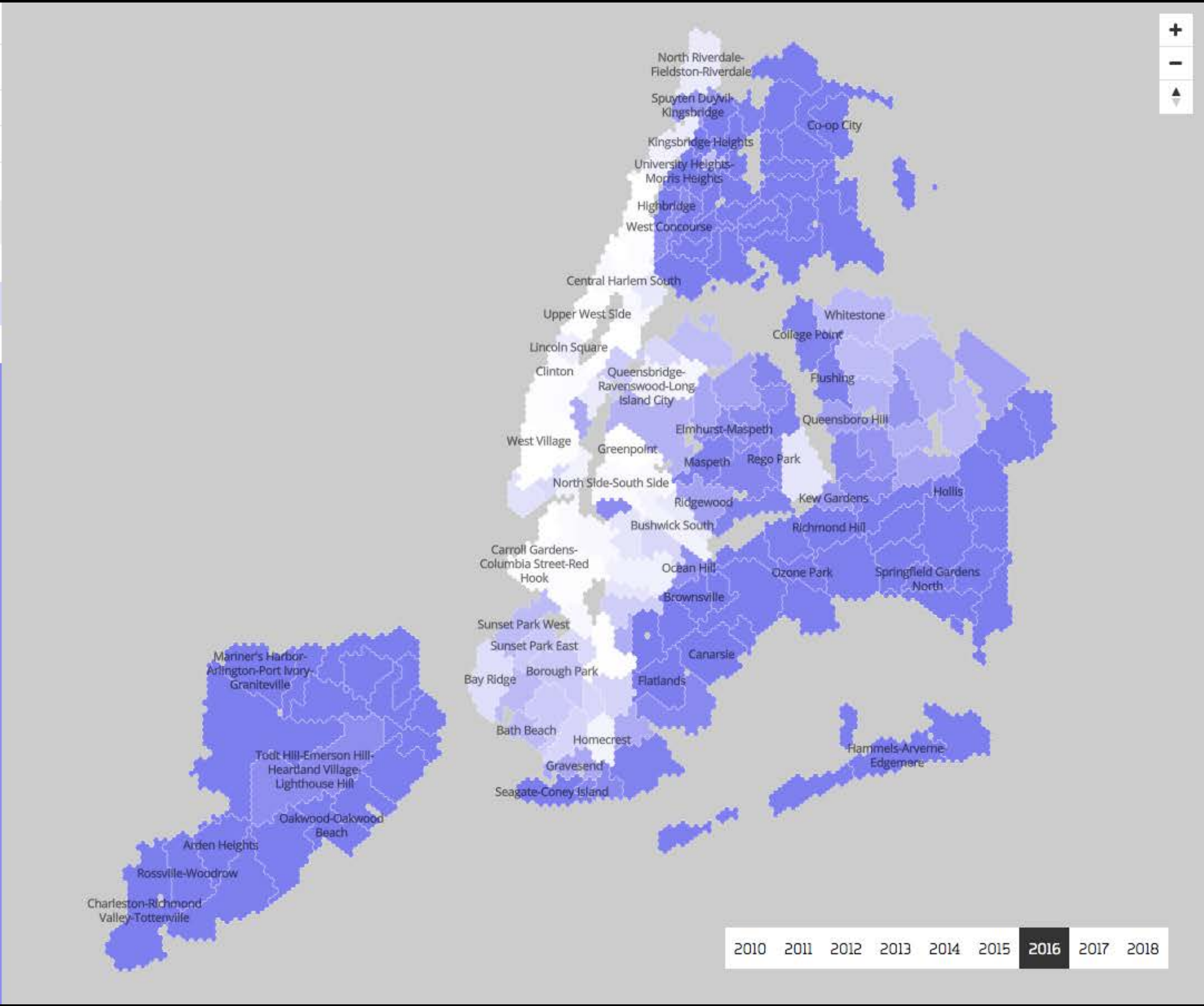
Brooklyn Bronx Manhattan Queens Staten Isl.

FILTER LIST BY PROBABILITY

High 111 Average 22 Low 54

NEIGHBORHOODS

BX Crotona Park East	100%
BK East New York	100%
SI Great Kills	100%
SI Grymes Hill-Clifton-Fox Hills	100%
QN Hammels-Arverne-Edgemere	100%
BX Hunts Point	100%
QN Laurelton	100%
SI Oakwood-Oakwood Beach	100%
BX Parkchester	100%
SI Port Richmond	100%
BX Soundview-Castle Hill-Clason Point-Harding Park	100%
QN St. Albans	100%
BK Starrett City	100%
SI West New Brighton-New Brighton-St. George	100%
SI Westerleigh	100%
BX Williamsbridge-Olinville	100%
SI Arden Heights	100%
BX Bronxdale	100%
BK Brownsville	100%
BK East Flatbush-Farragut	100%
BK Flatlands	100%



Probabilistic Real Estate

Where can I buy a property for

\$1,000,000

NEIGHBORHOOD **x**

Bedford

Brooklyn

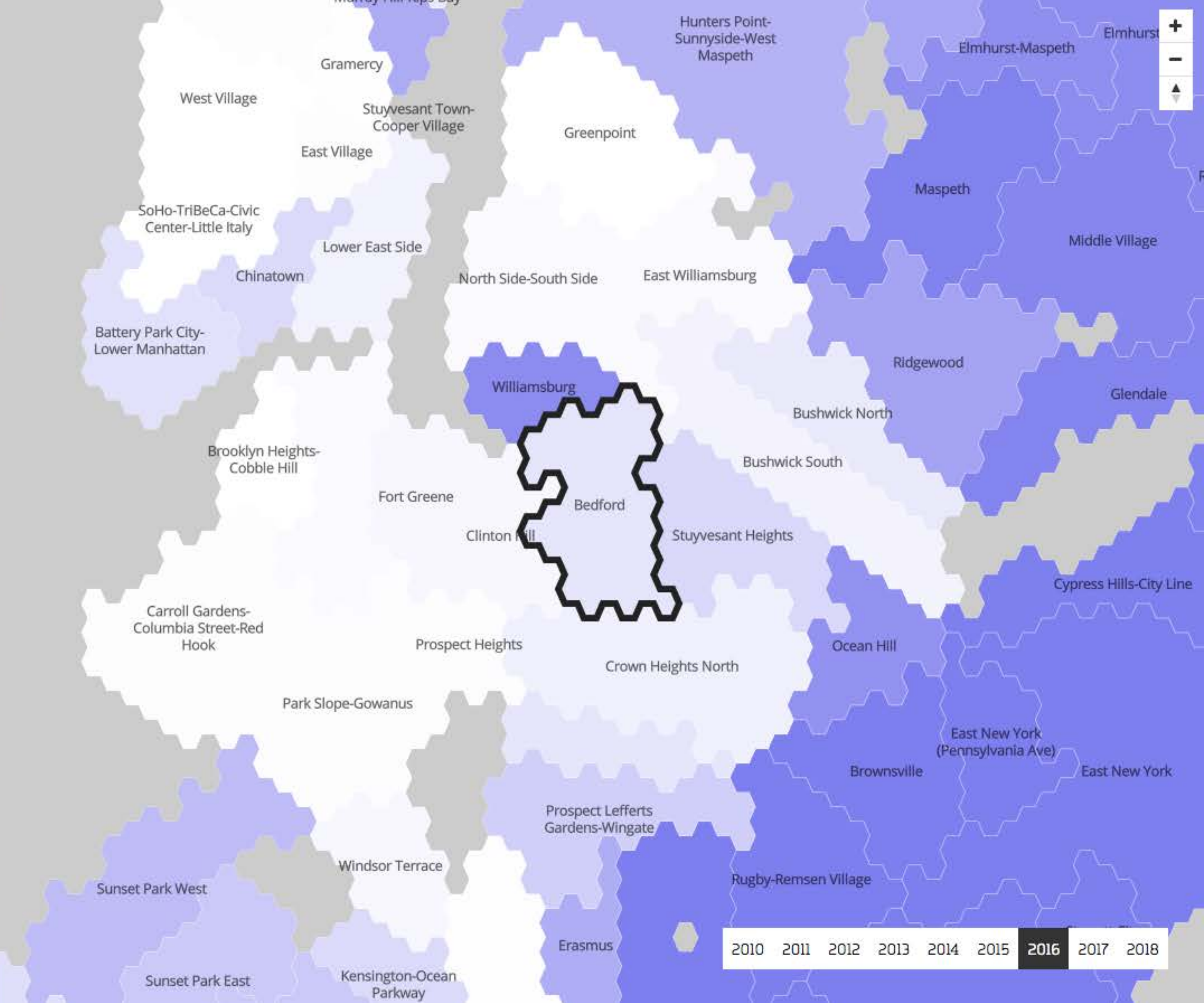
PROBABILITY

Low **19%**

PROBABILITY DISTRIBUTION FOR 2016

PROBABILITY BY YEAR

Year	Probability
2010	90%
2011	90%
2012	90%
2013	90%
2014	85%
2015	55%
2016	20%
2017	5%
2018	5%



Tools for combining domain
knowledge with small data to
make risky decisions.



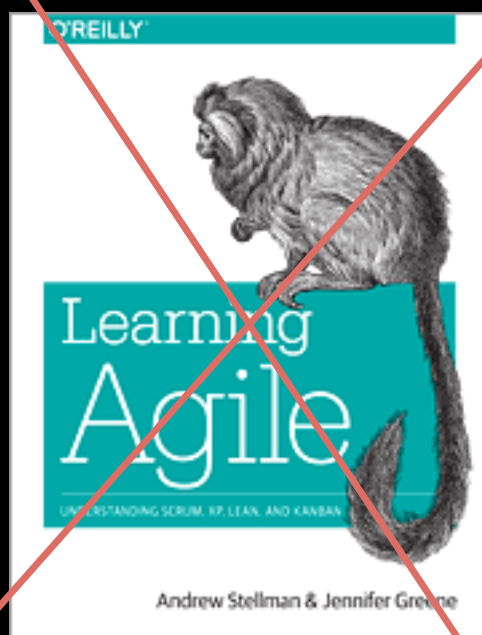
Practical advice.

First, find high-value tasks we already do.

Then, apply to things that were never viable
before.

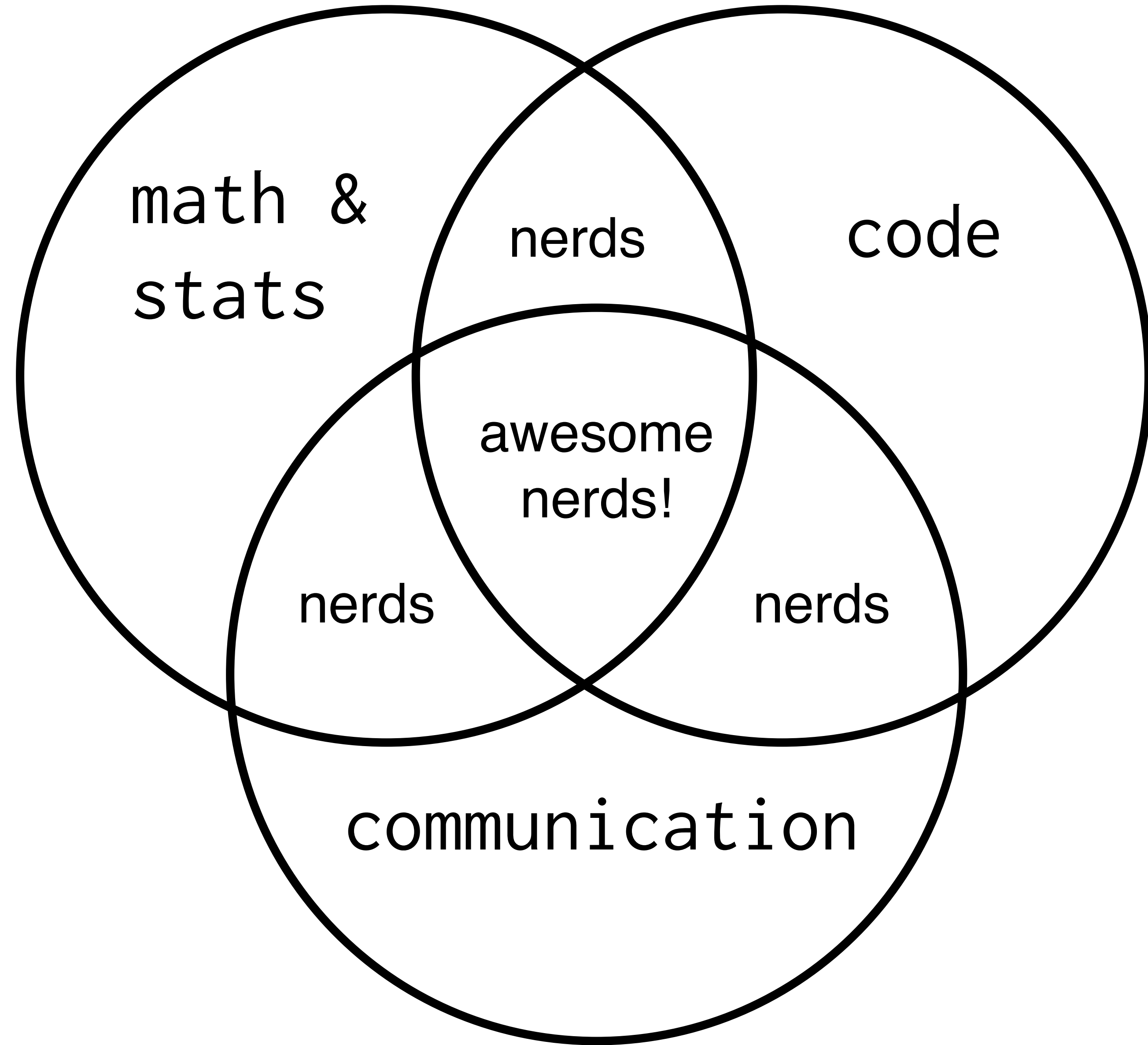
Create feedback loops.

Use an **Experimental**
Development Process.



New Talent Requirements.





CAUTION



THIS IS SPARTA

There's potential for mischief.

Human relationships have value.



AI Product Gap!



And now: The Future!





- STATS
1. PICKING THE BEST DRINK
 2. DRINKING THE DRINK
 3. TASTING THE DRINK
 4. RECOMMENDING THE DRINK
 5. SERVING THE DRINK

03:51
CURRENT WAITING TIME



STOP

DECREASING
SPEED

STEADY

INCREASING
SPEED

FULL
SPEED

MAKE
SMOKE

STOP
PUMPING
OVER.

UNITS

STOP
MAKING
SMOKE

PUMP
OVER

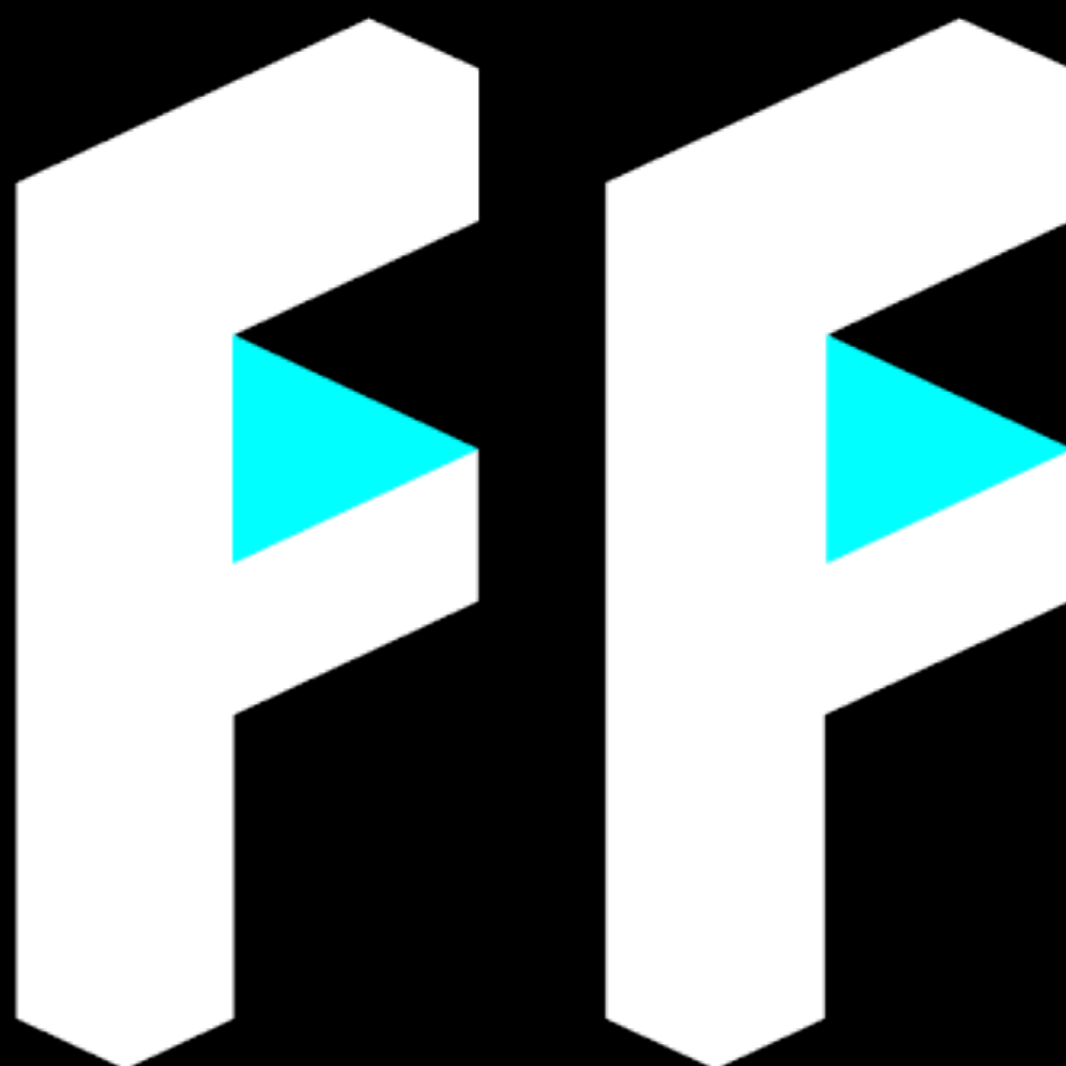
CONNECTED

A. ROBINSON & CO. LTD.

LIVERPOOL & GLASGOW

720





FAST FORWARD
LABS

Hilary Mason

@hmason
hilary@fastforwardlabs.com