

Data Science: The Big Picture

@MatthewRenze
#SQLServerUserGroupDubai







The Economist

FEBRUARY 27TH-MARCH 5TH 2010

Gordon Brown's pitch
What went wrong at RBS
Genetically modified crops blossom
The EU woos Russia
The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LOMK
Published: August 5, 2009

MOUNTAIN VIEW, Calif. — At Harvard, Carrie Grimes majored in anthropology and archaeology and ventured to places like Honduras, where she studied Mayan settlement patterns by mapping where artifacts were found. But she was drawn to what she calls “all the computer and math stuff” that was part of the job.

- [TWITTER](#)
- [LINKEDIN](#)
- [COMMENTS \(58\)](#)
- [SIGN IN TO E-MAIL](#)

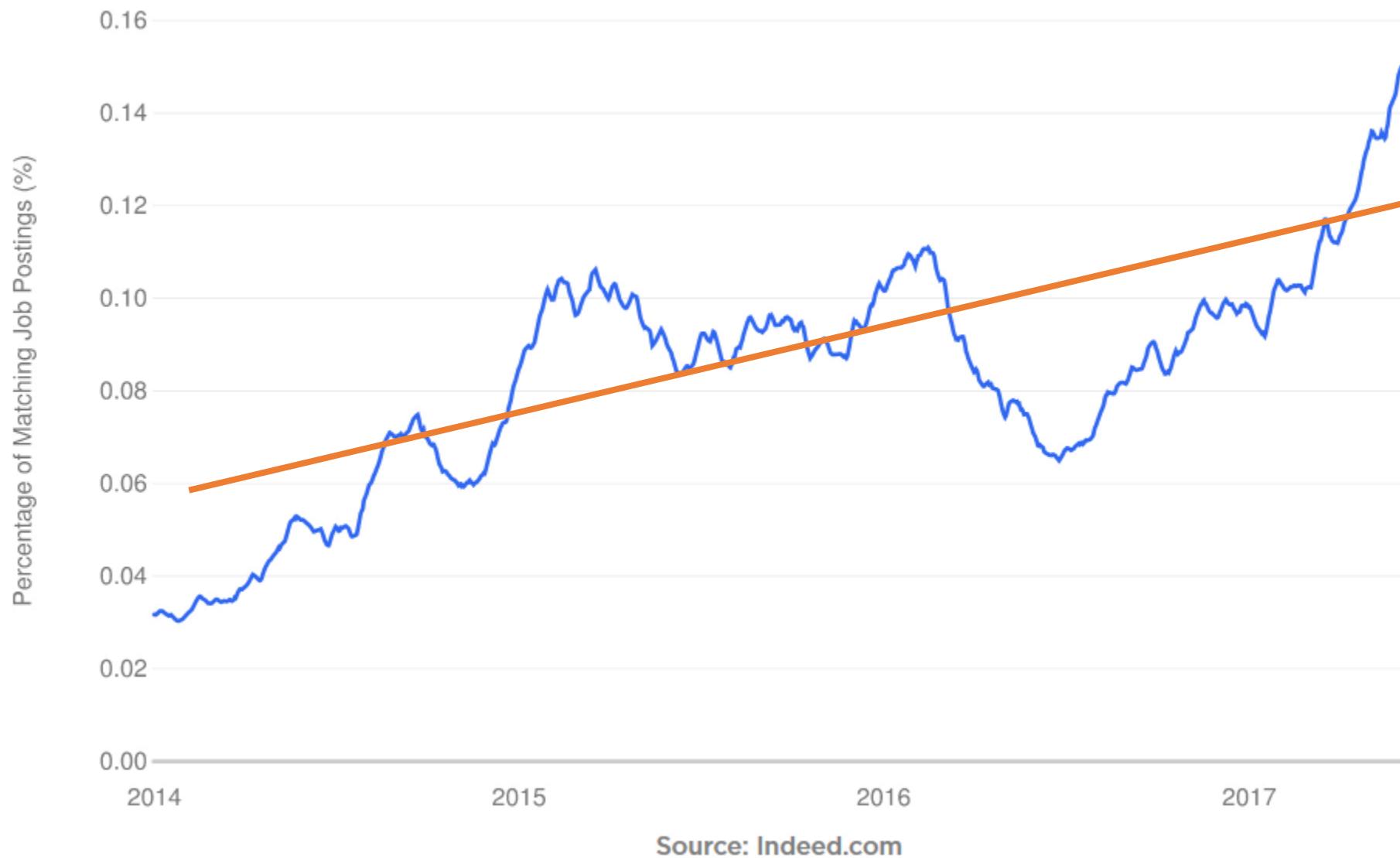
Data Scientist: *The Sexiest Job of the 21st Century*

Meet the people who can coax treasure out of messy, unstructured data.

by Thomas H. Davenport
and D.J. Patil

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—and you probably leave early.”

Job Postings for Data Scientists

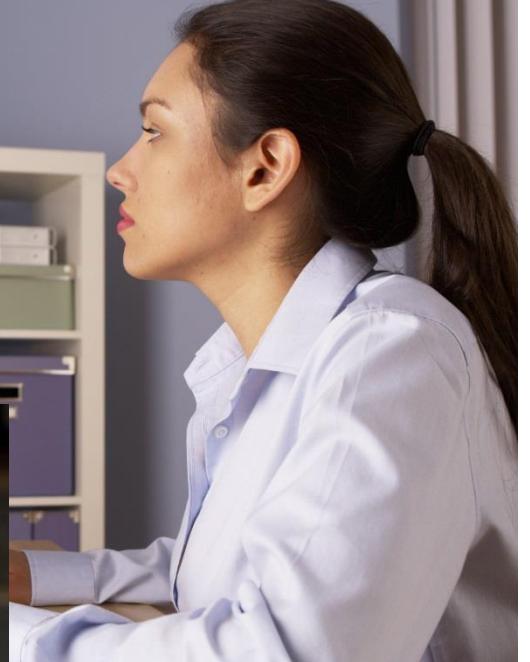
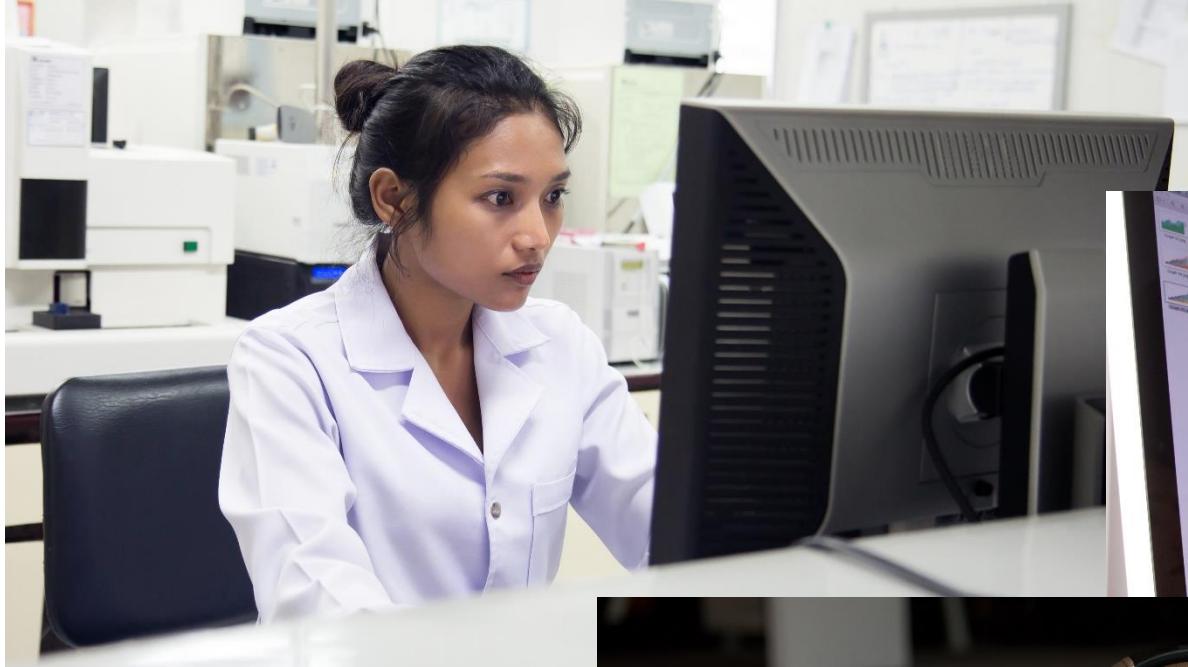


Top-paying Tech Skills

Skill	2016	Change
HANA (High Performance Analytical Application)	\$ 128,958	-3.3%
MapReduce	\$ 125,009	-0.3%
Cloud Foundry	\$ 124,038	n/a
Hbase	\$ 123,934	5.7%
Omnigraffle	\$ 123,782	-1.9%
Cassandra	\$ 123,459	2.2%
Apache Kafka	\$ 122,728	n/a
SOA (Service Oriented Architecture)	\$ 122,094	-1.9%
Ansible	\$ 121,382	n/a
Jetty	\$ 120,978	1.3%
PaaS (Platform as a Service)	\$ 120,403	-4.4%
Elasticsearch	\$ 120,002	n/a
ABAP (Advanced Business Application Programming)	\$ 119,961	0.5%
NoSQL	\$ 119,498	1.3%
CMMI (Capability Maturity Model Integration)	\$ 119,466	-0.6%
Amazon Redshift	\$ 119,197	n/a
Pig	\$ 119,118	-4.2%
Solr	\$ 119,032	0.1%
Cloudera	\$ 118,896	-9.0%
Docker	\$ 118,873	0.2%

Skill	2016	Change
Amazon Route 53	\$ 118,828	n/a
Hadoop	\$ 118,625	-2.5%
Hive	\$ 118,589	-1.3%
Korn Shell	\$ 118,273	1.4%
PMBok (Project Management Body of Knowledge)	\$ 118,233	0.7%
Dynamo DB	\$ 118,119	n/a
Groovy	\$ 117,897	-0.1%
IaaS (Infrastructure as a Service)	\$ 117,422	n/a
JAX-RS (Java API RestFUL Services)	\$ 116,997	n/a
RabbitMQ	\$ 116,909	n/a
JDBC (Java Database Connectivity)	\$ 116,833	2.0%
SOX (Sarbanes Oxley)	\$ 116,743	0.6%
Objective C	\$ 116,667	2.5%
FCoE (Fibre Channel over Ethernet)	\$ 116,145	7.2%
UML (Unified Modeling Langauge)	\$ 115,285	-3.6%
XSLT (Extensible Stylesheet Language Transformations)	\$ 115,089	3.5%
Redis	\$ 114,922	2.8%
ETL (Extract Transform and Load)	\$ 114,892	2.6%
SDN (Software Defined Network)	\$ 114,739	-2.3%
Informatica	\$ 114,143	1.1%

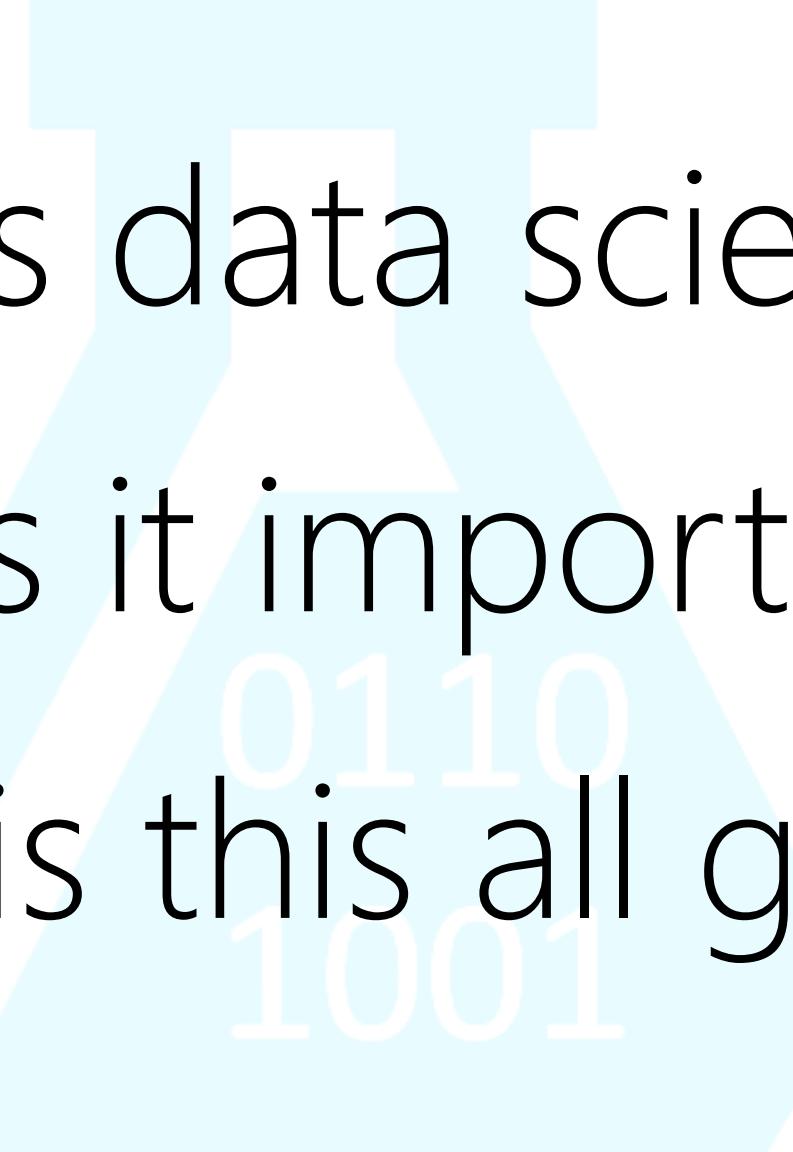
Source: Dice Salary Survey 2017









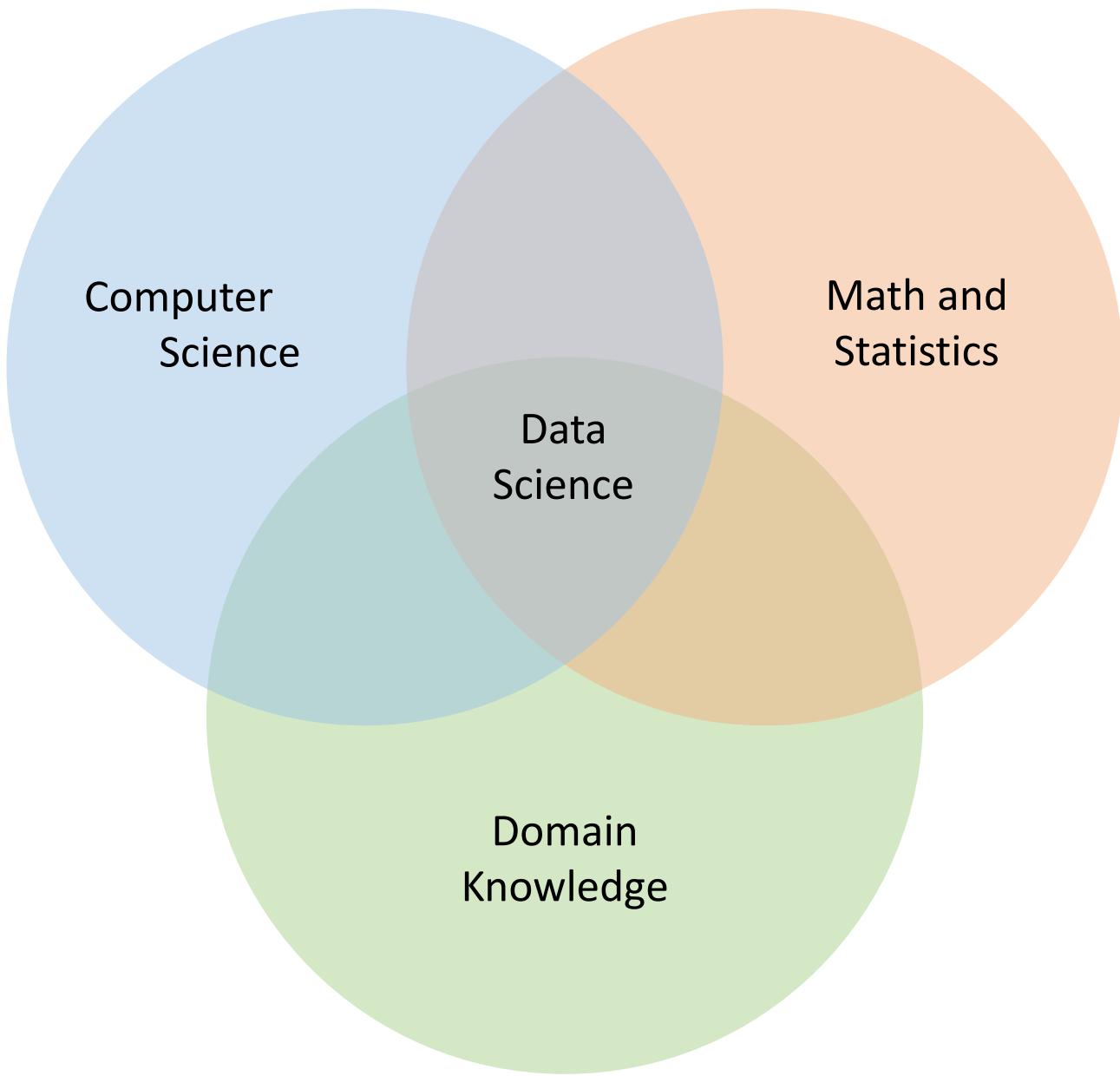


What is data science?

Why is it important?

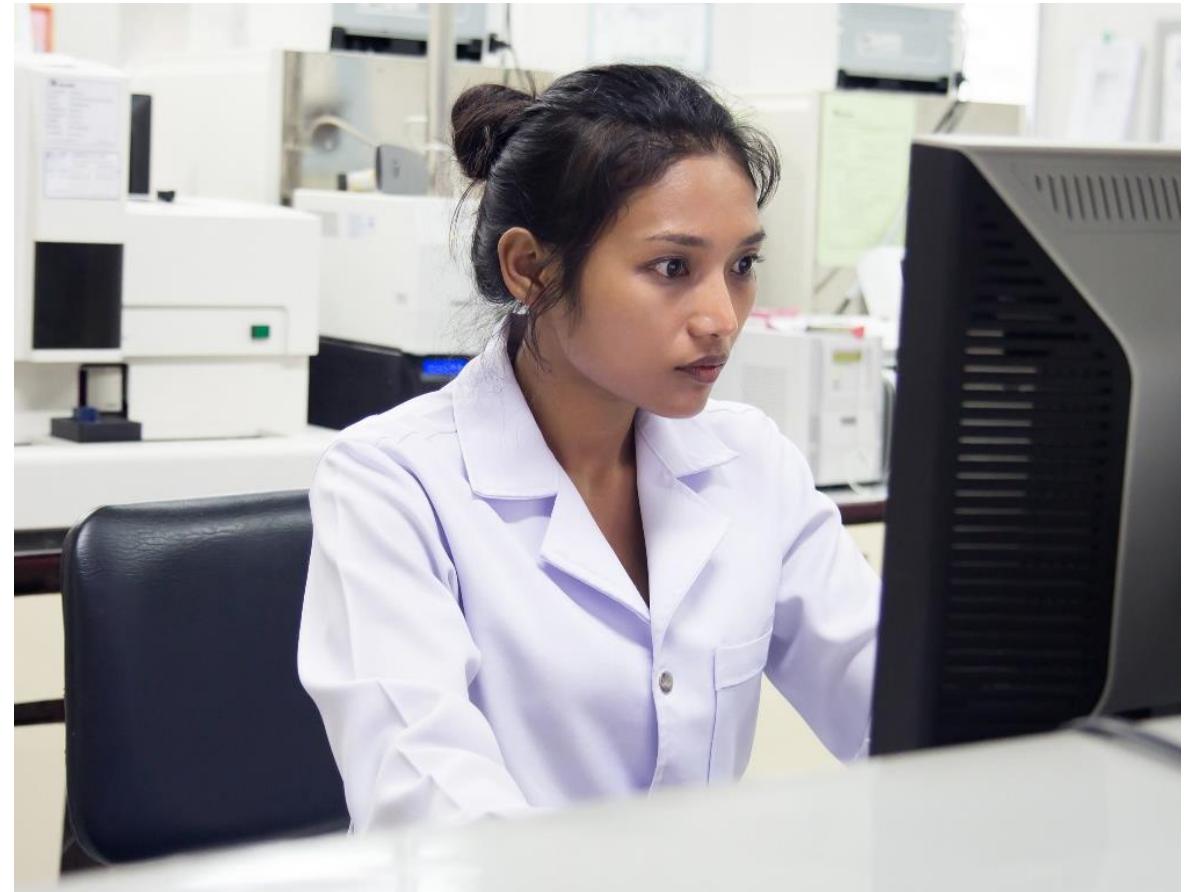
Where is this all going?

What is data science?



What Is a Data Scientist?

Performs data science
More than a scientist
More than an analyst
More than a developer



What skills are necessary?

Data Science Skills

Programming

Working with data

Descriptive statistics

Data visualization

Data Science Skills

Programming

Working with data

Descriptive statistics

Data visualization

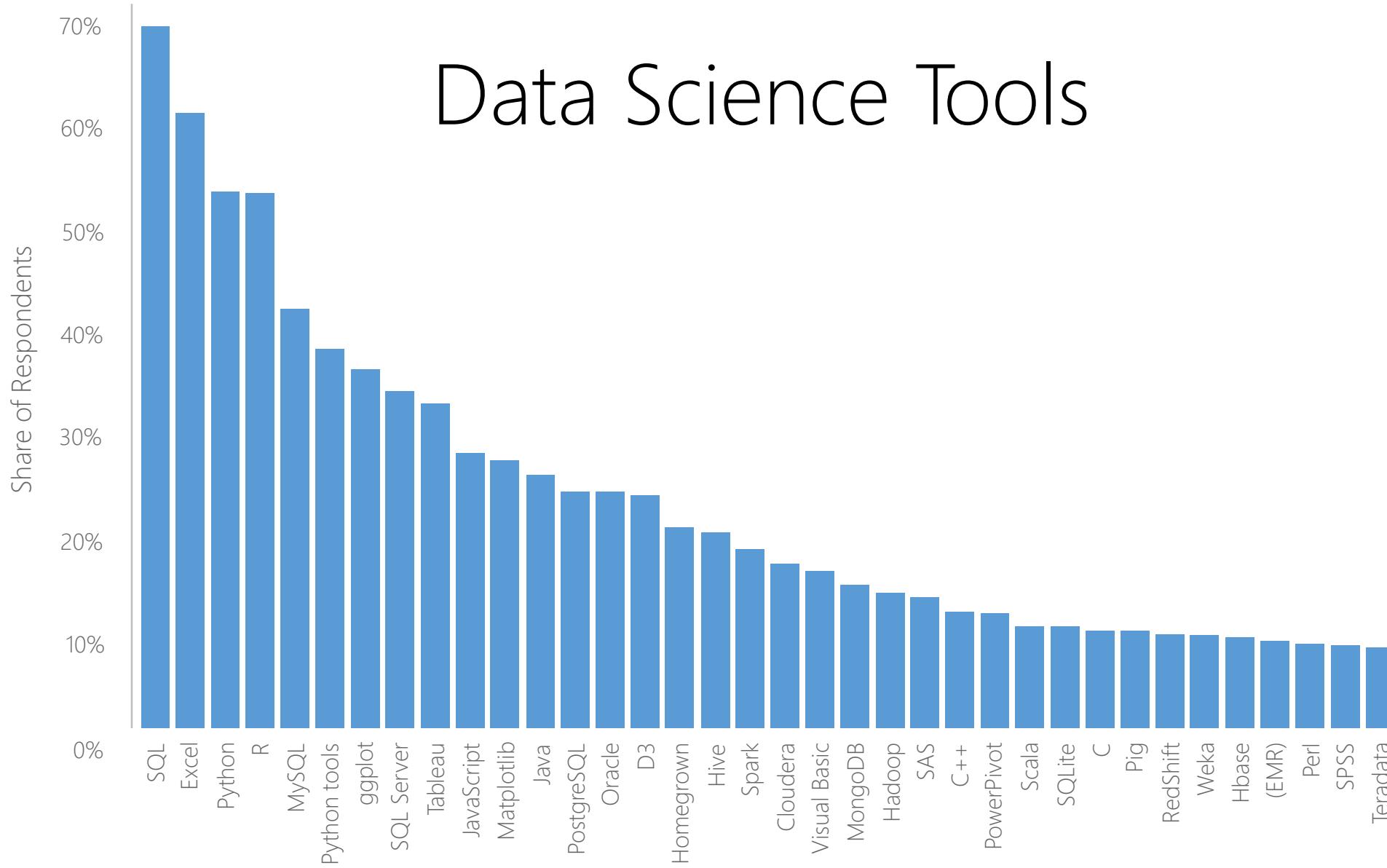
Statistical modeling

Handling Big Data

Machine learning

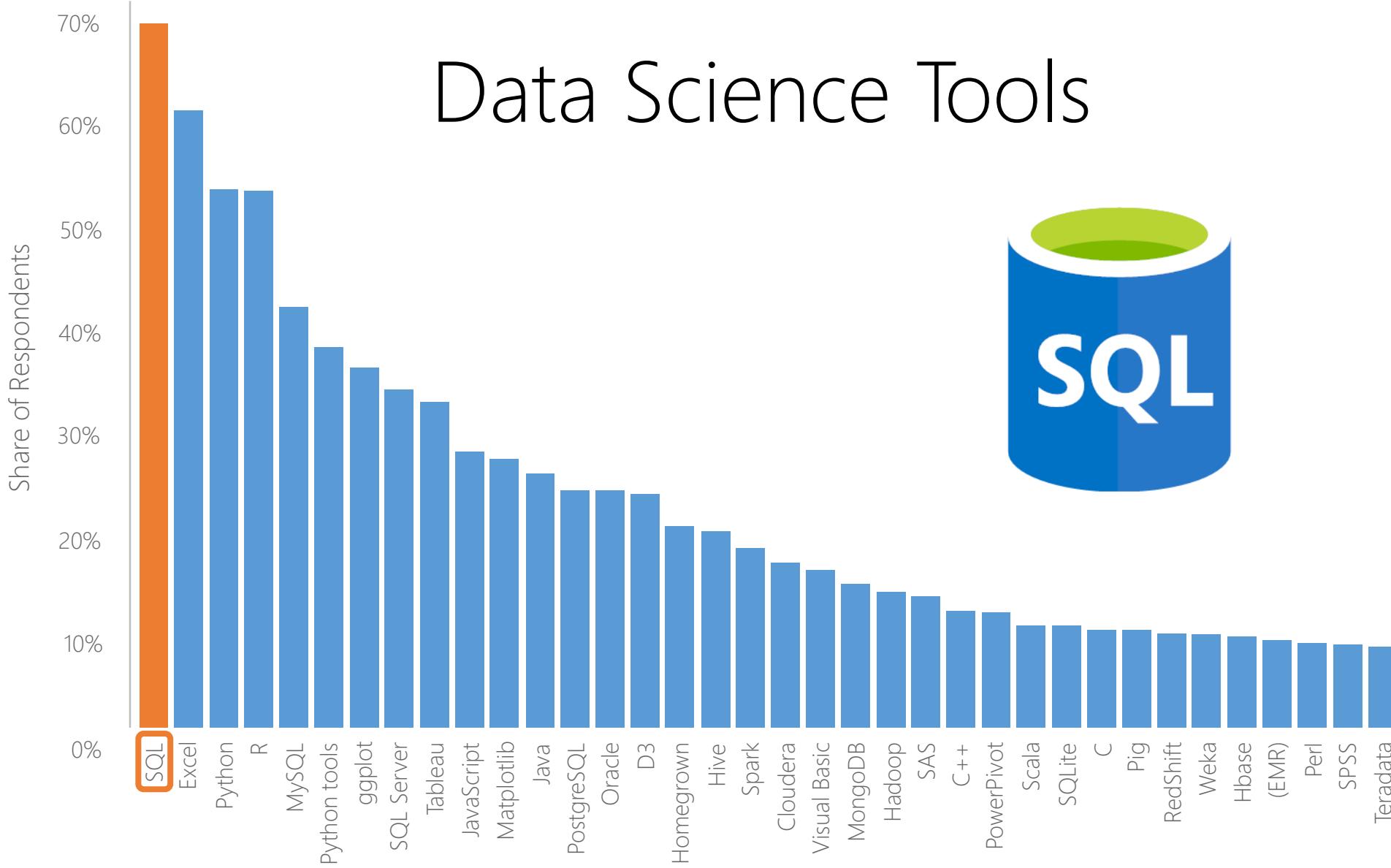
Deploying to production

What tools are used?



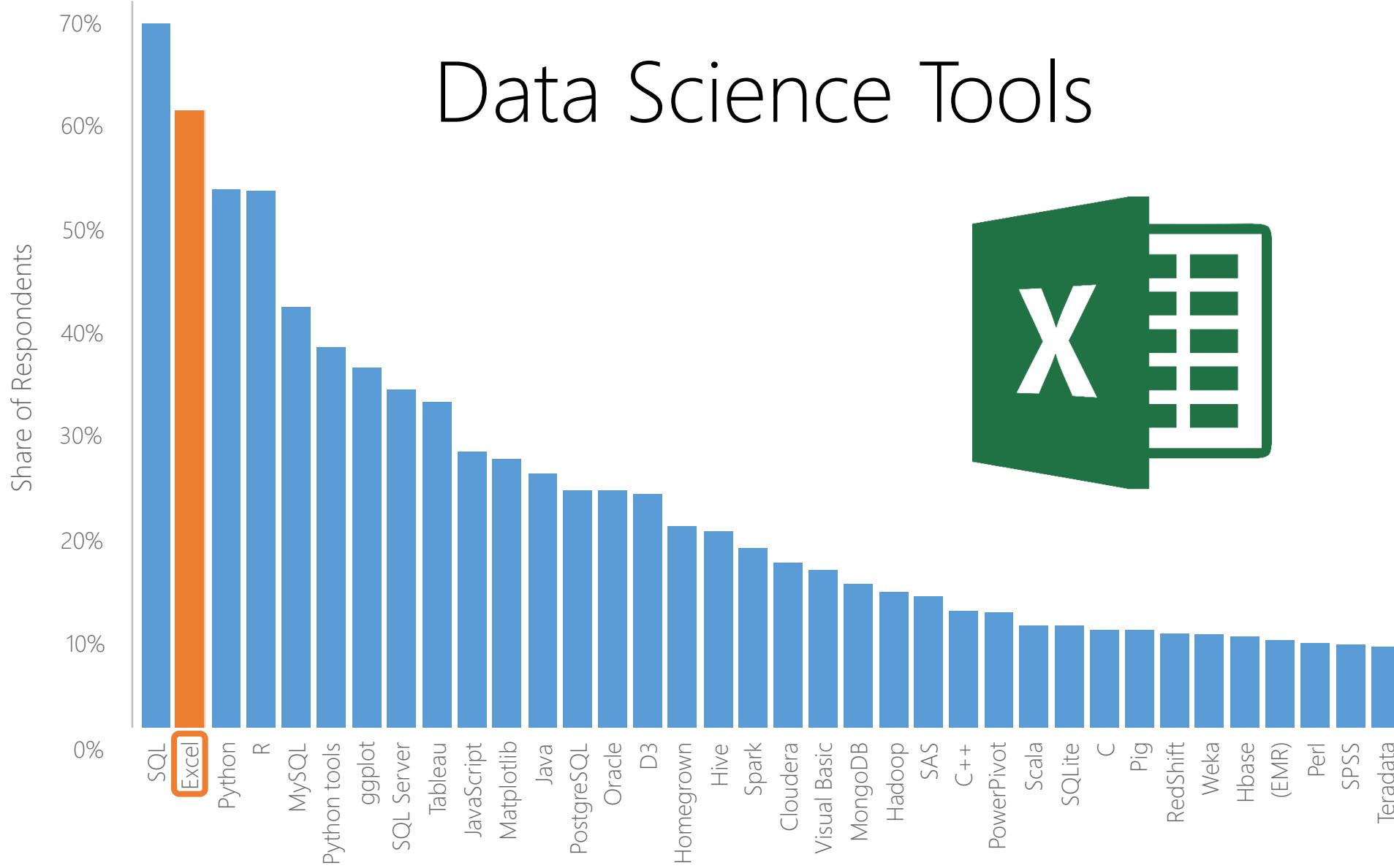
Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey



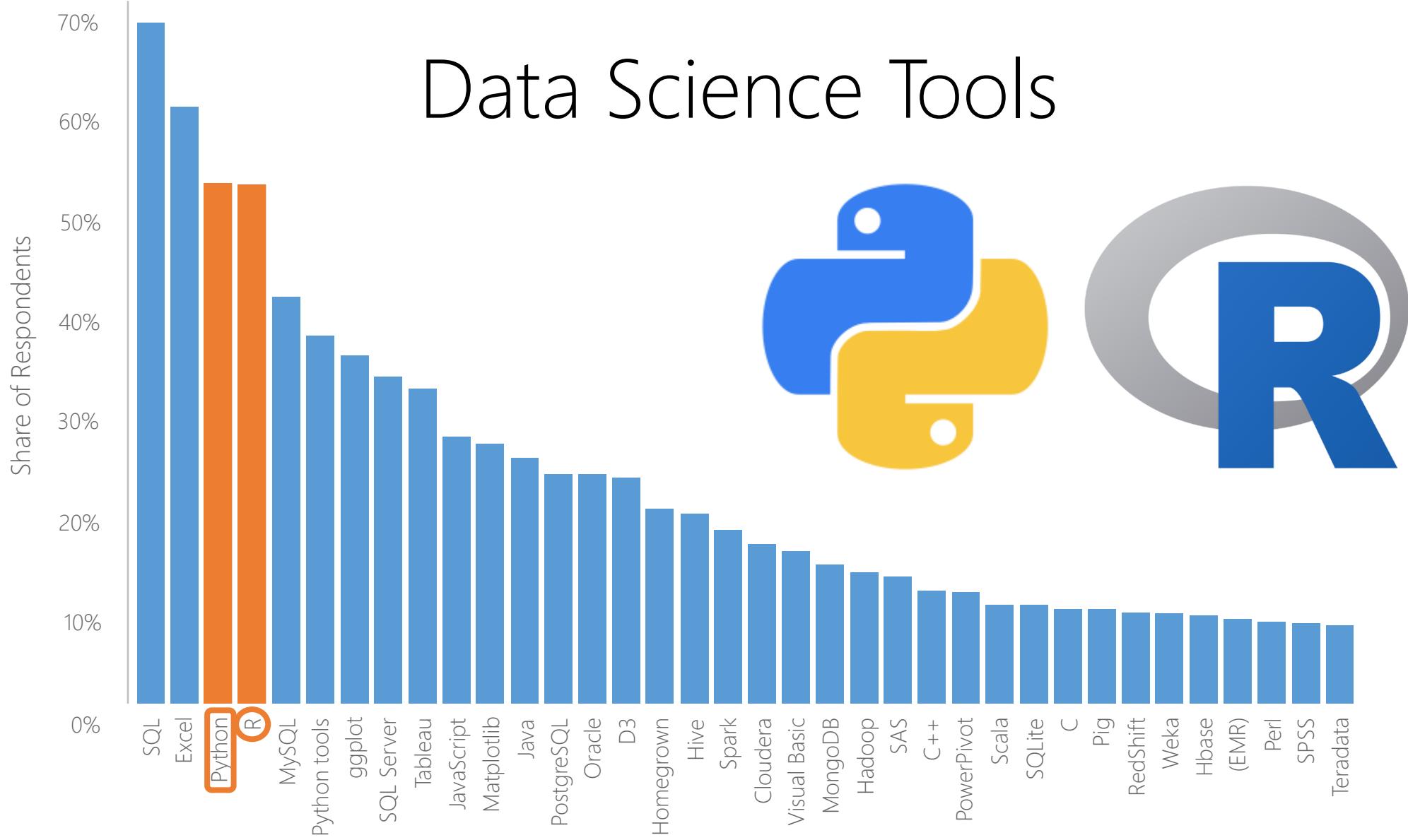
Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey



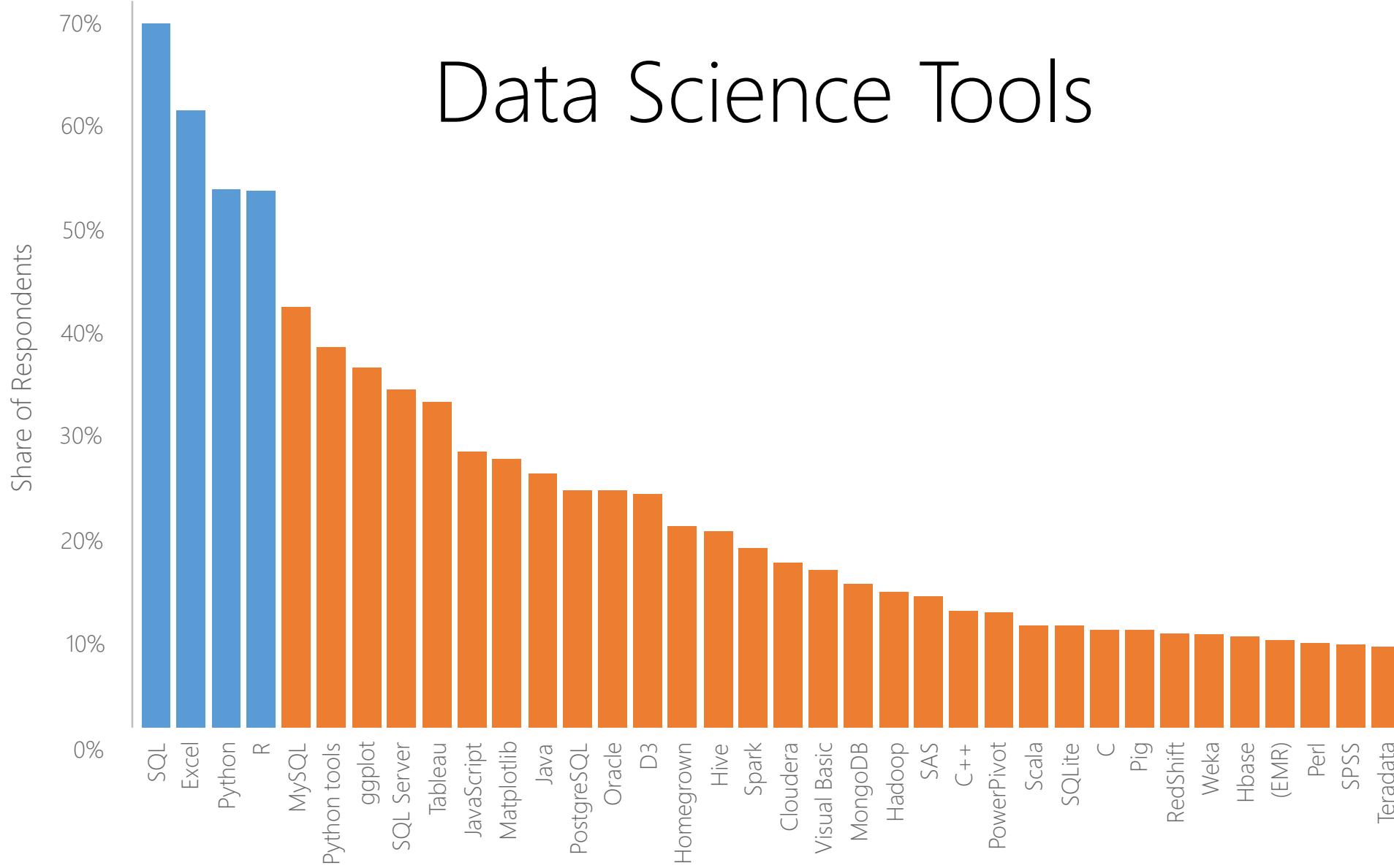
Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey



Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey



Tool: language, platform, analytics

Source: O'Reilly 2015 Data Science Salary Survey

How is data science performed?

The Data Science Process



Data

The Data Science Process

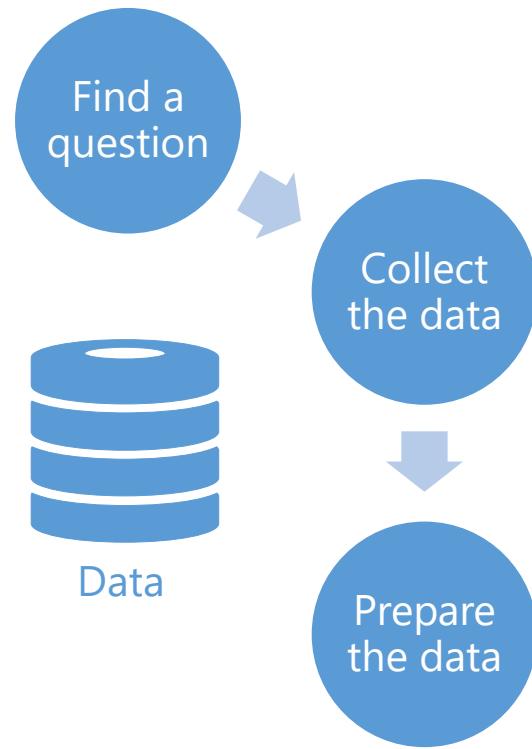


Data

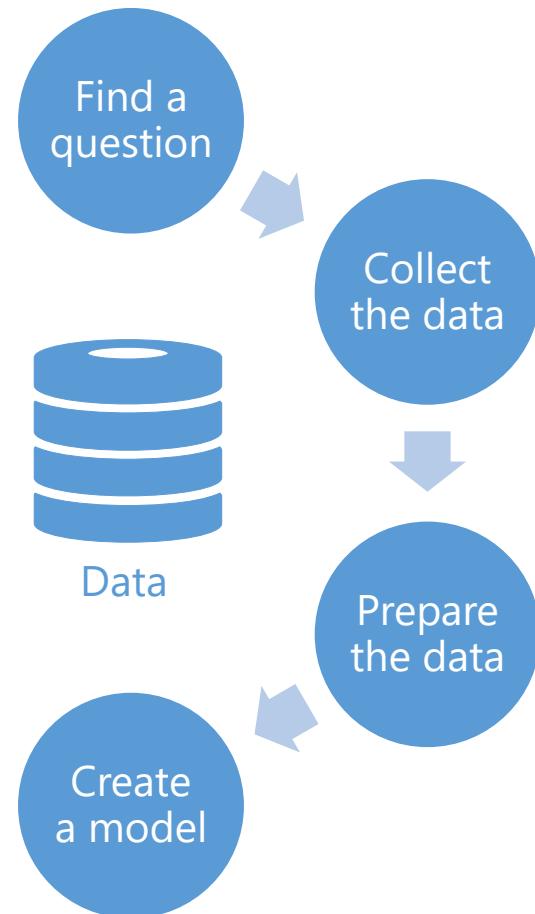
The Data Science Process



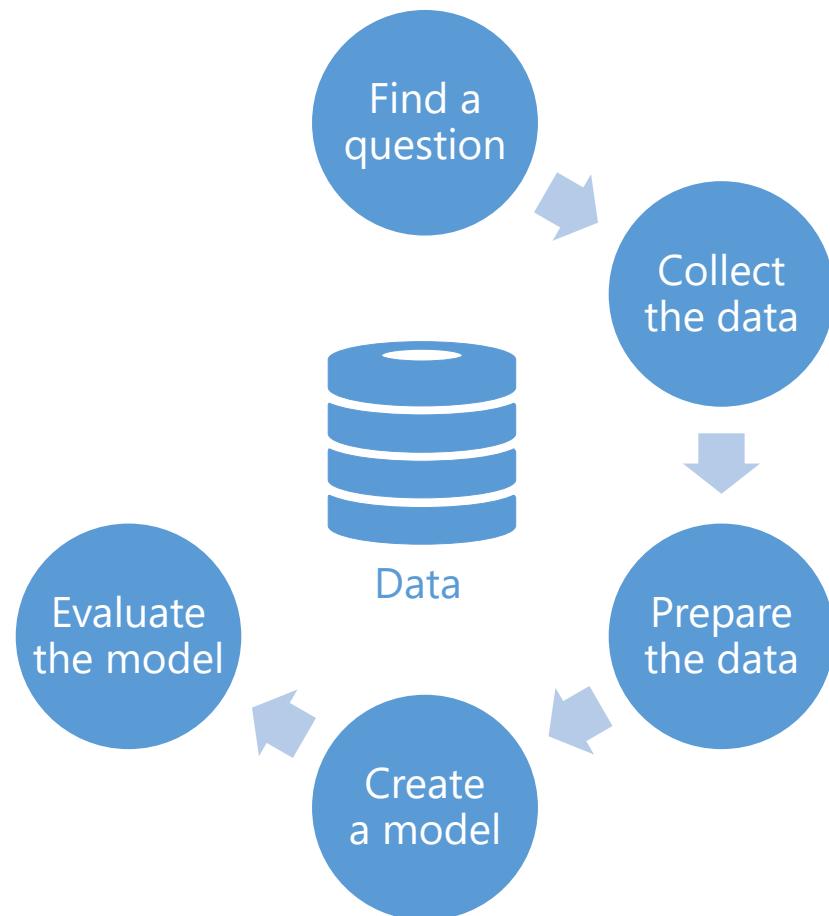
The Data Science Process



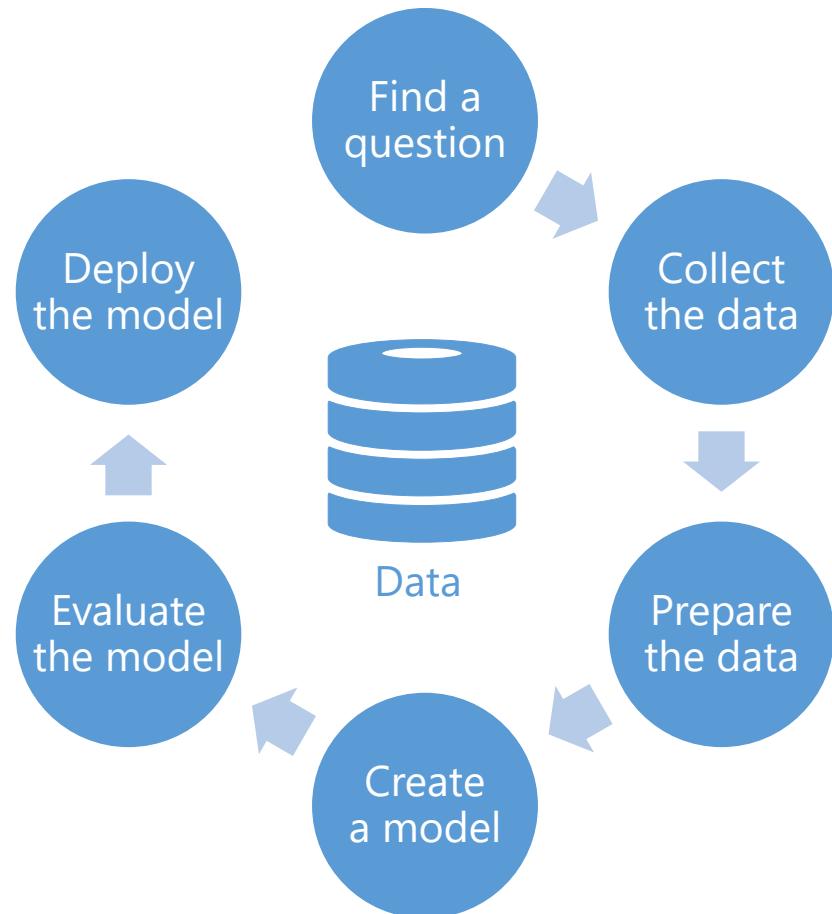
The Data Science Process



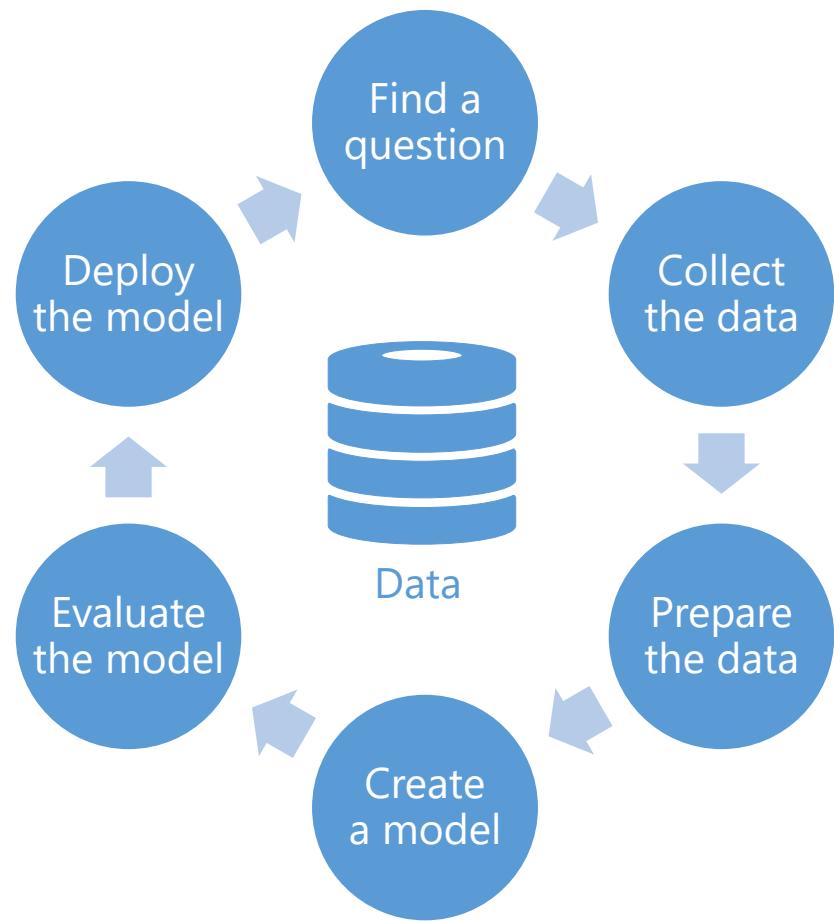
The Data Science Process



The Data Science Process

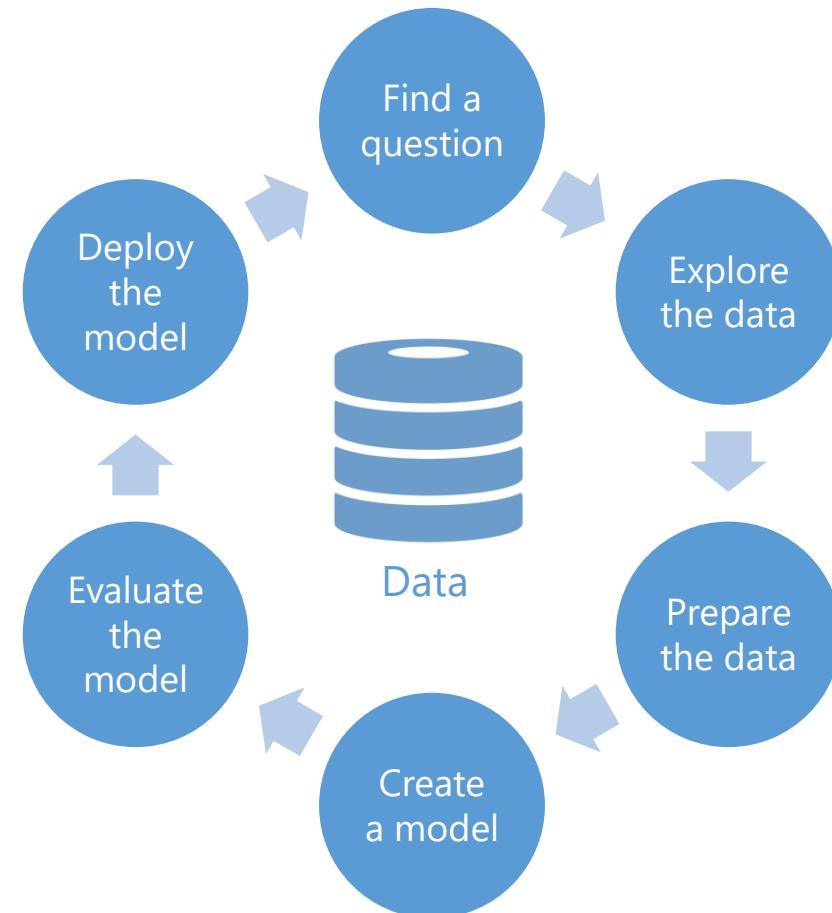


The Data Science Process



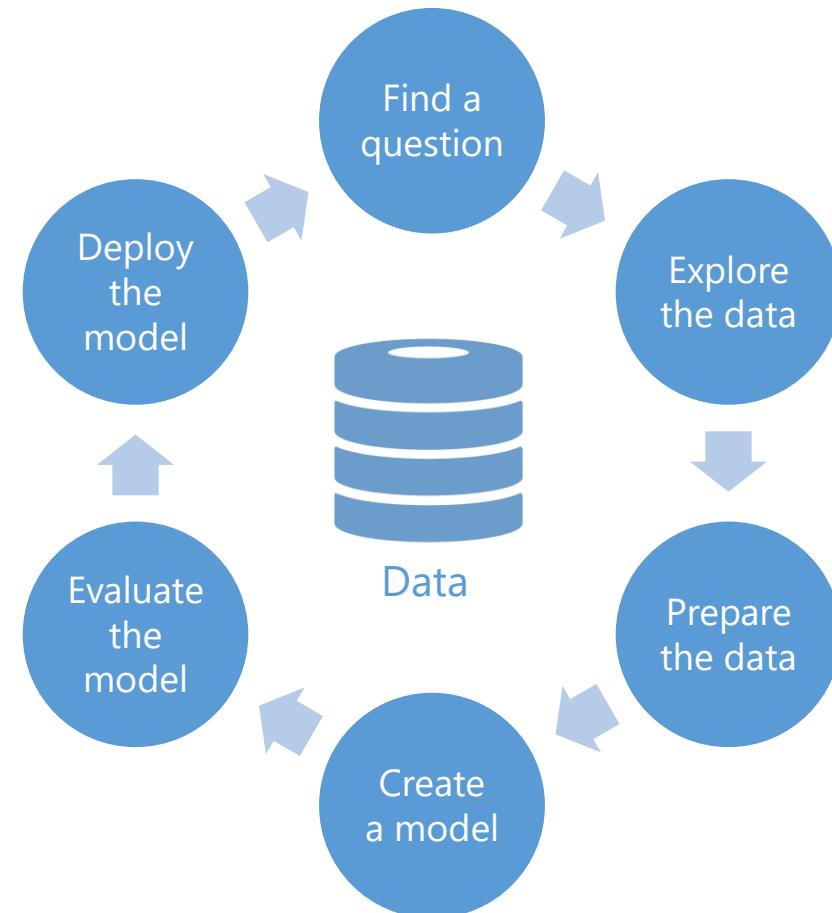
The Data Science Process

Iterative process



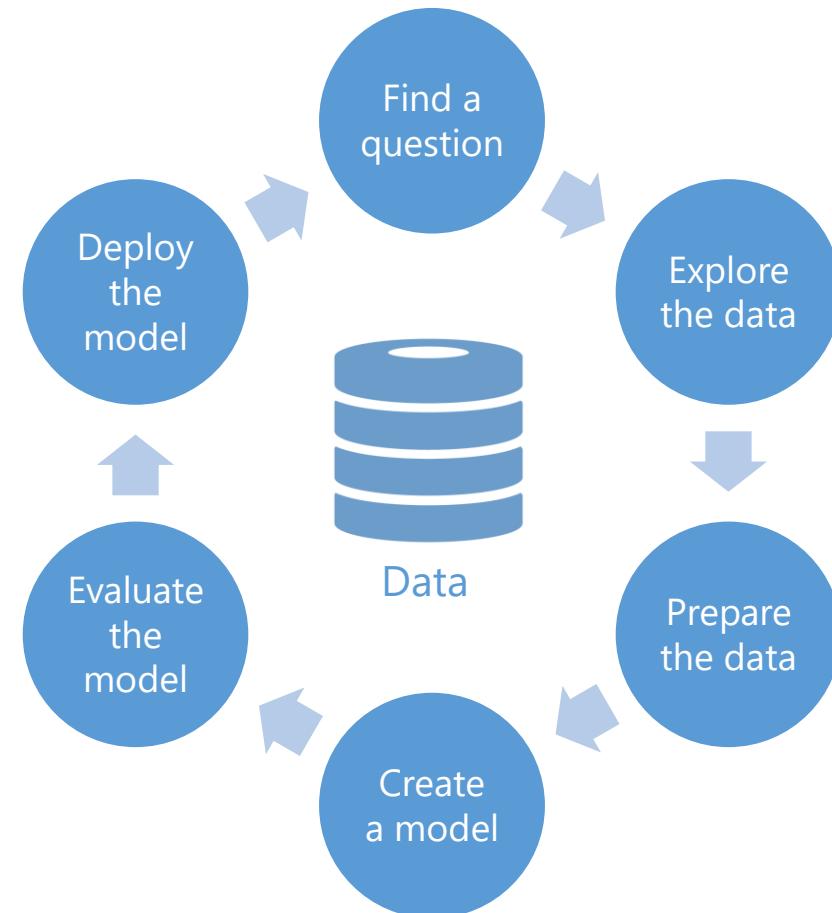
The Data Science Process

Iterative process
Non-sequential



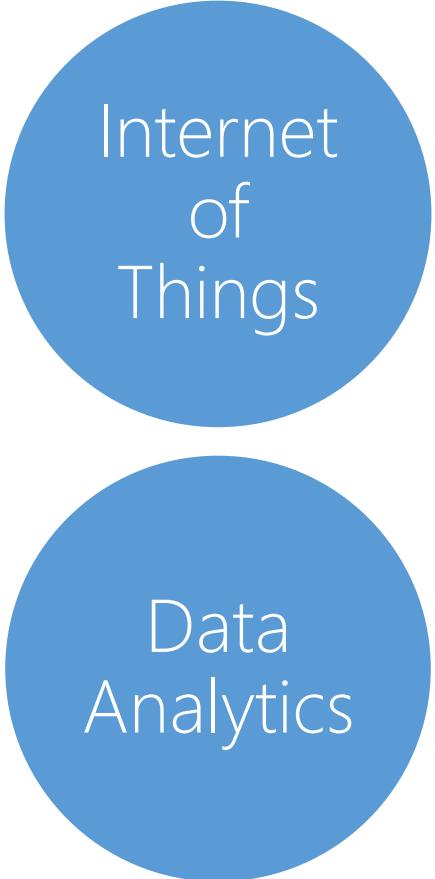
The Data Science Process

Iterative process
Non-sequential
Early termination



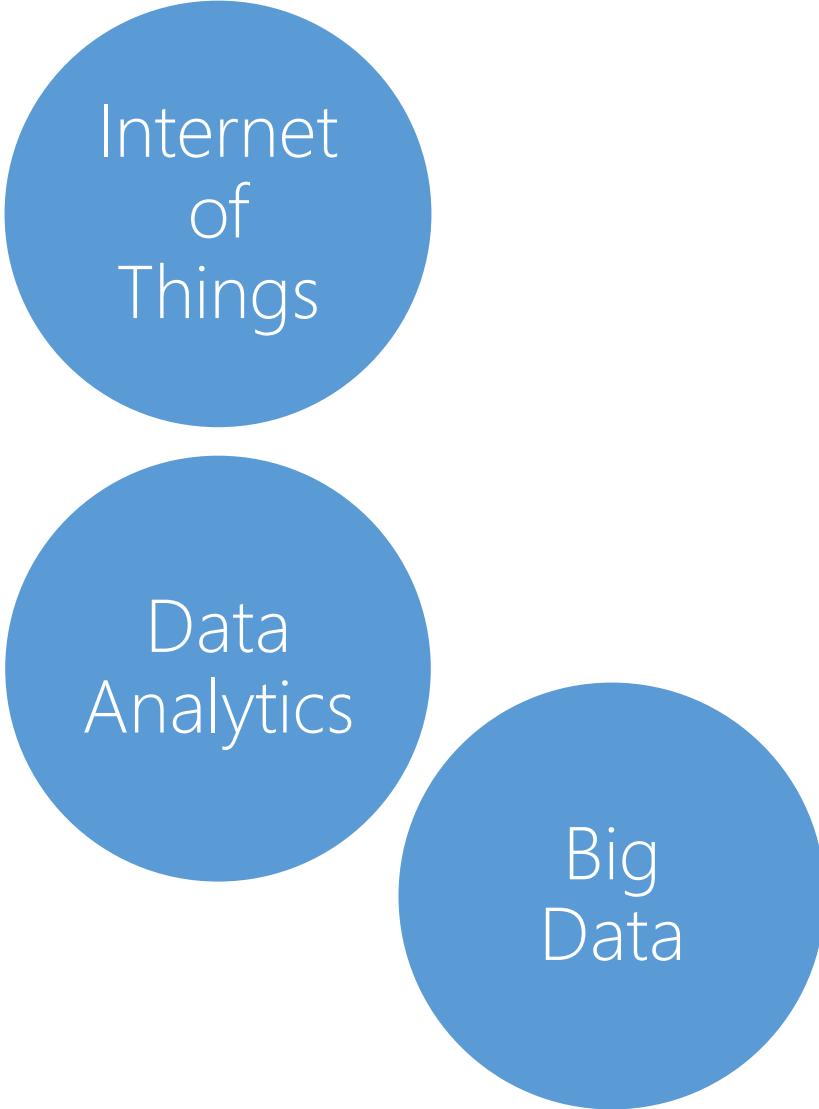
Why is data science important?

Data
Analytics



Internet
of
Things

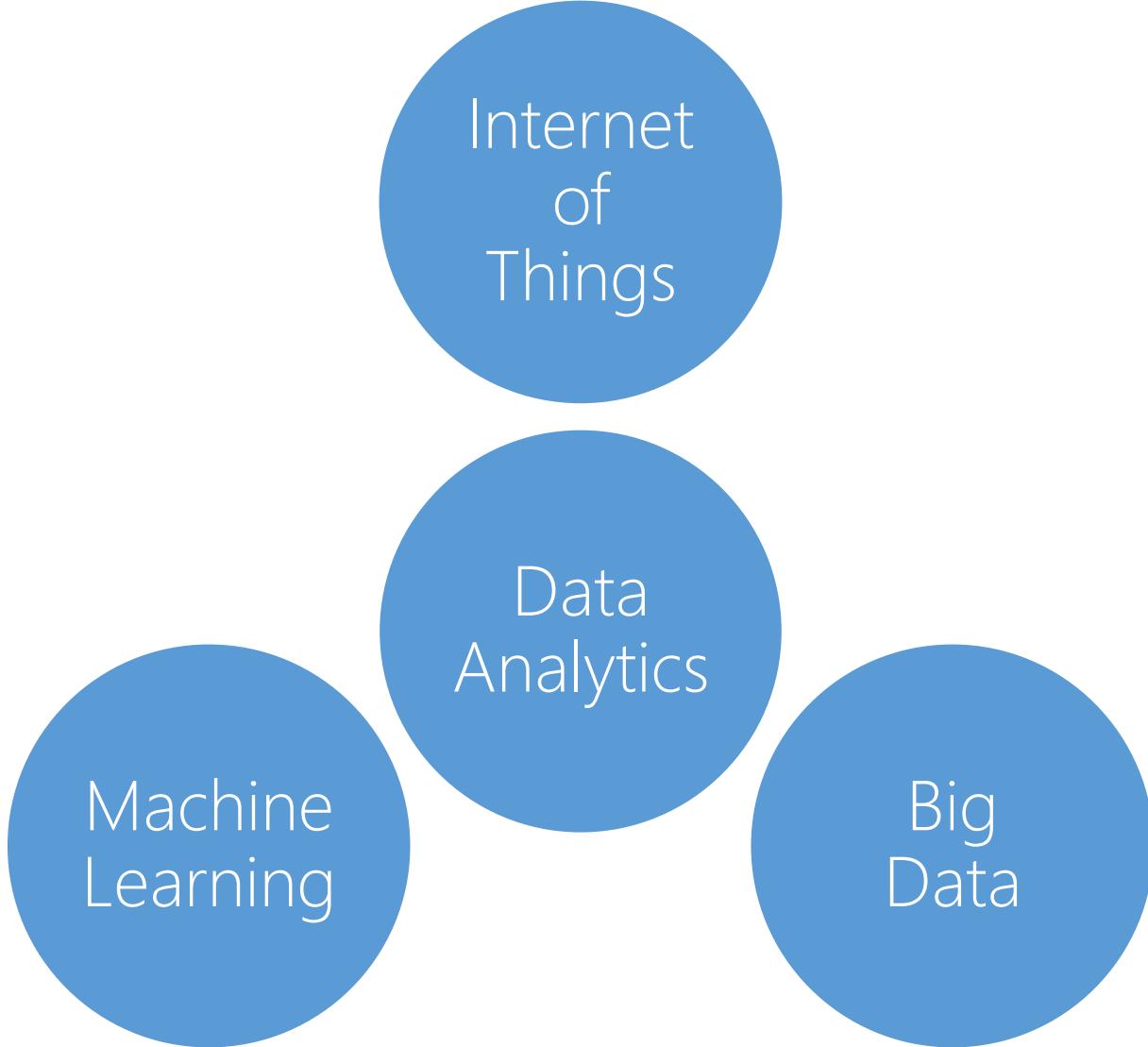
Data
Analytics



Internet
of
Things

Data
Analytics

Big
Data

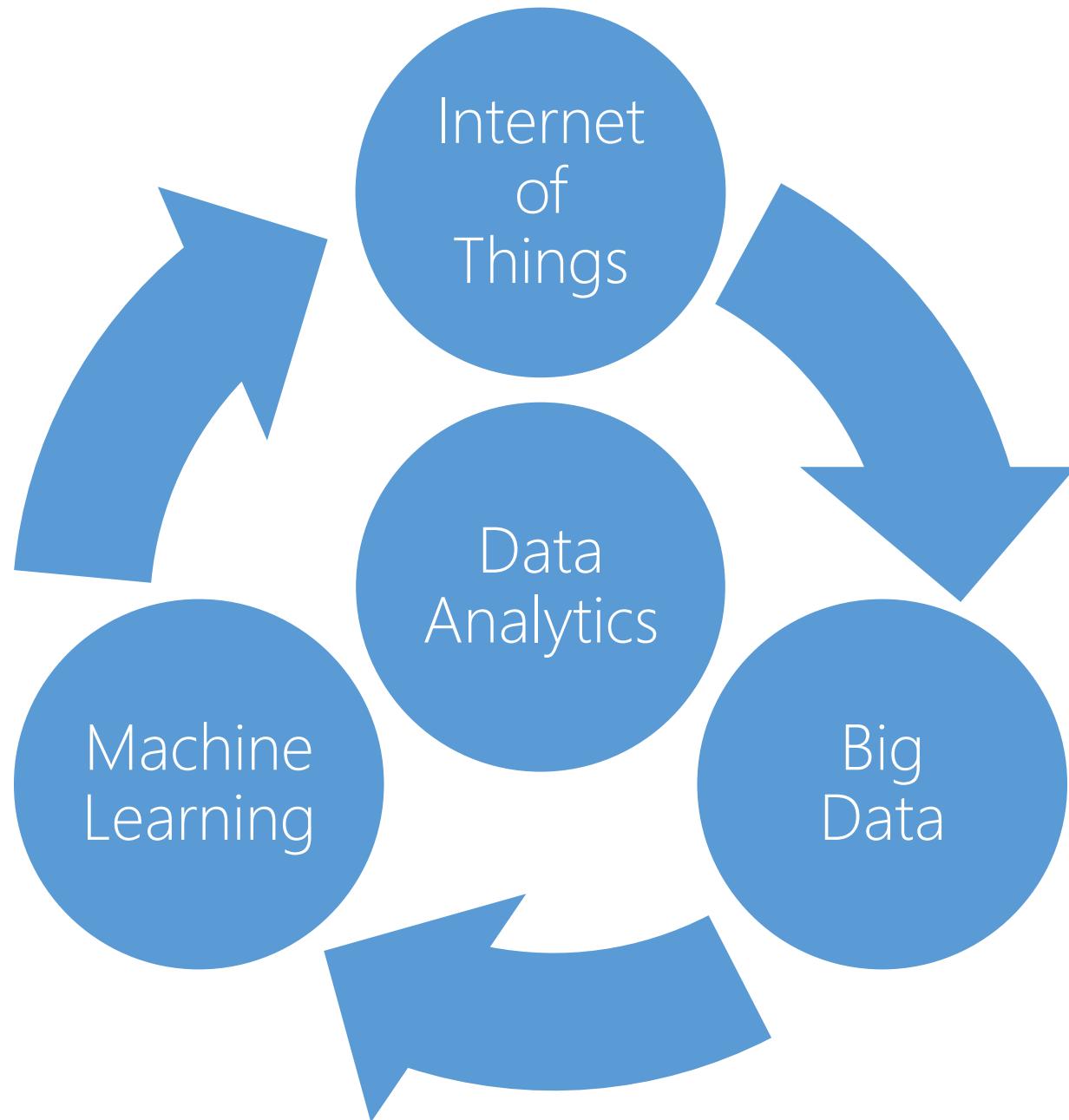


Internet
of
Things

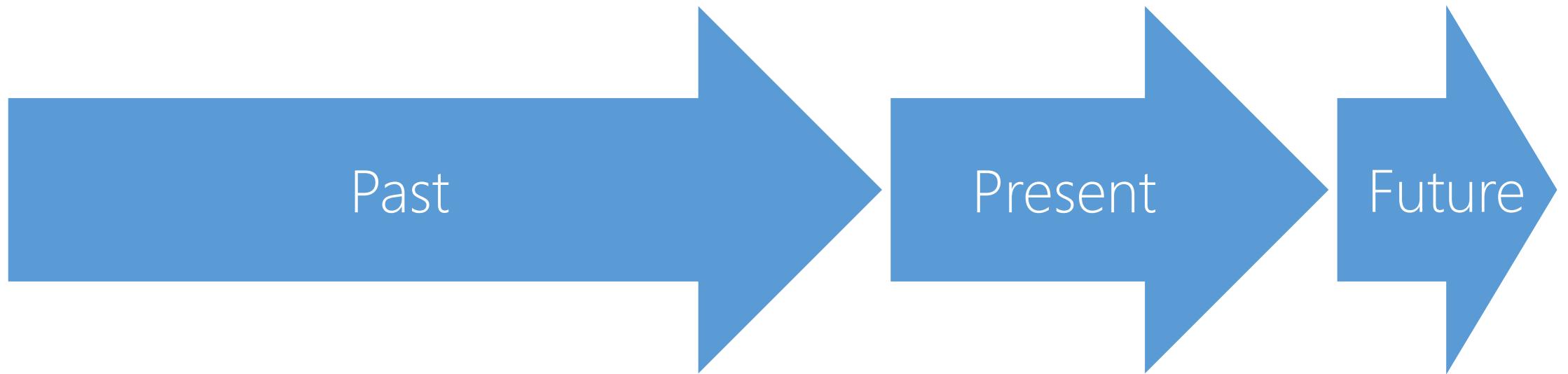
Data
Analytics

Machine
Learning

Big
Data



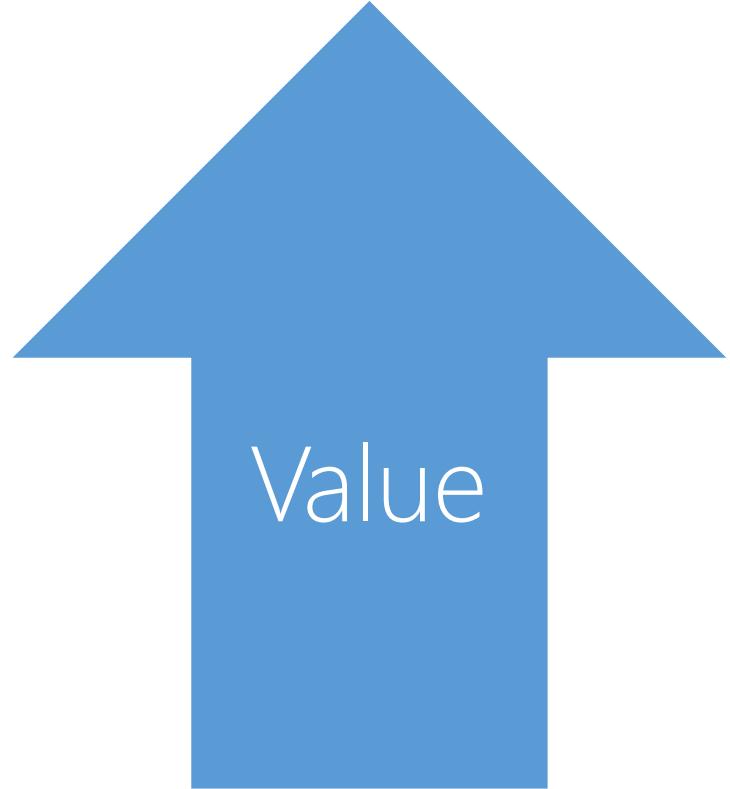
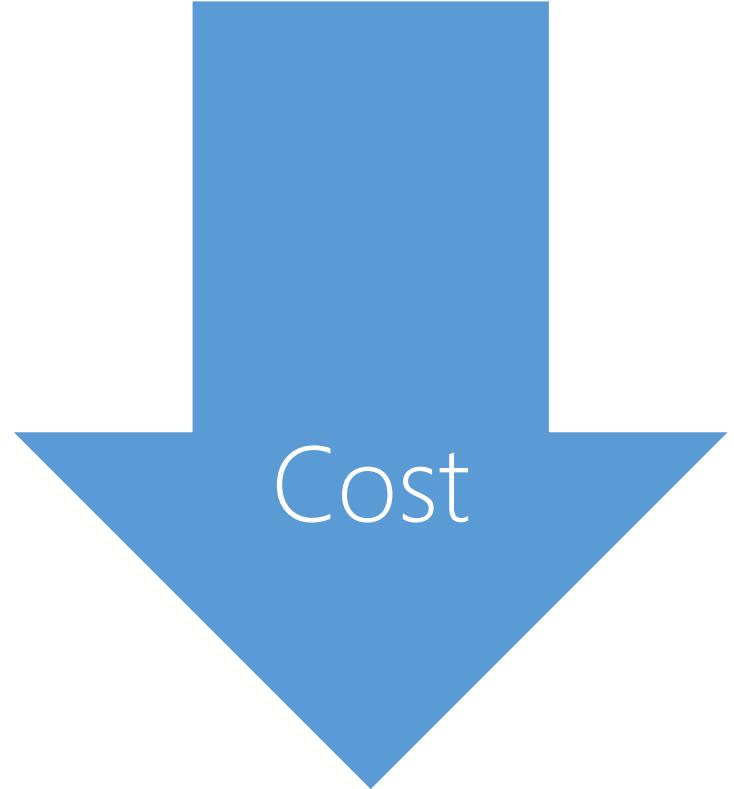
Trends

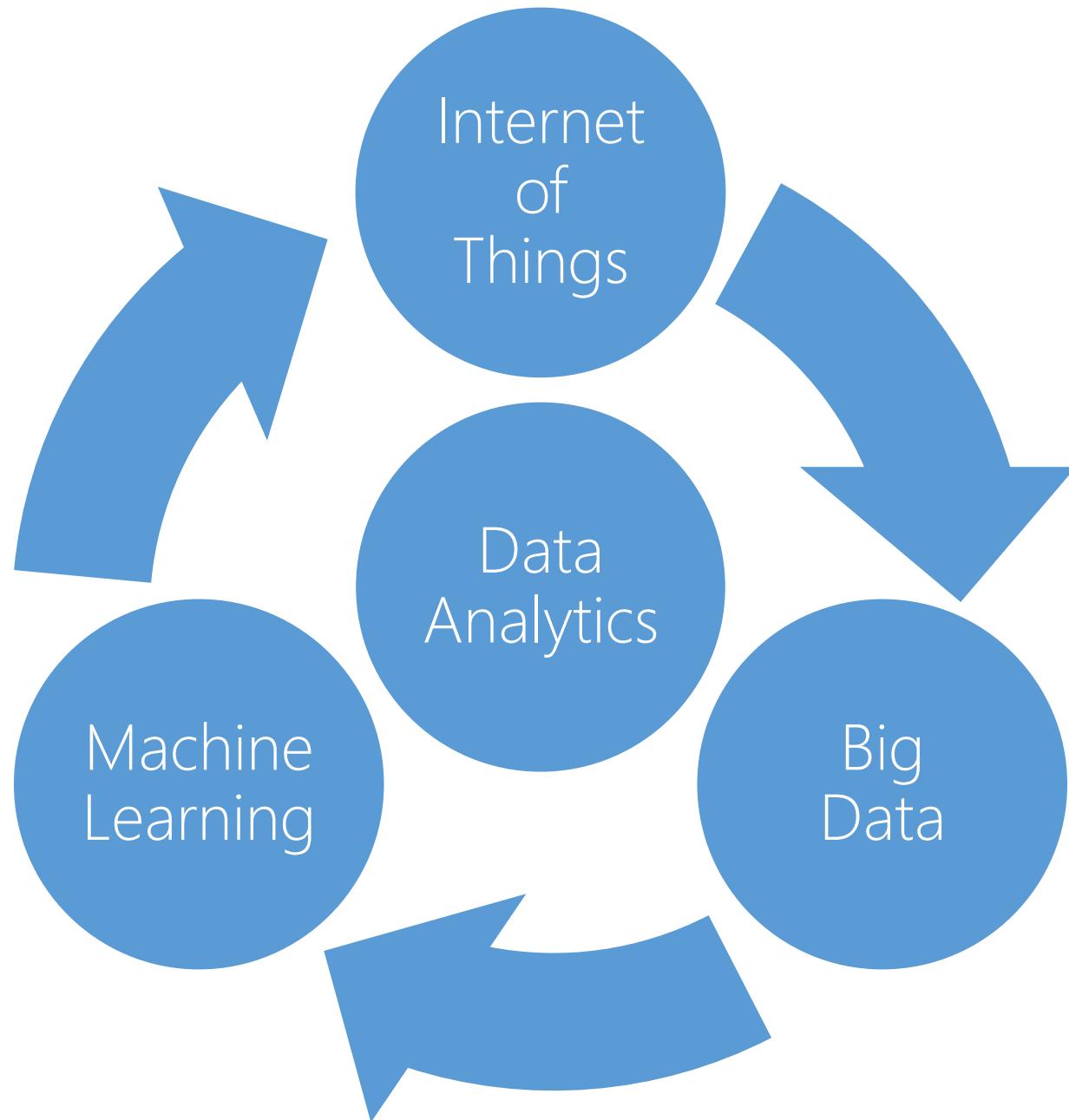


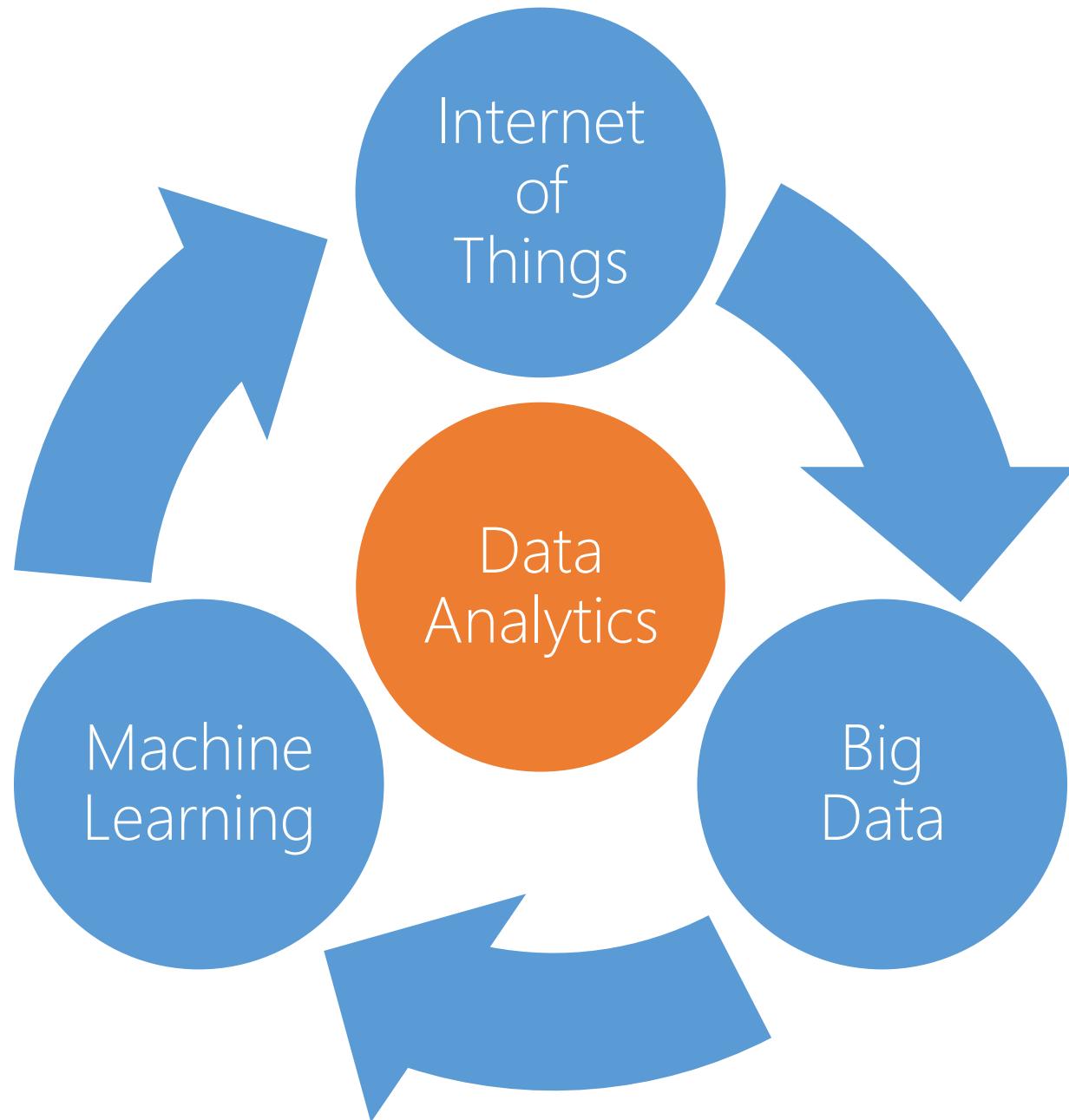


Driven by economics

Possible by technology



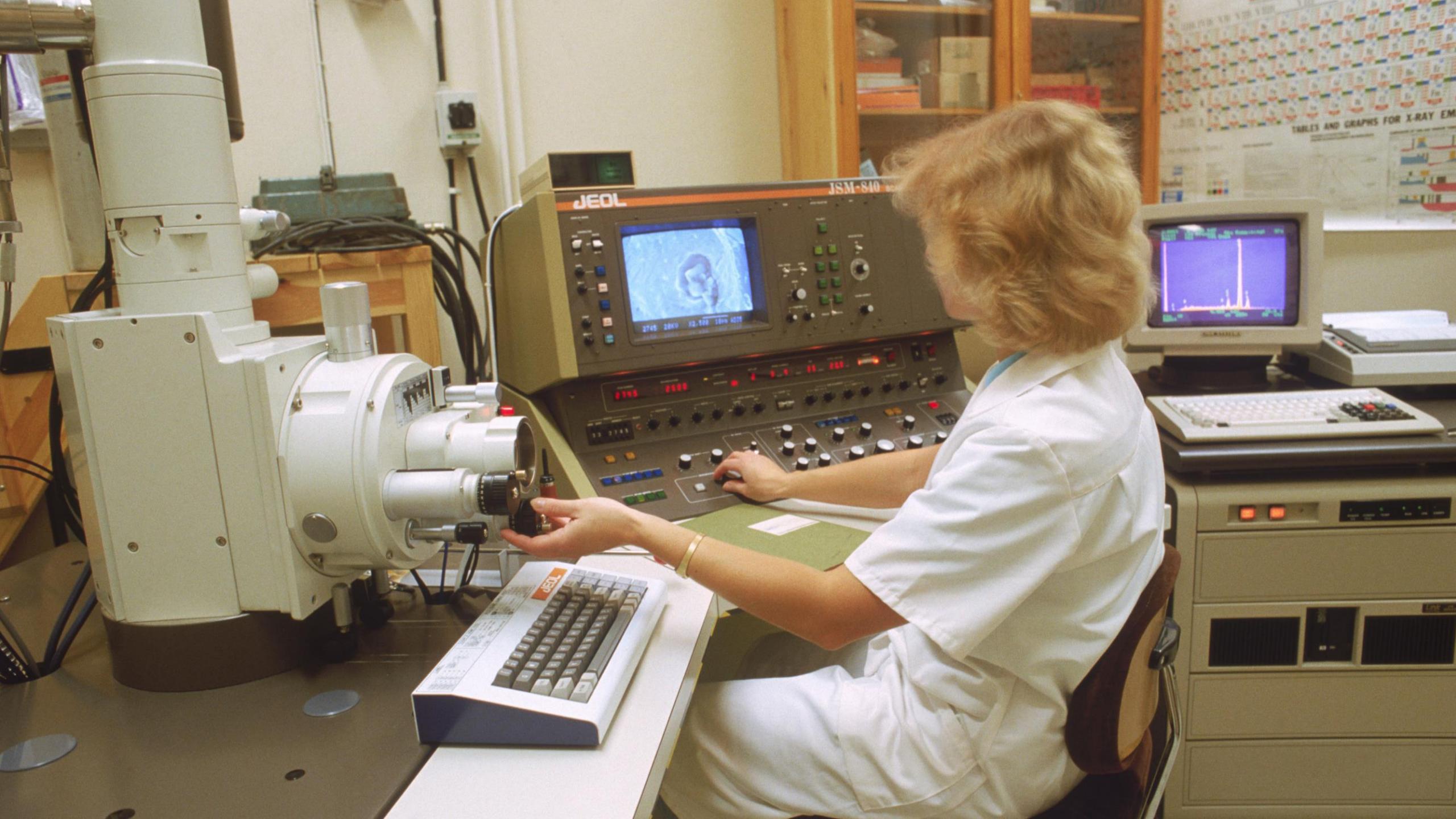






Data Analysis (The Past)





JEOL

JSM-840



2745 2849 32 322 189 422



SUMITOMO

JEOL



My Computer



Network Neighborhood



Recycle Bin

STATS 1.0

File Edit View Tools Run Solutions Window Help

Explorer

Contents of Environment

Libraries File Shortcuts

Editor - Untitled1

```
/** Set CSV file name **/
filename csv "/folders/myfolders/Movies.csv"
termstr = crlf;

/** Import the CSV file.  **/
proc import
  datafile = CSV
  out = movies
  dbms = csv
  replace;
run;

/** Create a scatterplot **/
title "Critic Score and Box Office Revenue";
```

Log - (Untitled)

PLOT 1

HISTOGRAM

FREQUENCY

VARIABLE 1

Bin Range (VARIABLE 1)	Frequency
-2.5 to -2.0	~5
-2.0 to -1.5	~35
-1.5 to -1.0	~65
-1.0 to -0.5	~135
-0.5 to 0.0	~155
0.0 to 0.5	~145
0.5 to 1.0	~145
1.0 to 1.5	~90
1.5 to 2.0	~20

Results Explorer Output - (Untitled) Log - (Untitled) Editor - Untitled1

C:\Documents and Settings Ln 1, Col 1



CHODE

PAGE 174

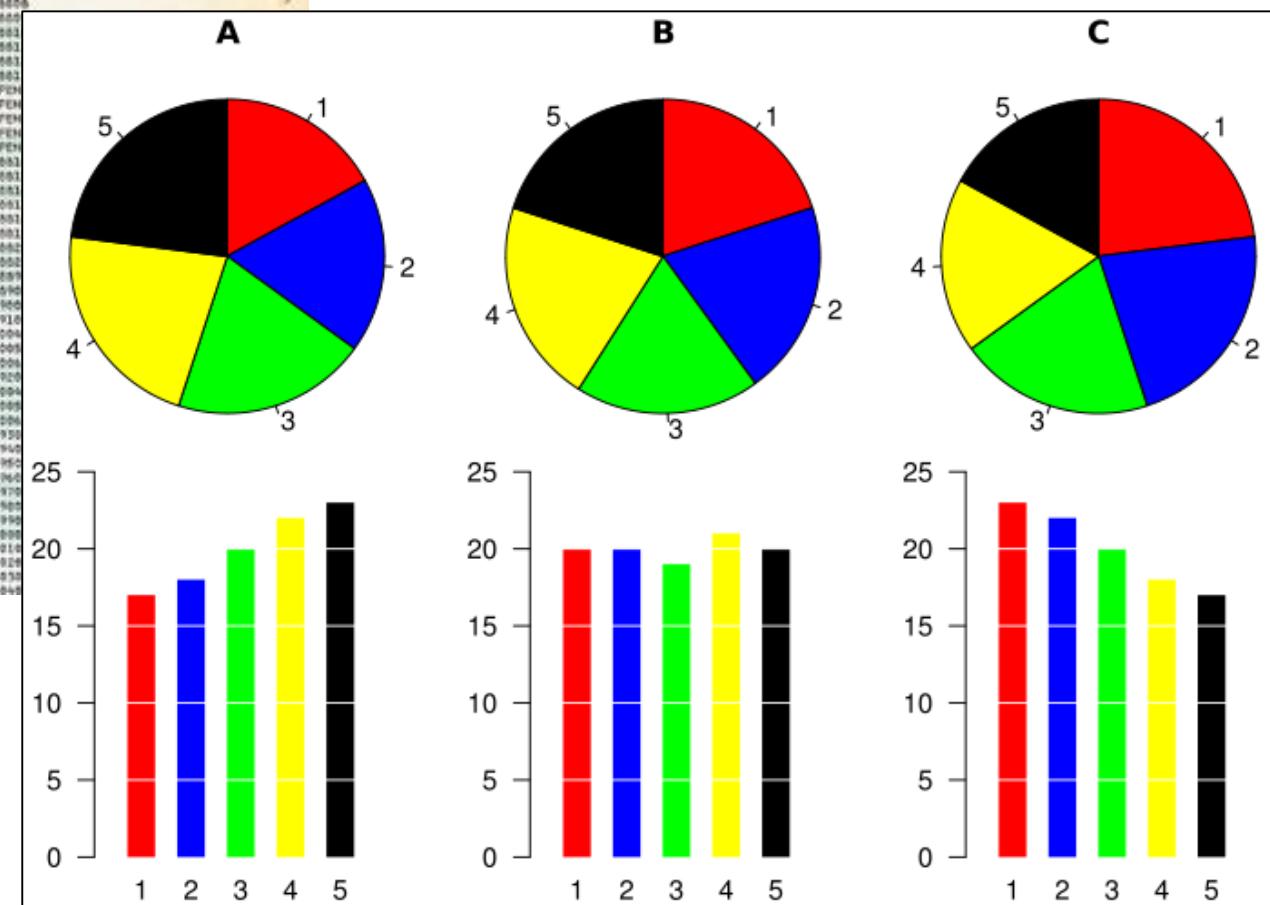
LOC OBJECT CODE ADDR1 ADDR2 STMT SOURCE STATEMENT

ARM R V 08 07.31 02/21/78

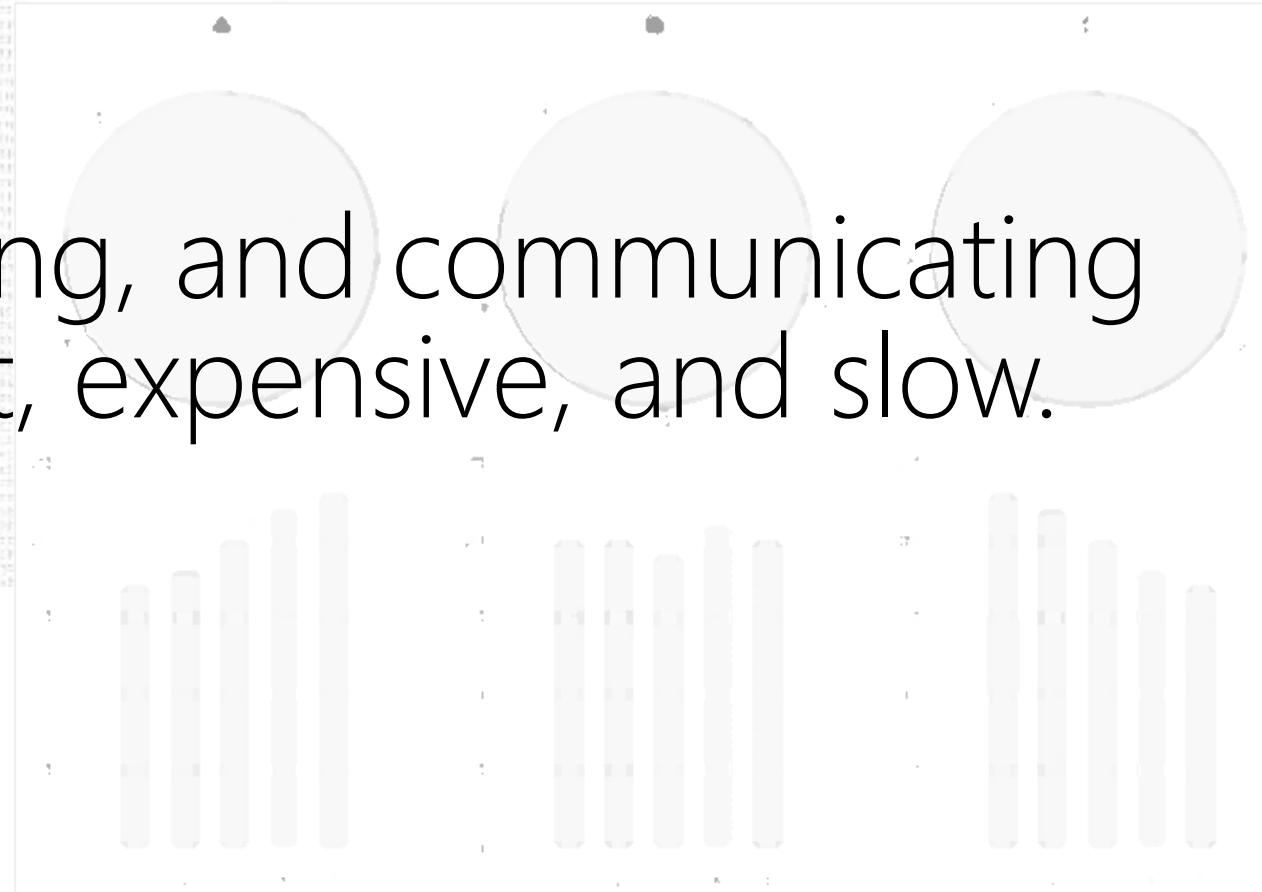
```

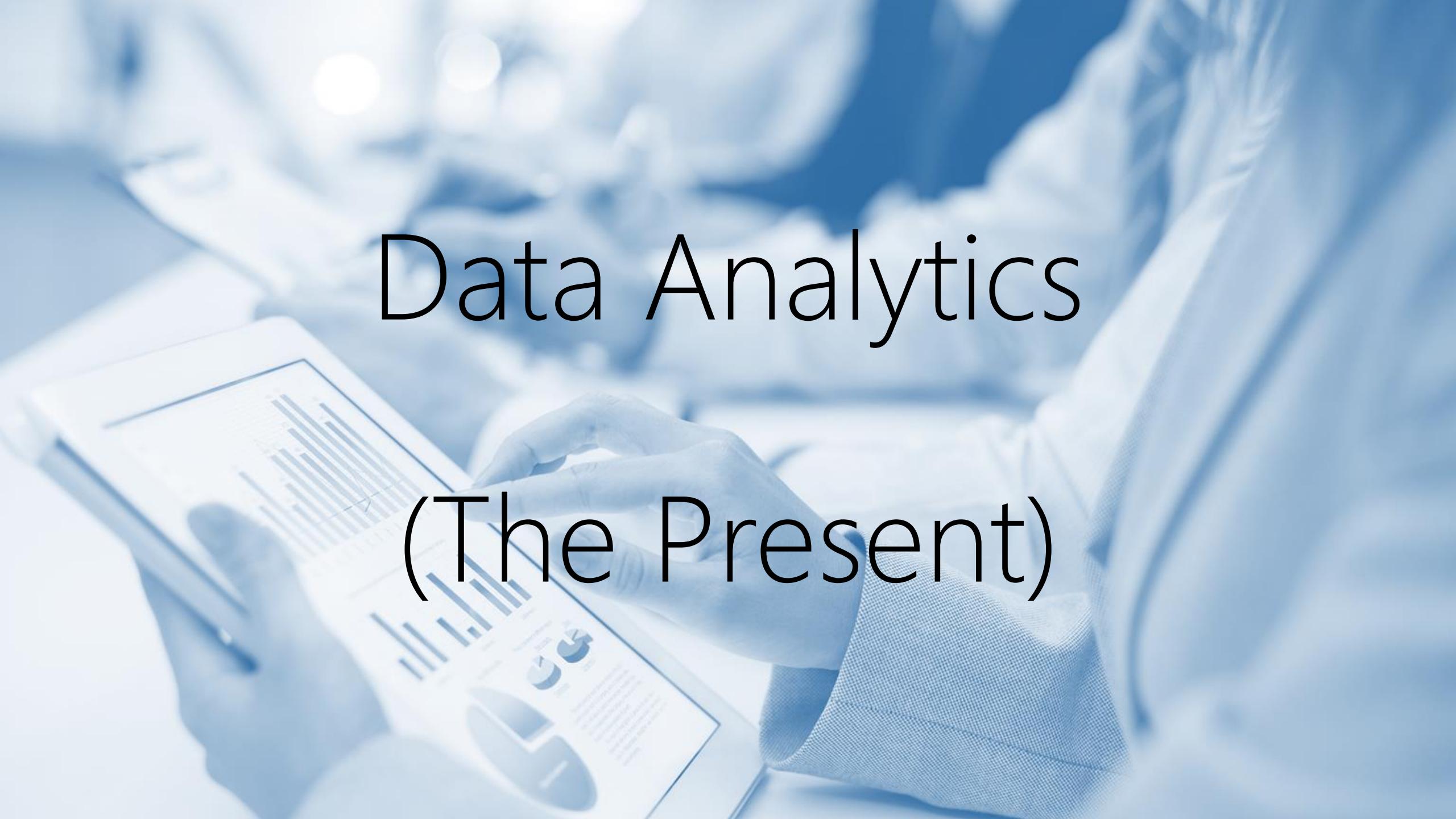
003492 ****
003492 4110 2248 00268 0100 CHODE DS DH
003492 0463 002 002 002,RSPTB ADDRESS OF ADDRESS OF FB 00238300
003492 0463 002 002 SVC 99 CALL DYNAMIC ALLOCATION 00238300
003492 12FF 002 002 LTR CORPCODE,COMPCode SEE IF FAILED 00238300
003492 4770 3009 03000 0106 BNZ AER3 YES ISSUE MESSAGE 00238300
003492 0207 20CC 2240 002AD 0105 MVC JESACB+ACBDRN-SPACB(16),DNAME MOVE DNAME 00238300
003492 0207 20CC 2240 002CC 0106 OPEN (JESACB) OPEN INTERNAL CONSOLE 00238300
003492 0207 20CC 2240 002CC 0106 CNOP 8,4 ALIGN LIST TO FULLWORD 01-0F0H
003492 4519 34AC 03HAC 0106 BAL 3,X#0 LOAD REG1 W/ LIST ADDR. 01-0F0H
003492 00 01094 DC A1211201 OPTION BYTE 01-0F0H
003492 0000A4 01104 DC A131JESACB DCR ADDRESS 01-0F0H
003492 0A13 01114 SVC 1V ISSUE OPEN SVC 01-0F0H
003492 12FF 01124 LTR CORPCODE,COMPCode ISSUE OPEN MSG 00238301
003492 4770 3008 03008 01135 BNZ AER3 ADDRESS OF ACB 00238301
003492 4110 2244 00544 01145 LA 01,JESACB STORE IN RPL 00238301
003492 2979 00570 01155 ST 01,JESRPL+RPLD(AE-IFERPL) 00238301
003492 4110 429C 00219C 01165 LA 01,CPO ADDRESS OF CPO BUFFER 00238301
003492 5010 2978 00576 01175 ST 01,JESRPL+RPLD(AE-IFERPL) 00238301
003492 5010 42EC 0123C 01185 L 01,CHLEN LENGTH OF COMMAND BUFFER 00238301
003492 5010 2958 00500 01195 ST 01,JESRPL+RPLD(AE-IFERPL) 00238301
003492 4110 3108 005405 01205 S 01,CIR 00238301
003492 4110 3582 01002 01811 SYS 01,BSTATS,BSTS 00238301
003492 4770 34E4 03HAC 01212 S 01,ARH 00238301
003492 4110 4874 020F4 01213 C19 PDS PDSPT,,PUT 00238301
003492 4100 3856 00506 01214 LA 01,PROMPT SET ADDRESS 01-0F0H
003492 4100 3856 00506 01215 LA 01,PROMPT SET LENGTH 01-0F0H
003492 4100 3856 0100C 01216 BAL L3H2,PUT SET TO PUT 01-0F0H
003492 4110 429C 0029C 01217 ARHK PDS CMD75,SET 00238301
003492 4110 429C 0029C 01218 ARHK LA 01,CPO SET ADDRESS 01-0F0H
003492 4100 384E 00545 01219 LA 01,78 SET LENGTH 01-0F0H
003492 4285 3846 01046 01219 BAL LDH2,GET 00238301
003492 4110 3782 01082 01311 TH $19STAT2,BSTS IS 'SYS' IN EFFECT? 00238301
003492 4715 5545 00542 01312 ED CLEART YES, PROCESS 00238301
003492 0581 429C 3E4F 0329C 01313 CLC CMD131,=C'00' 00238301
003492 4758 551A 00514 01314 BE EXIT 00238301
003492 0513 429C 3E2B 0329C 01315 CLC CMD291,HCL2B' 00238301
003492 4758 5405 00504 01316 BE CIB 00238301
003492 0593 429C 5E3C 0329C 01317 CLC CMD141,=C'SYS' 00238301
003492 4798 5405 00402 01318 BE BT3 00238301
003516 0591 429C 3E44 0329C 01319 CLC CMD121,=C'BT' 00238301
003492 4798 558A 00584 01408 BE INTCON 00238301
003529 0591 429C 3E46 03246 01411 CLC CMD121,=C'BT' 00238301
003492 4798 558A 00584 01412 BE INTCON 00238301

```



Collecting, analyzing, and communicating data was difficult, expensive, and slow.



A close-up photograph of a person's hands and face. The person is wearing a light blue surgical mask and a pair of blue nitrile gloves. They are holding a white piece of paper with several blue bar charts printed on it. The background is blurred, showing more of the same blue medical equipment.

Data Analytics

(The Present)





File Home Insert Page Layout Formulas Data Review View Add-ins Team Tell me what you want to do

Paste Font Calibri 11 A A Wrap Text General \$ % , Conditional Formatting Merge & Center Number Format as Table Insert AutoSum

Font Font Style Alignment Text Direction Orientation Merge & Center Format Cells Format Table Format Styles Format Cells Format Table Format Styles Insert Delete Format Cells

Clipboard Font Font Style Text Direction Orientation Merge & Center Format Cells Format Table Format Styles Insert Delete Format Cells

Font Font Style Text Direction Orientation Merge & Center Format Cells Format Table Format Styles Insert Delete Format Cells

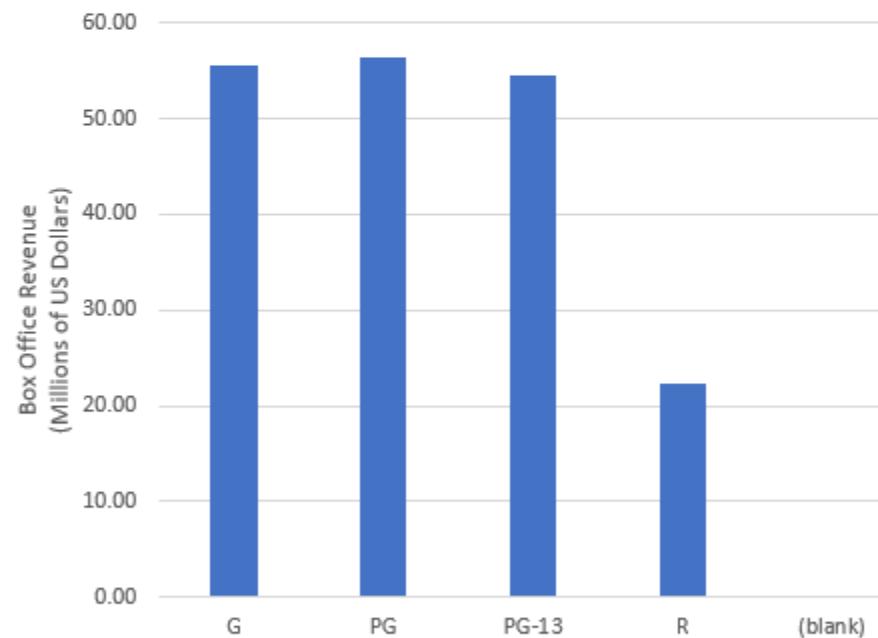
Font Font Style Text Direction Orientation Merge & Center Format Cells Format Table Format Styles Insert Delete Format Cells

T25

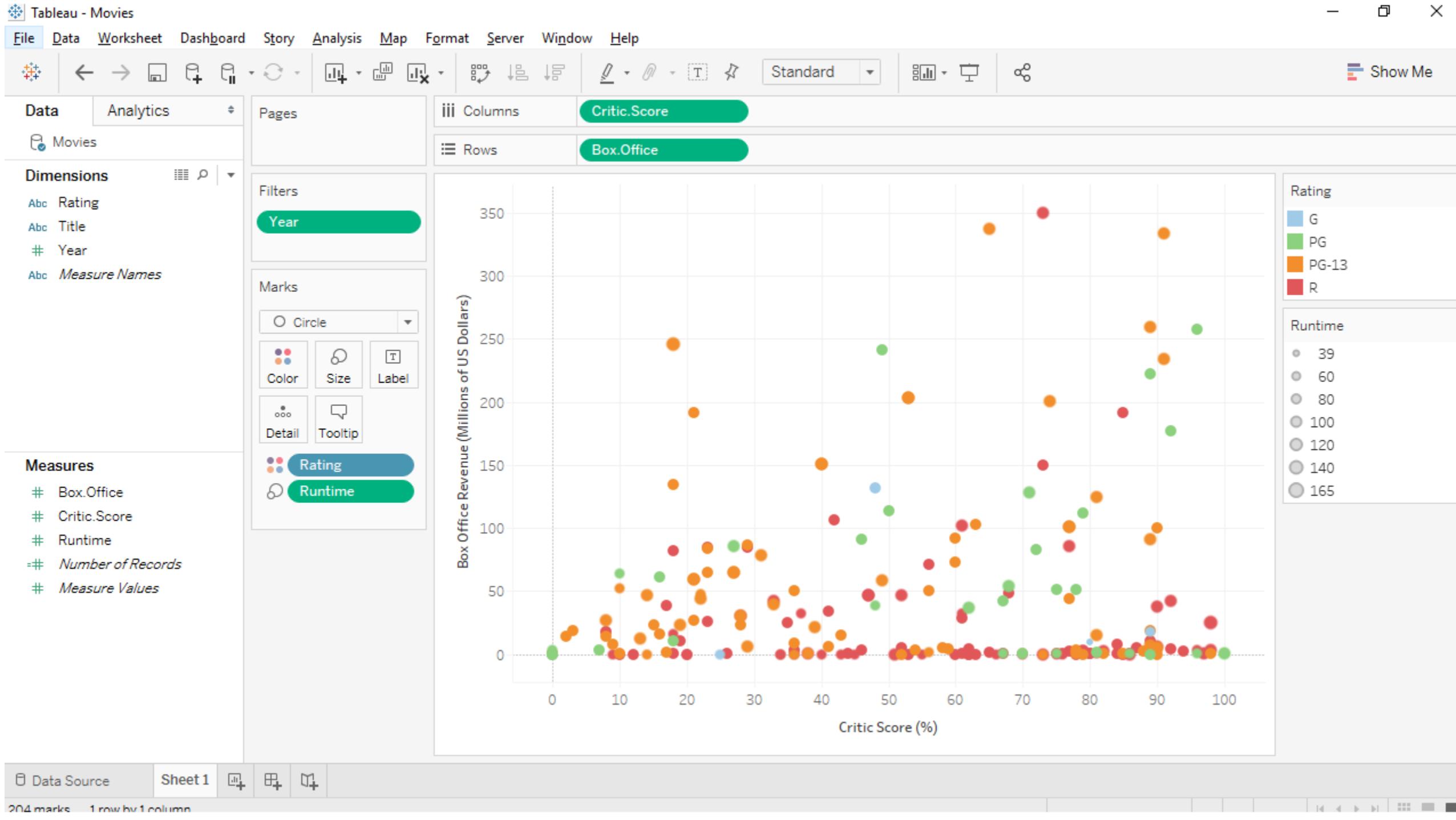
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	Title	Year	Rating	Runtime	Critic Scor	Box Office		Row Labels	Average of Box Office									
2	The Whole	2000	R		98	45	57.3	G		55.48								
3	Cirque du	2000	G		39	45	13.4	PG		56.40								
4	Gladiator	2000	R		155	76	187.3	PG-13		54.56								
5	Dinosaur	2000	PG		82	65	135.6	R		22.26								
6	Big Momm	2000	PG-13		99	30	0.5	(blank)										
7	Gone in Si	2000	PG-13		118	24	101	Grand Total	40.68									
8	Shaft	2000	R		99	68	70.3											
9	Chicken R	2000	G		84	97	106.6											
10	Me, Myse	2000	R		116	48	0.6											
11	The Patric	2000	R		165	62	2.3											
12	The Adver	2000	PG		92	43	0.4											
13	The Kid	2000	PG		104	49	68.5											
14	X-Men	2000	PG-13		104	82	156.2											
15	What Lies	2000	PG-13		130	46	155											
16	Thomas ar	2000	G		85	19	2											
17	Nutty Prof	2000	PG-13		106	26	122.4											
18	Coyote Ug	2000	PG-13		100	22	59.8											
19	Hollow Ma	2000	R		112	27	72.1											
20	Space Cov	2000	PG-13		130	78	89.4											
21	The Repla	2000	PG-13		118	40	44											
22	Bless the	2000	R		107	3	28.1											

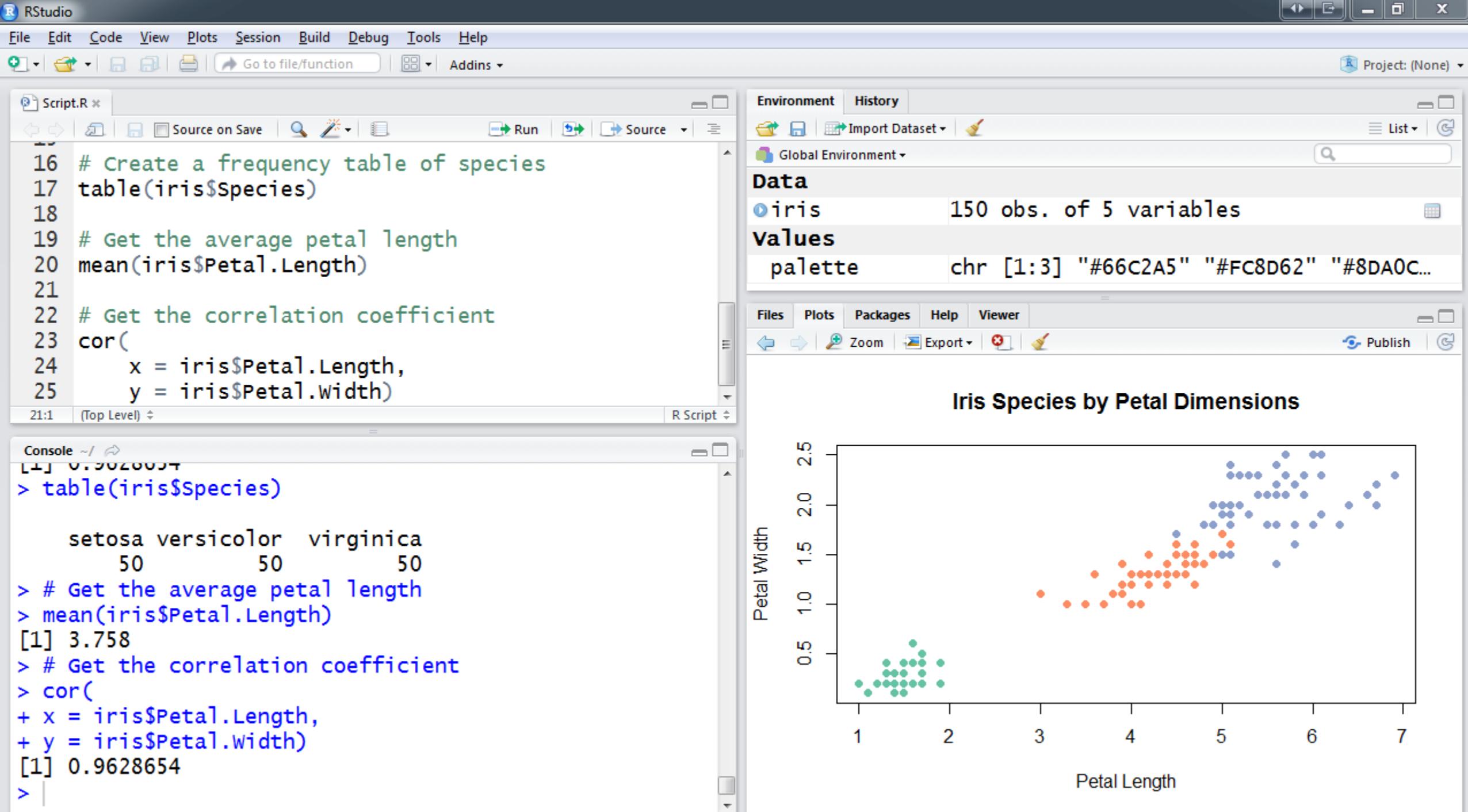
Average of Box Office

Average Box Office Revenue by Rating Category



Rating







SEARCH

MENU

- Suspendisse
- Velit Sit Amet
- Pellentesque
- Voluptat
- Pretium Nulla
- Eget Turpis

PROJECTS

- Nunc venenatis
- Maecenas
- Curabitur

Dashboard > Ornare Orci Velit

Consectetur Adipiscing

Donec sollicitudin, ipsum non

\$1,310

Nunc et metus magna



Aenean Porta

Donec sollicitudin, ipsum non

\$24,815

Nunc et metus magna



Aenean Porta

Donec sollicitudin, ipsum non

94.2%

Nunc et metus magna



Aenean Porta

Donec sollicitudin, ipsum non

Nunc et metus magna

Integer

Nam ultrices

Phasellus vulputate

Vivamus

< NULLAM SAPIEN >

Mon	Tue	Wed	Thu	Fri	Sat	Sun
30	31	1	2	3	4	5
6	7	8	9	10	11	12
13	14	15	16	17	18	19
20	21	22	23	24	25	26
27	28	29	30	31	1	2

Nullam Ac Leo Vitae

5000

4000

3000

2000

1000

JAN

FEB

1,283

Cursus Id Arcu Eget

3,155

Cursus Id Arcu Eget

MAR

APRIL

MAY

JUN

JULY

Lorem Ipsum

Nunc faucibus dignissim nibh

Cras Pharetra



Aliquam

Nunc faucibus dignissim nibh

Donec orci nulla

872

Lobortis non nisi quis

650

Phasellus Interdum

Dignissim nibh

Bibendum

Nunc Leo Tortor

Nunc faucibus dignissim nibh

Felis ligula rutrum orci 2016

Phasellus interdum ligula eget turpis

Sed id egestas 2016

47

Aspernatur aut odit aut,
sed quia consequuntur
magni dolores eos,
voluptatem sequi
nesciunt

125 000

Sed quia non numquam
eius modi tempora
incident, ut labore
et dolore

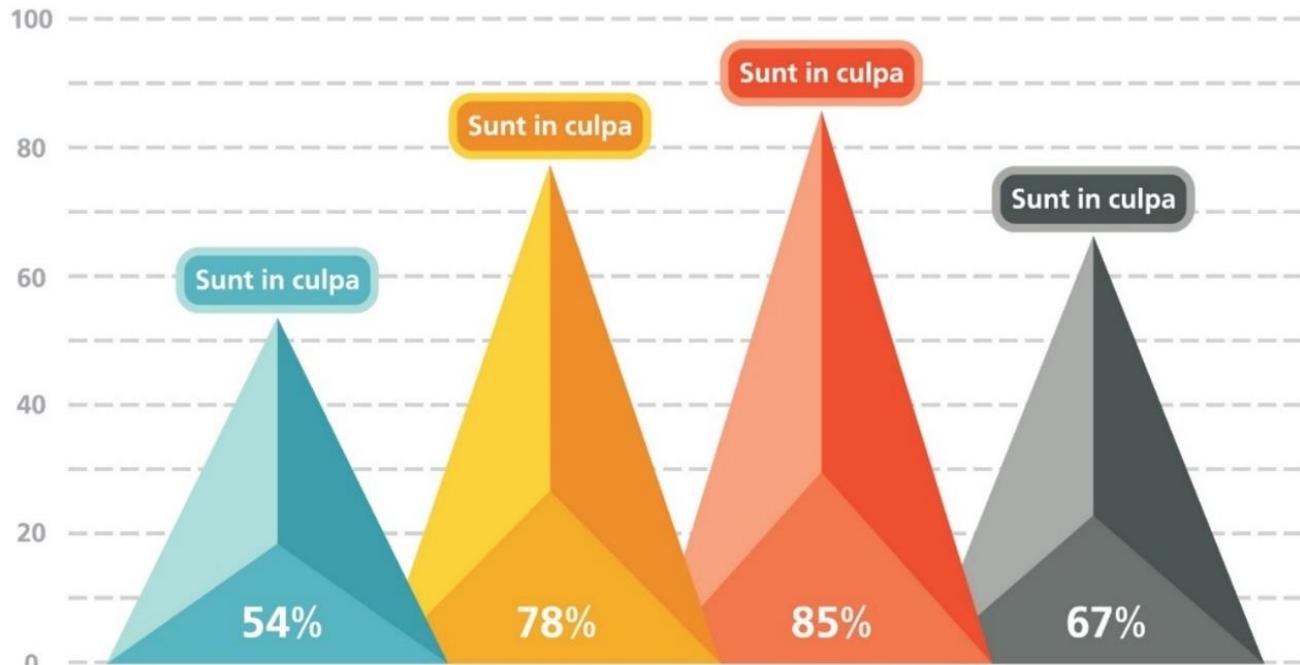
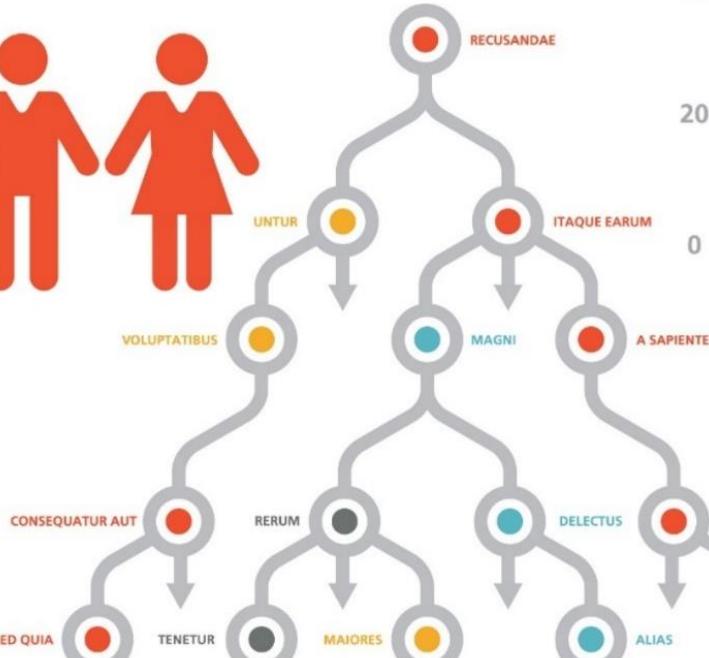
REPUDIANDAE NON RECUSANDAE QUIA VOLUNTAS SIT

Enim ipsam voluptatem,
quia voluptas sit, asperna
odit aut fugit, sed quia
consequuntur magni
dolores eos



ET HARUM QUIDEM

Rerum facilis est et expedita
distinctio. Nam libero tempore,
cum soluta nobis est eligendi optio.



156 000
Ut enim ad minima
veniam, quis nostrum
exercitationem ullam
corporis suscipit
laboriosam, ?

89 000
Nisi ut aliquid ex ea
commodi consequatur





Collecting, analyzing, and communicating
data is easy, inexpensive, and fast.

Data-Driven Decision Making

(The Future)







The home of the U.S. Government's open data

Here you will find data, tools, and resources to conduct research, develop web and mobile applications, design data visualizations, and [more](#).

GET STARTED

SEARCH OVER 196,465 DATASETS

Federal Student Loan Program Data



BROWSE TOPICS



Agriculture



Climate



Consumer



Ecosystems



Education



Energy



Finance



Health



Local
Government



Manufacturing



Maritime



Ocean



Public Safety

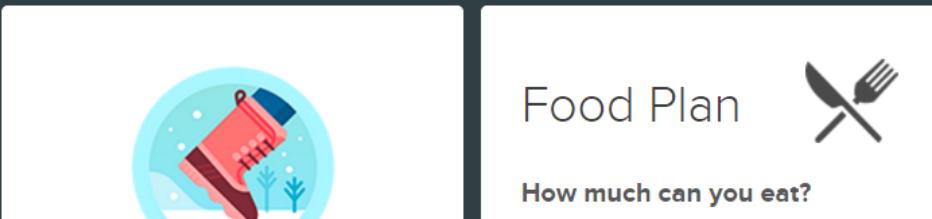
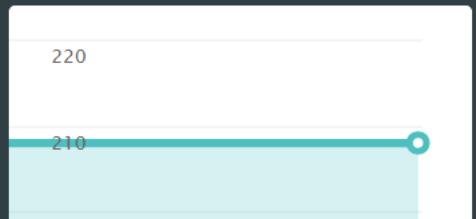
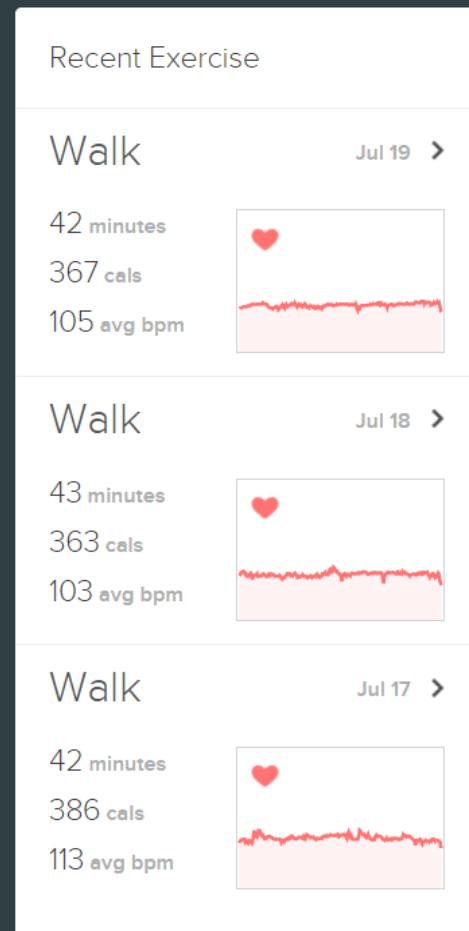
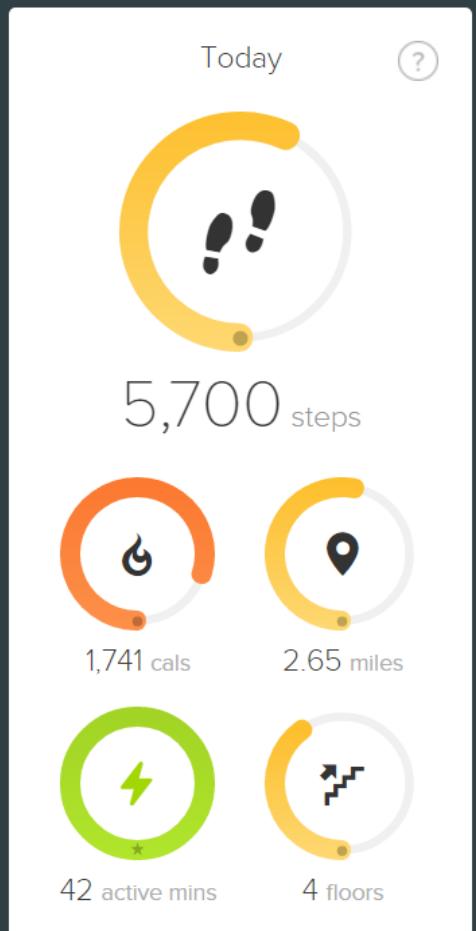


Science &
Research



Today JULY 19

Last Sync: 8 mins ago



Retail Sales

Ask a question about the data on this dashboard

Total Products

104

Total Sales

\$22.05M

Sales by Product Type

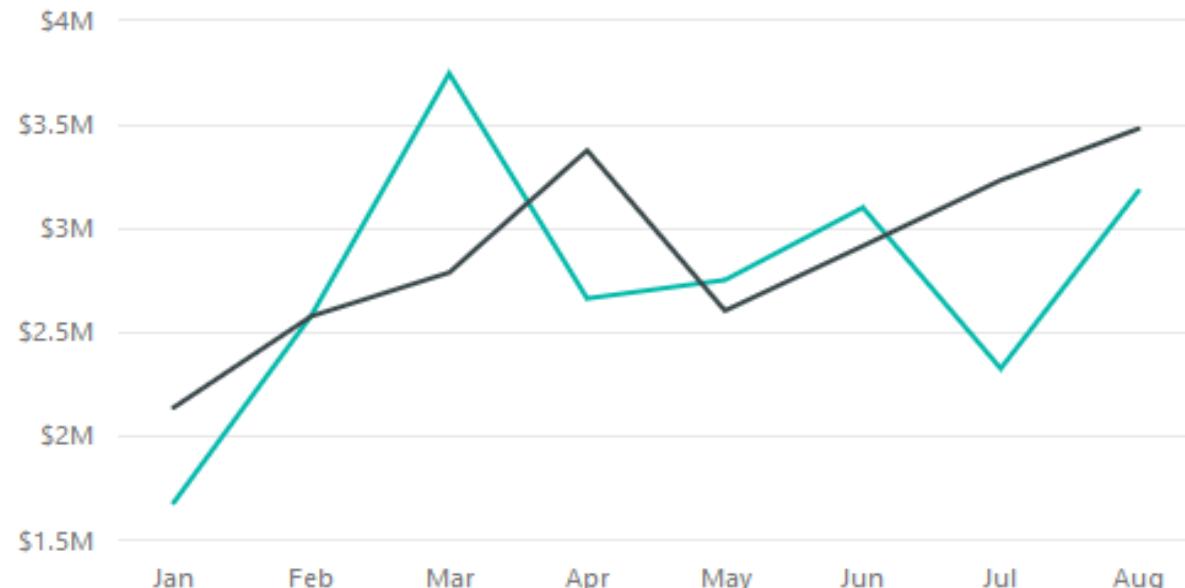


New Products This Year

10

Annual Sales Comparison by Month

This Year Sales Last Year Sales



Sales per Square Foot by Total Sales Variance and District

District FD - 01 FD - 02 FD - 03 FD - 04 LI - 01 LI - 02 LI - 03



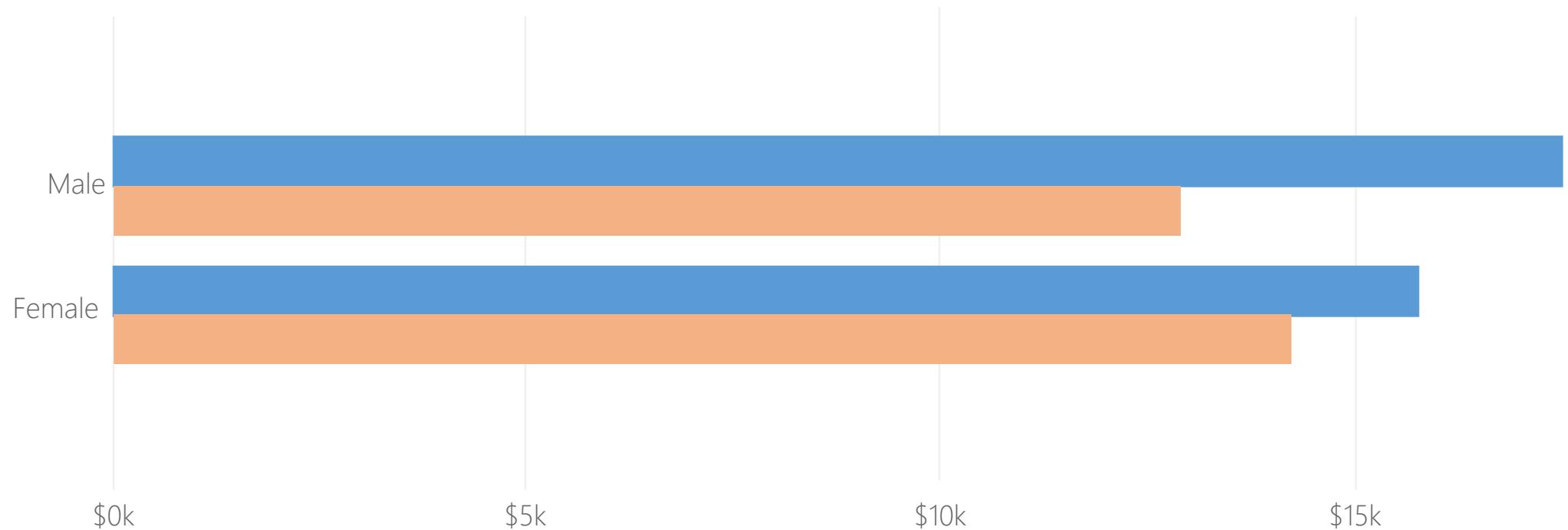
Internet Sales

Show me sales by gender and marital status.

Displaying sum of sales by gender and marital status

Marital Status:

- Married
- Single

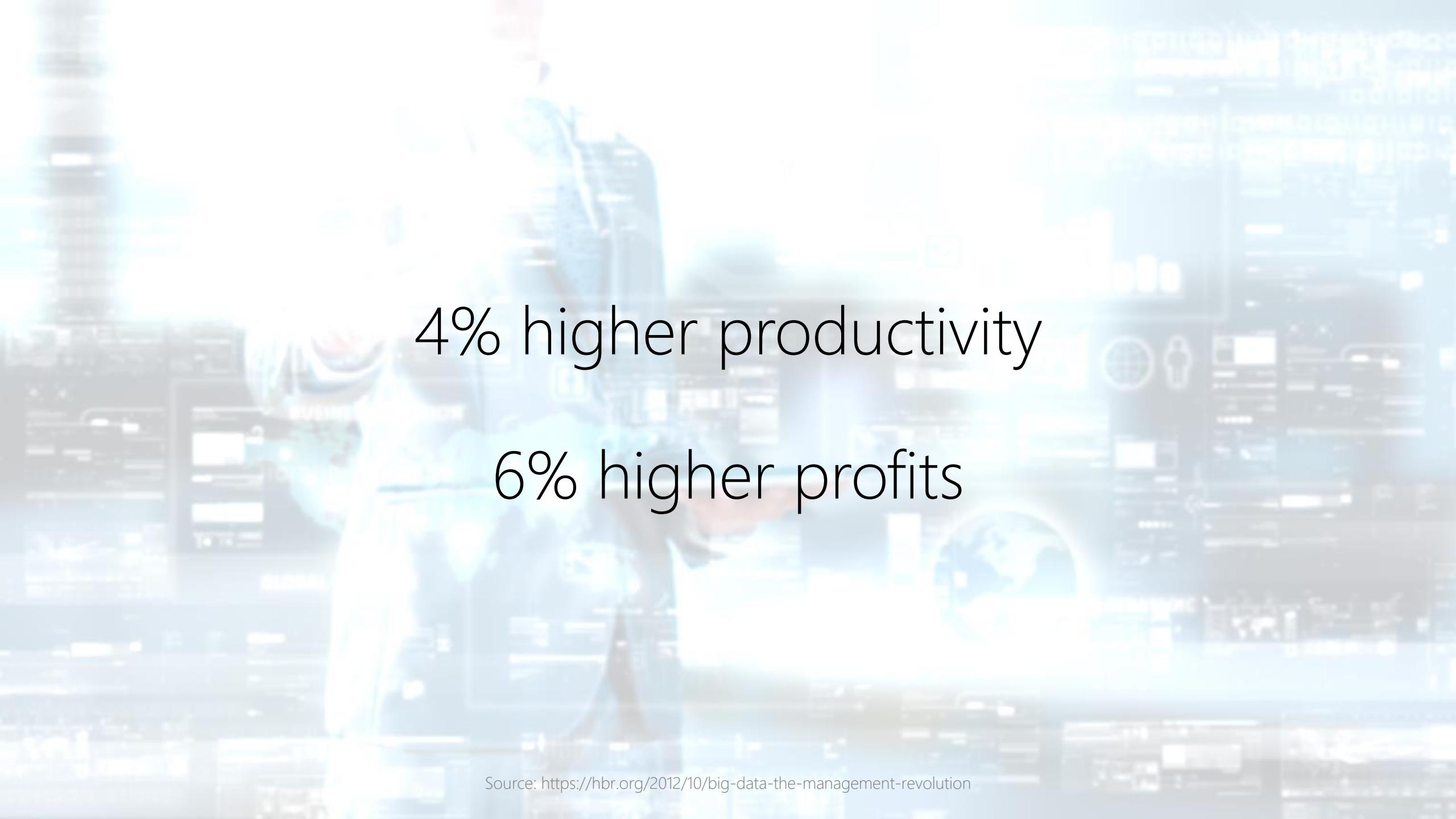


"Show me sales by
gender and marital
status."









4% higher productivity

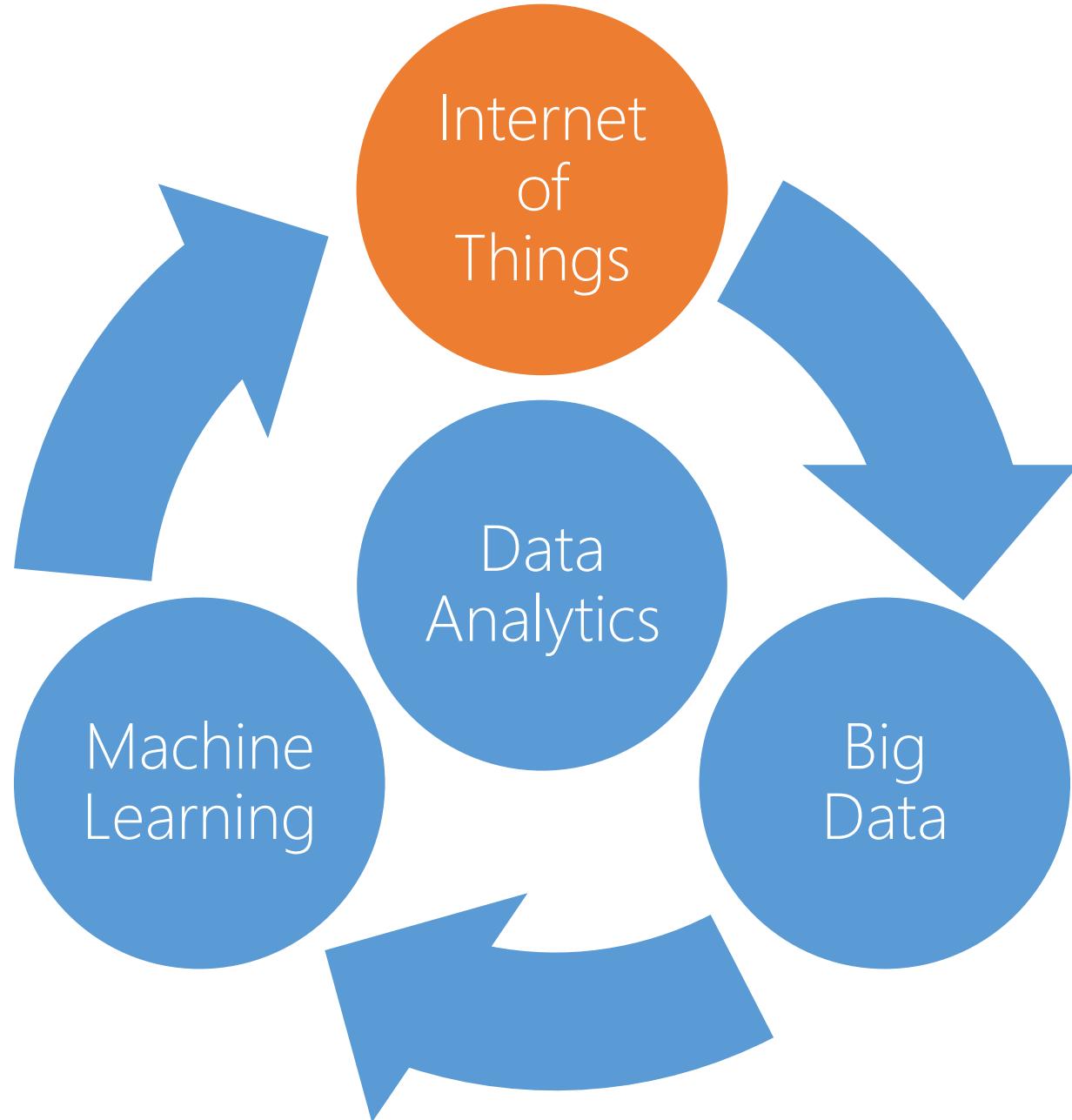
6% higher profits





Empowering people to make better decisions is not the end goal...

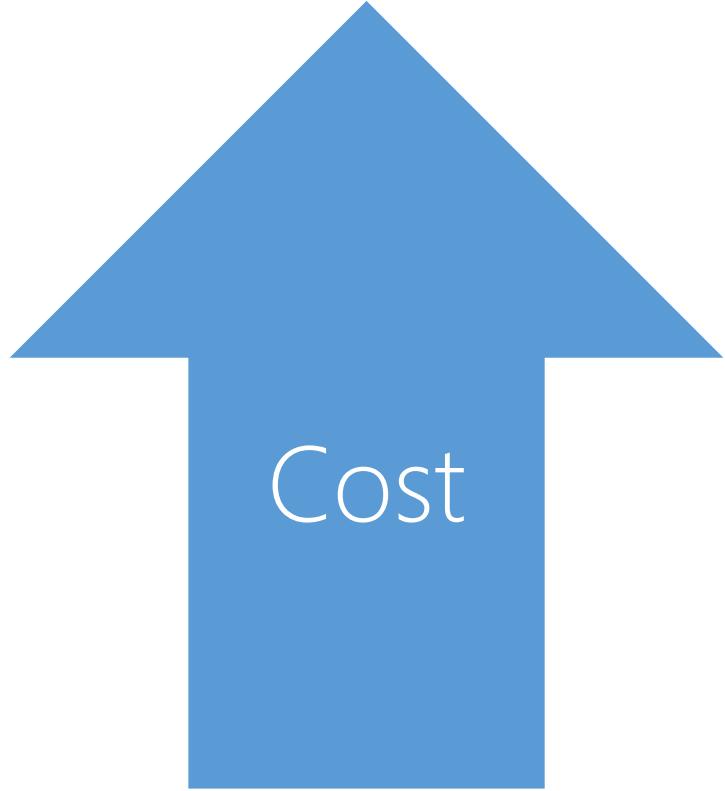
... it's just the beginning



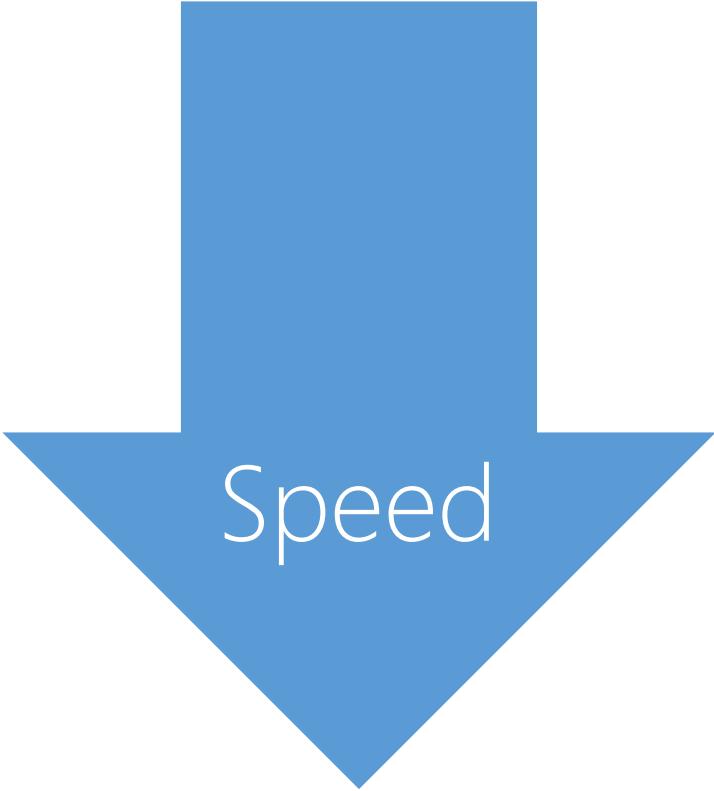
The Internet (The Past)



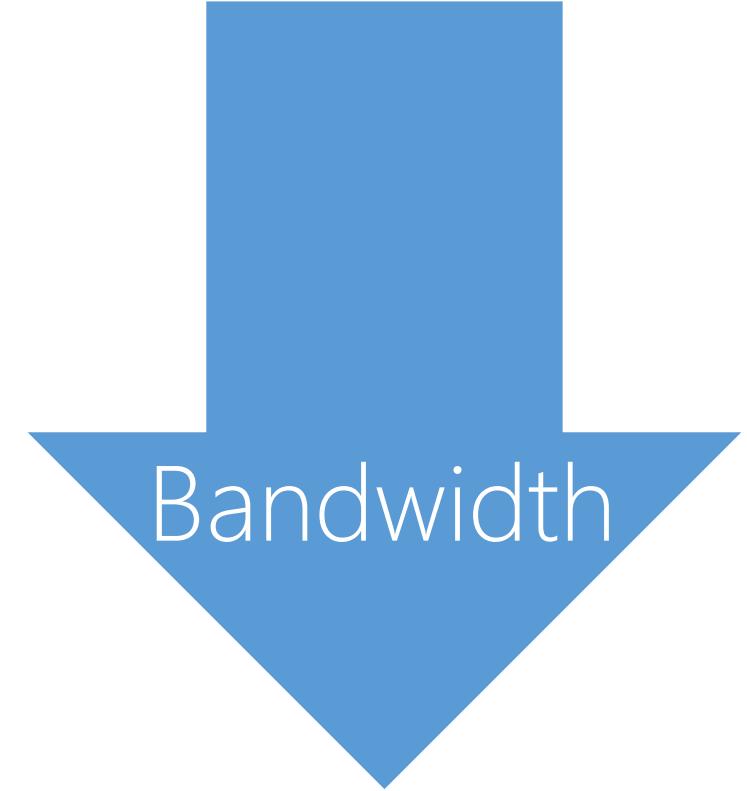




Cost



Speed



Bandwidth



World Wide Web

The WorldWideWeb (W3) is a wide-area [hypermedia](#) information retrieval initiative aiming to give universal access to a large universe of documents.

Everything there is online about W3 is linked directly or indirectly to this document, including an [executive summary](#) of the project, [Mailing lists](#), [Policy](#), November's [W3 news](#), [Frequently Asked Questions](#).

[What's out there?](#)

Pointers to the world's online information, [subjects](#), [W3 servers](#), etc.

[Help](#)

on the browser you are using

[Software Products](#)

A list of W3 project components and their current state. (e.g. [Line Mode](#), [X11 Viola](#), [NeXTStep](#), [Servers](#), [Tools](#), [Mail robot](#), [Library](#))

[Technical](#)

Details of protocols, formats, program internals etc

[Bibliography](#)

Paper documentation on W3 and references.

[People](#)

A list of some people involved in the project.

[History](#)

A summary of the history of the project.

[How can I help ?](#)

If you would like to support the web..

[Getting code](#)

Getting the code by [anonymous FTP](#), etc.

File Edit View Go Favorites Help



Address: http://info.cern.ch/hypertext/WWW/

World Wide

The WorldWideWeb (W3) i

Everything there is online a
[Policy](#), November's [W3 new](#)

[What's out there?](#)

Pointers to the world'

[Help](#)

on the browser you ar
[Software Products](#)

A list of W3 project c
[Technical](#)

Details of protocols, f

[Bibliography](#)

Paper documentation
[People](#)

A list of some people

[History](#)
A summary of the his

[How can I help ?](#)

If you would like to s

[Getting code](#)

Getting the code by [a](#)

File Edit View Go Favorites Help

Sign Up for our E-Mail Newsletter

First name:

Last name:

Title:

Organization:

nonprofit foundation consulting firm other

Street address:

City:

State:

Country:

Zip or postal code:

E-mail address: @

Voice phone: () Ext.

Fax: ()

I would prefer to receive information via

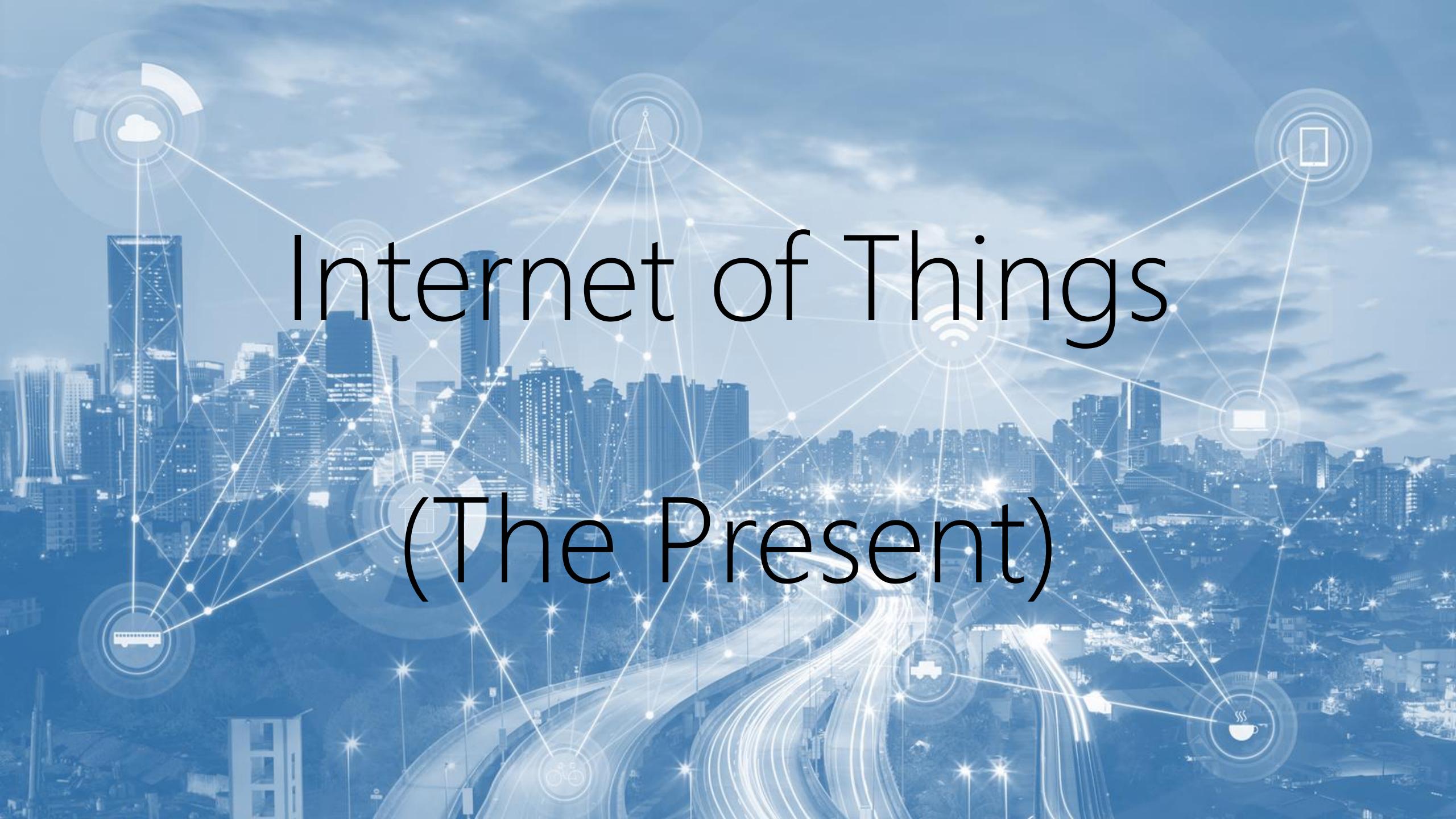
E-mail Fax

File Edit View Go Favorites Help



Address: <http://info.cern.ch/hypertext/WWW/TheProject.html>

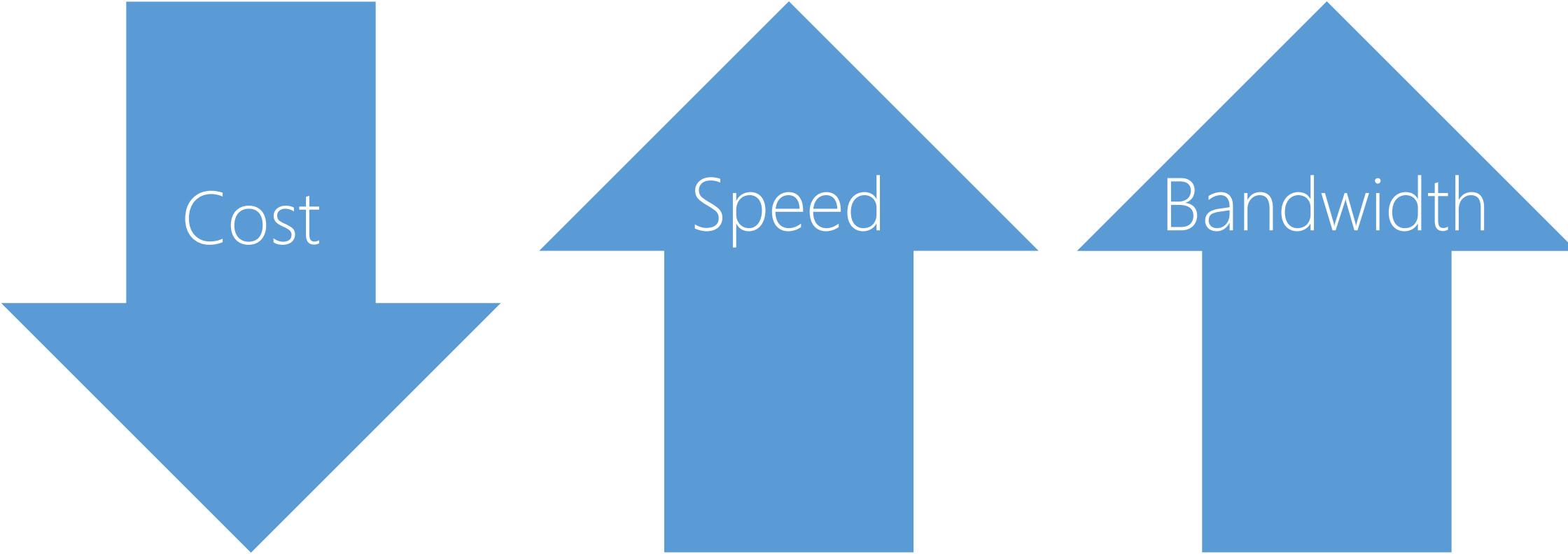
The internet was expensive, slow,
and not generating much data.



Internet of Things

(The Present)





Cost

Speed

Bandwidth







REMOTE HOME CONTROL

ONLINE HOME AUTOMATION



CLIMATE



LIGHTING



SECURITY



ENTERTAINMENT



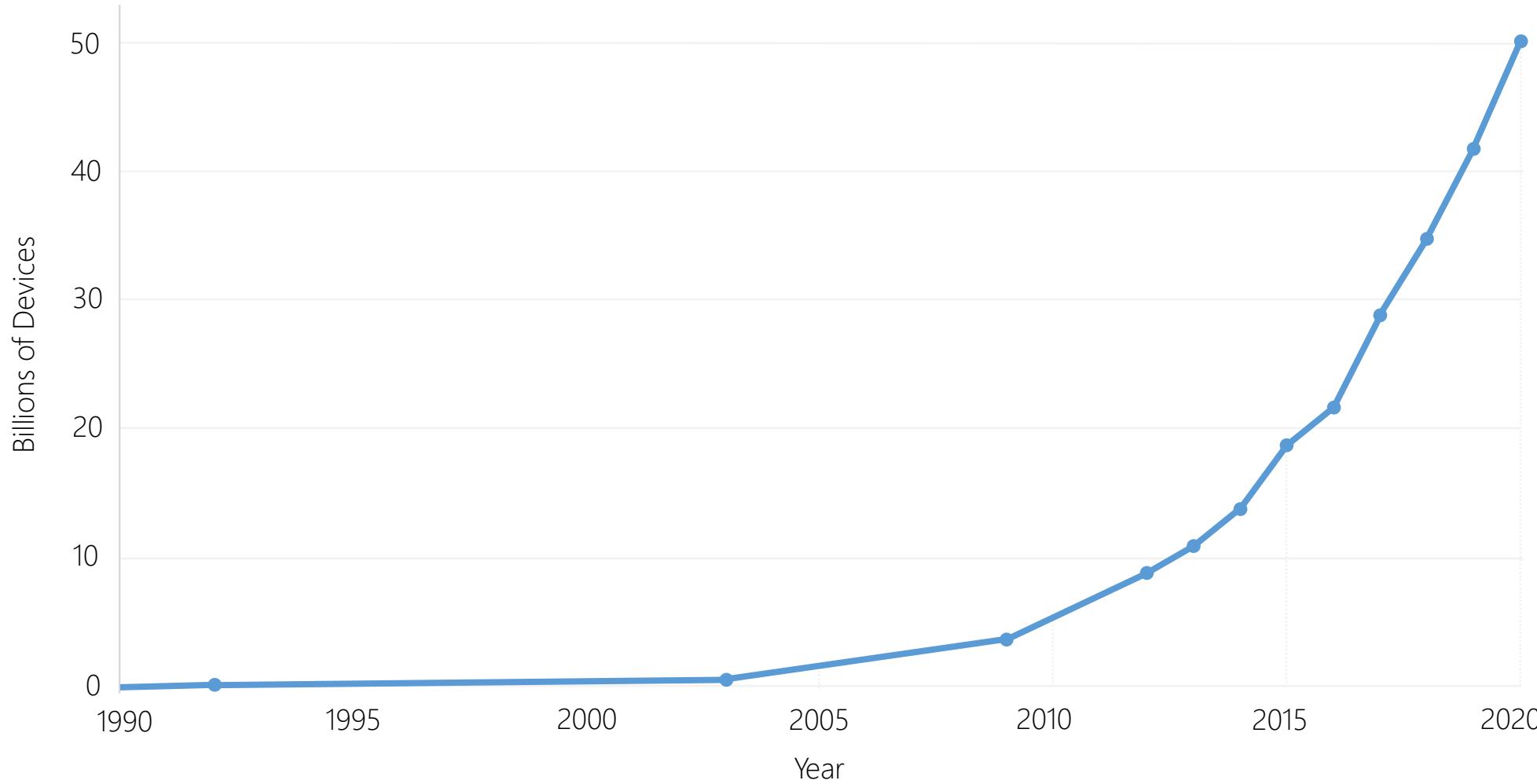
ECO MODE

NORMAL

OFF



Growth of the IoT Devices

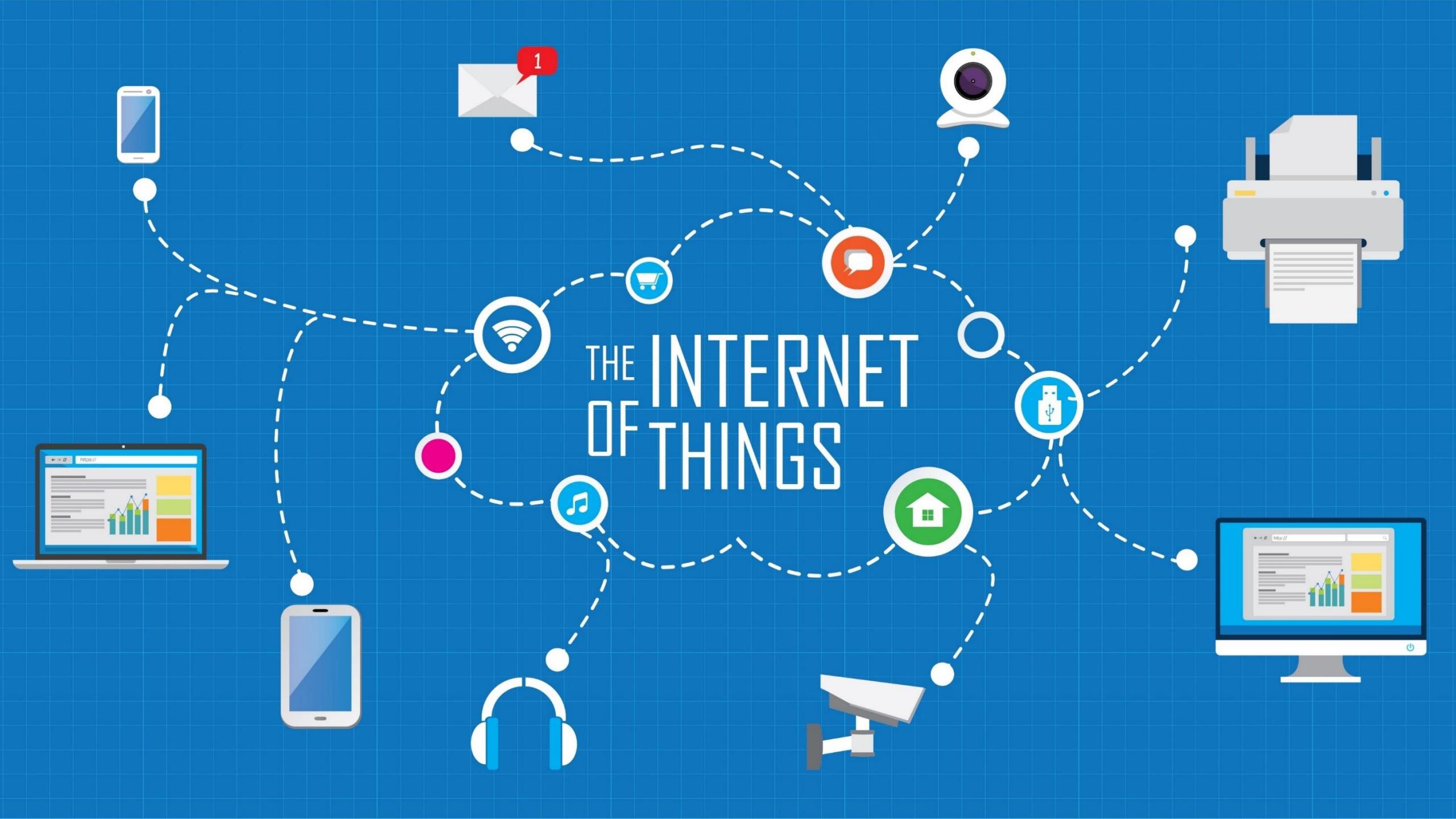


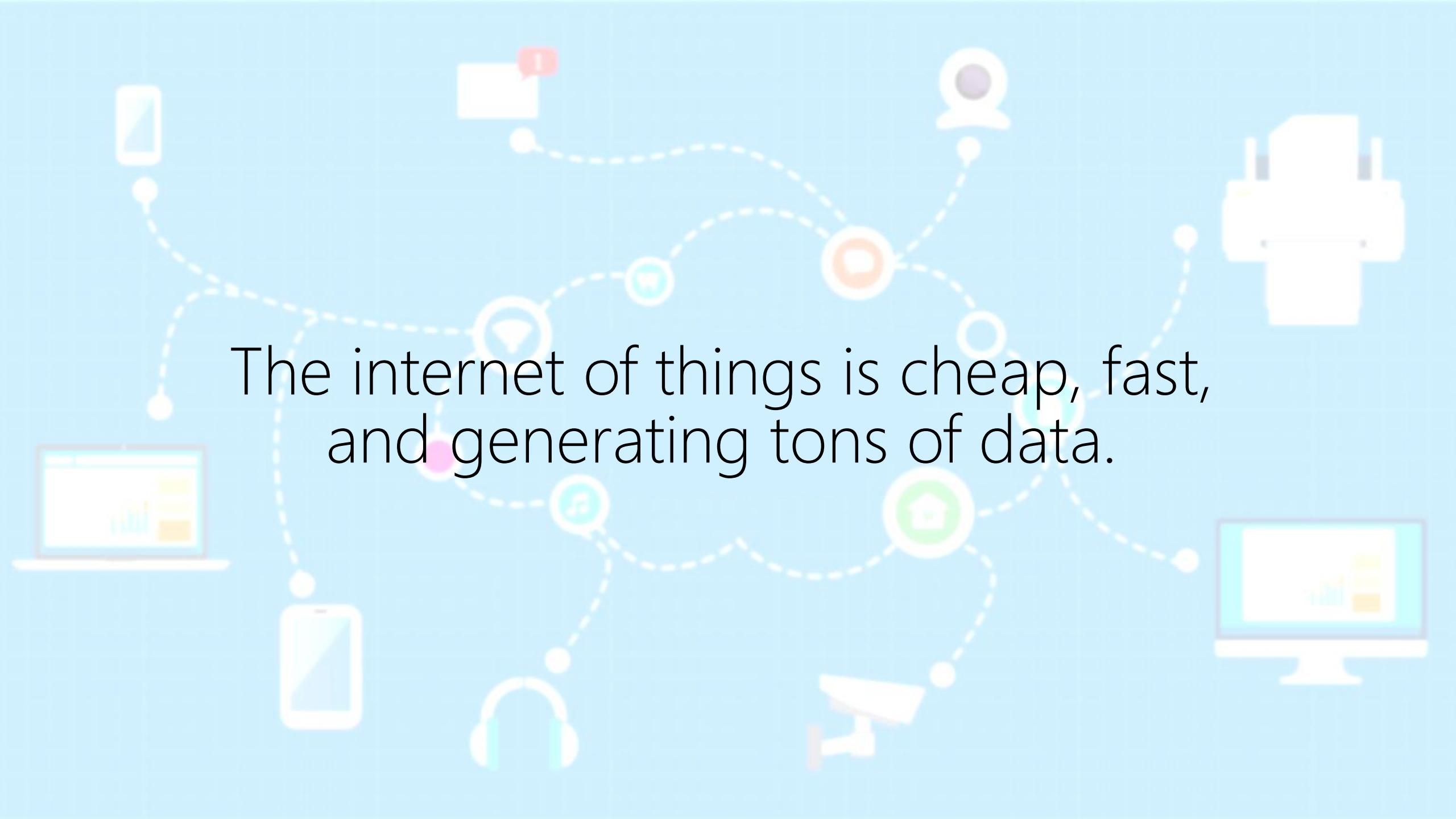
Source: NCTA, 2014

Growth of the IoT Devices

50 billion IoT devices by 2020

THE INTERNET OF THINGS

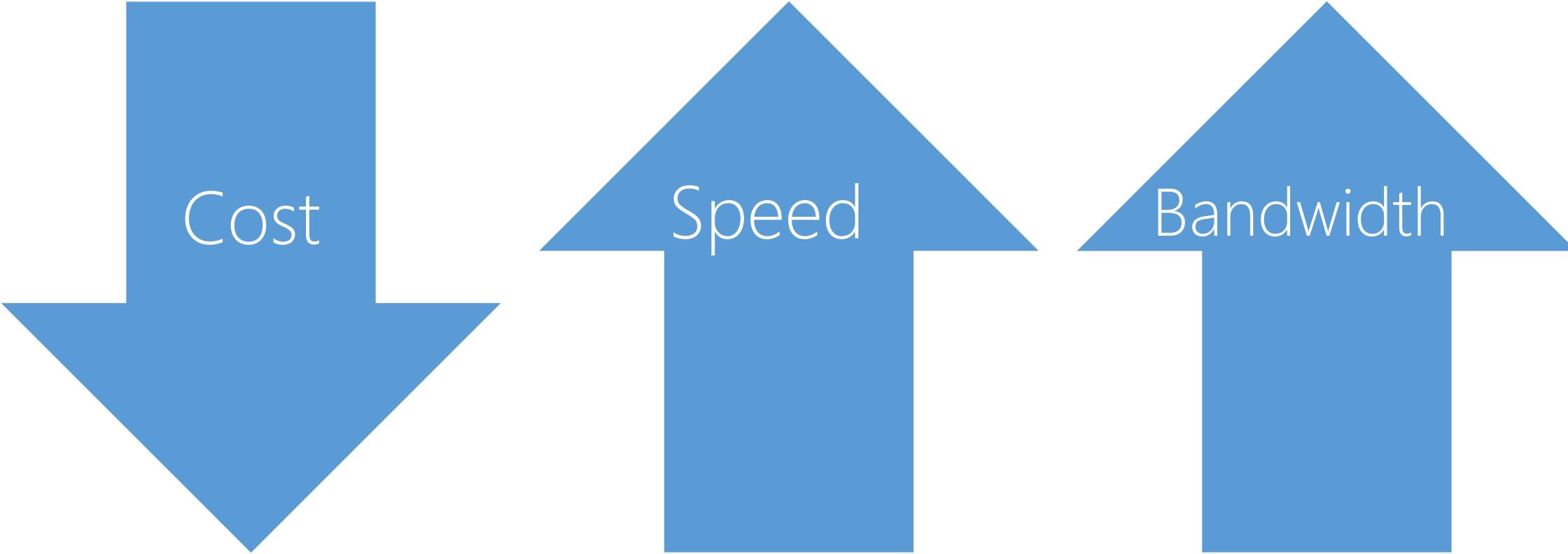




The internet of things is cheap, fast,
and generating tons of data.

Internet of Everything

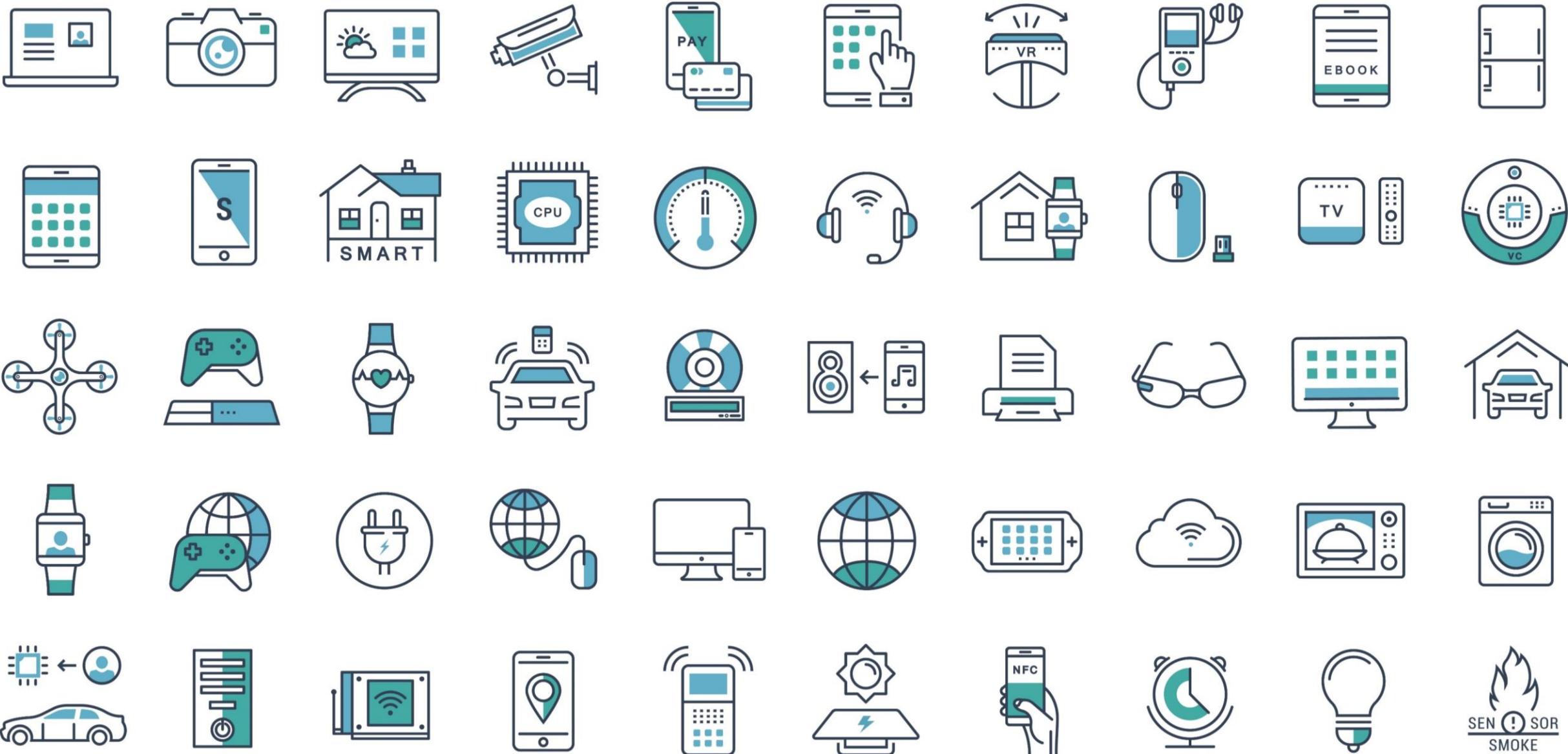
(The Future)

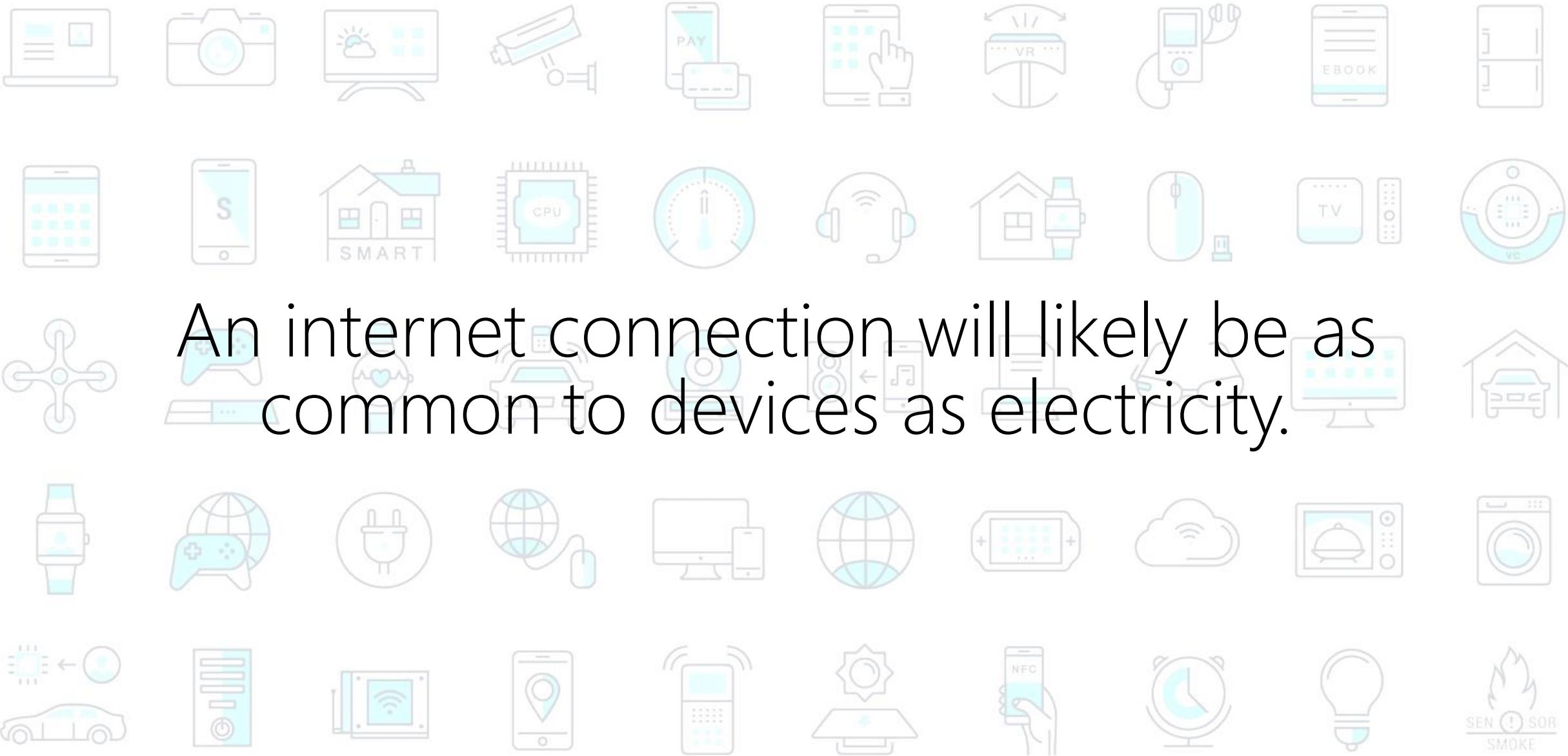


Cost

Speed

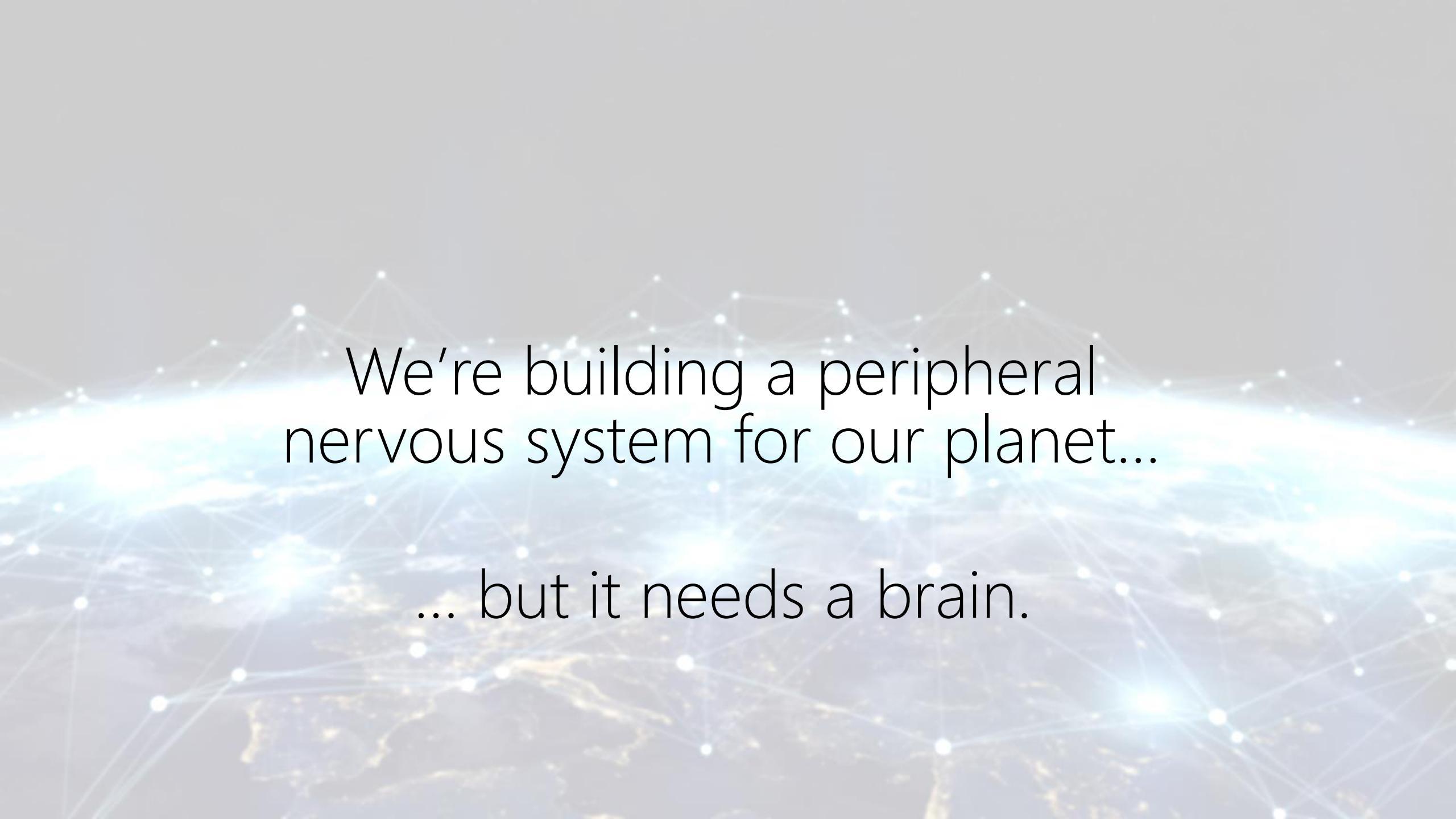
Bandwidth





An internet connection will likely be as common to devices as electricity.





We're building a peripheral
nervous system for our planet...

... but it needs a brain.

FUN GAME 1

Is It IoT?

Is it IoT?



YES!



Is it IoT?



YES!



Is it IoT?

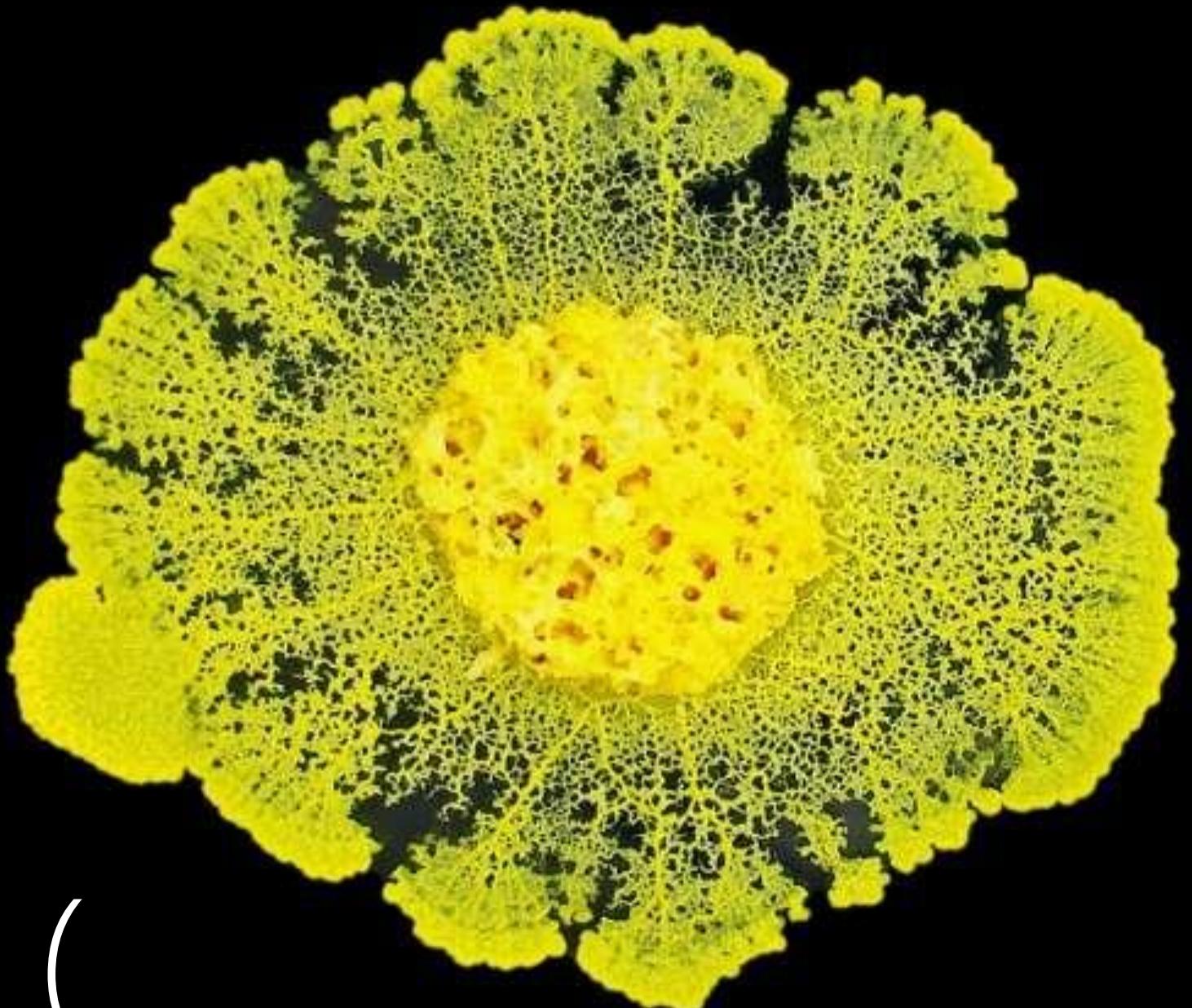


YES!

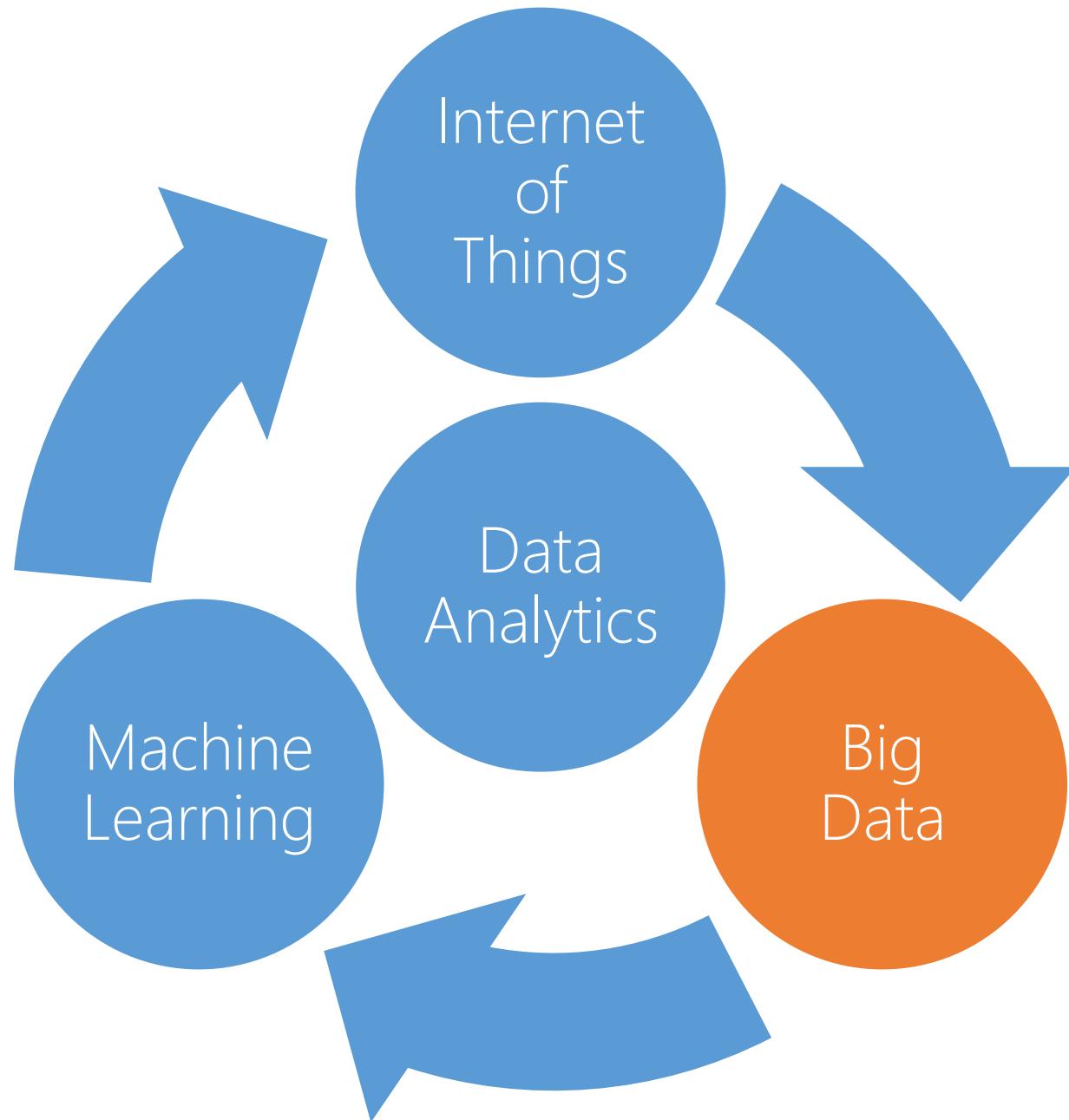




Is it IoT?



NO : (



Data (The Past)













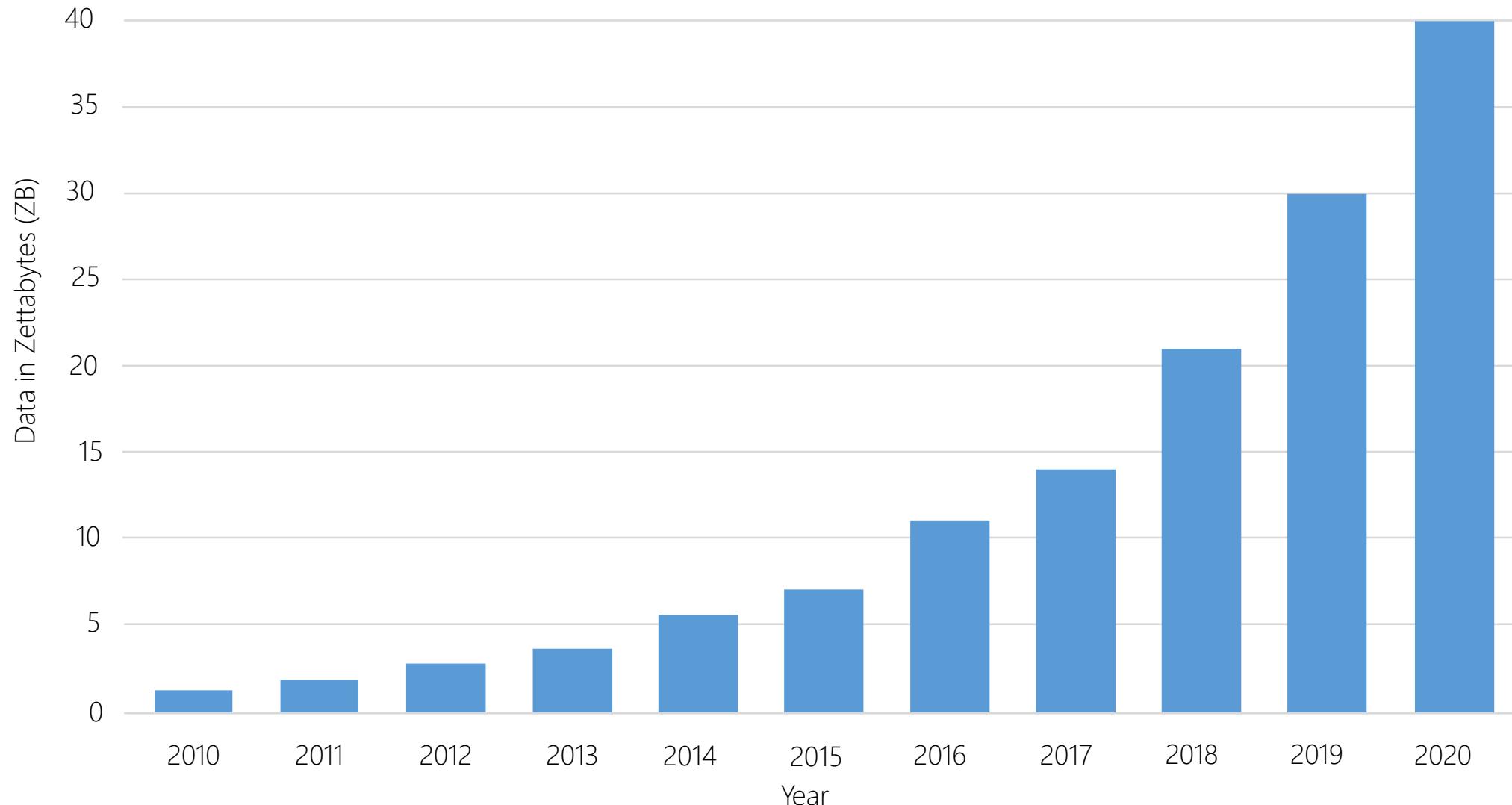
Data sets were small, slow,
and had little diversity.



Big Data

(The Present)

Global Data Growth

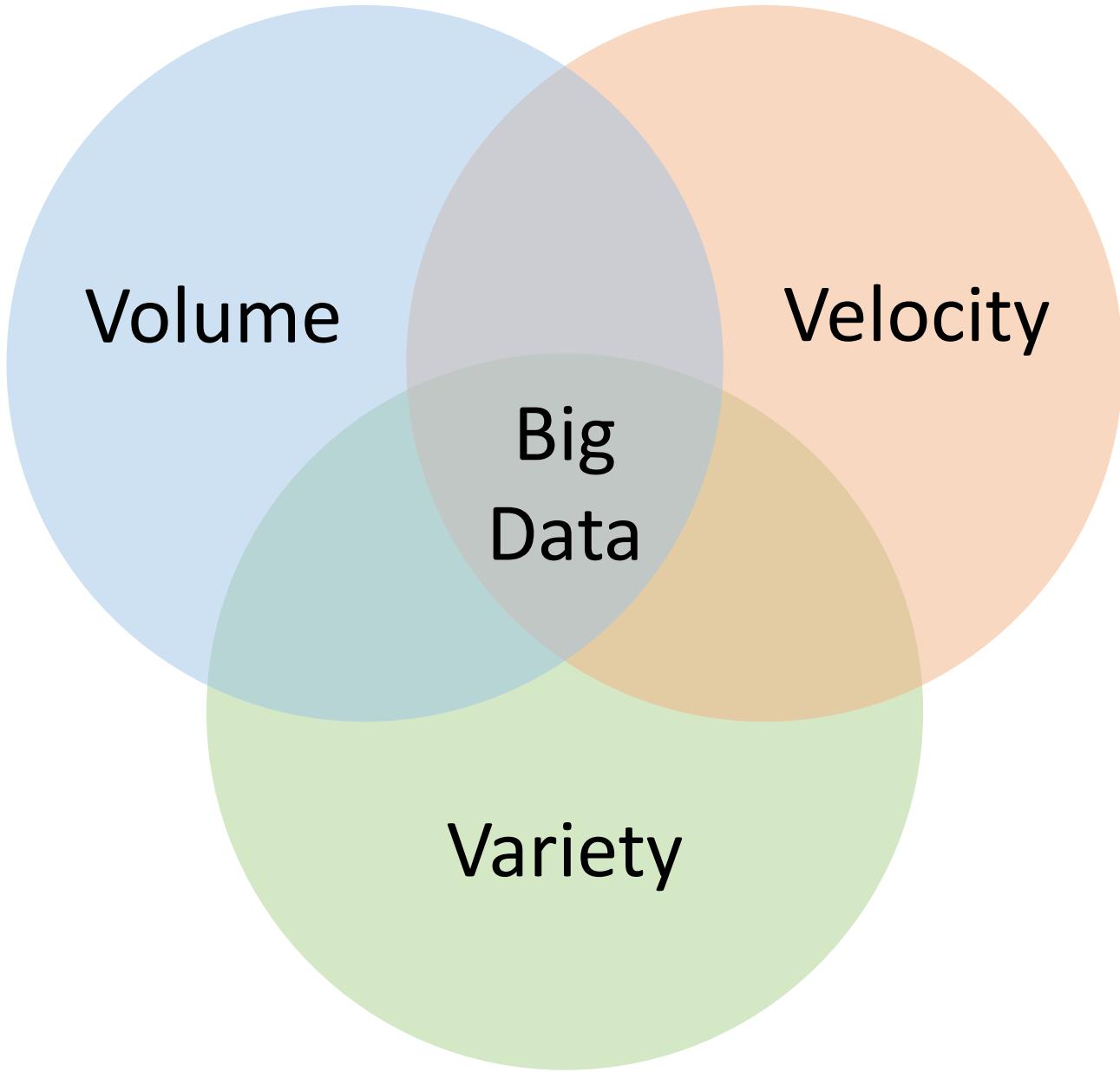


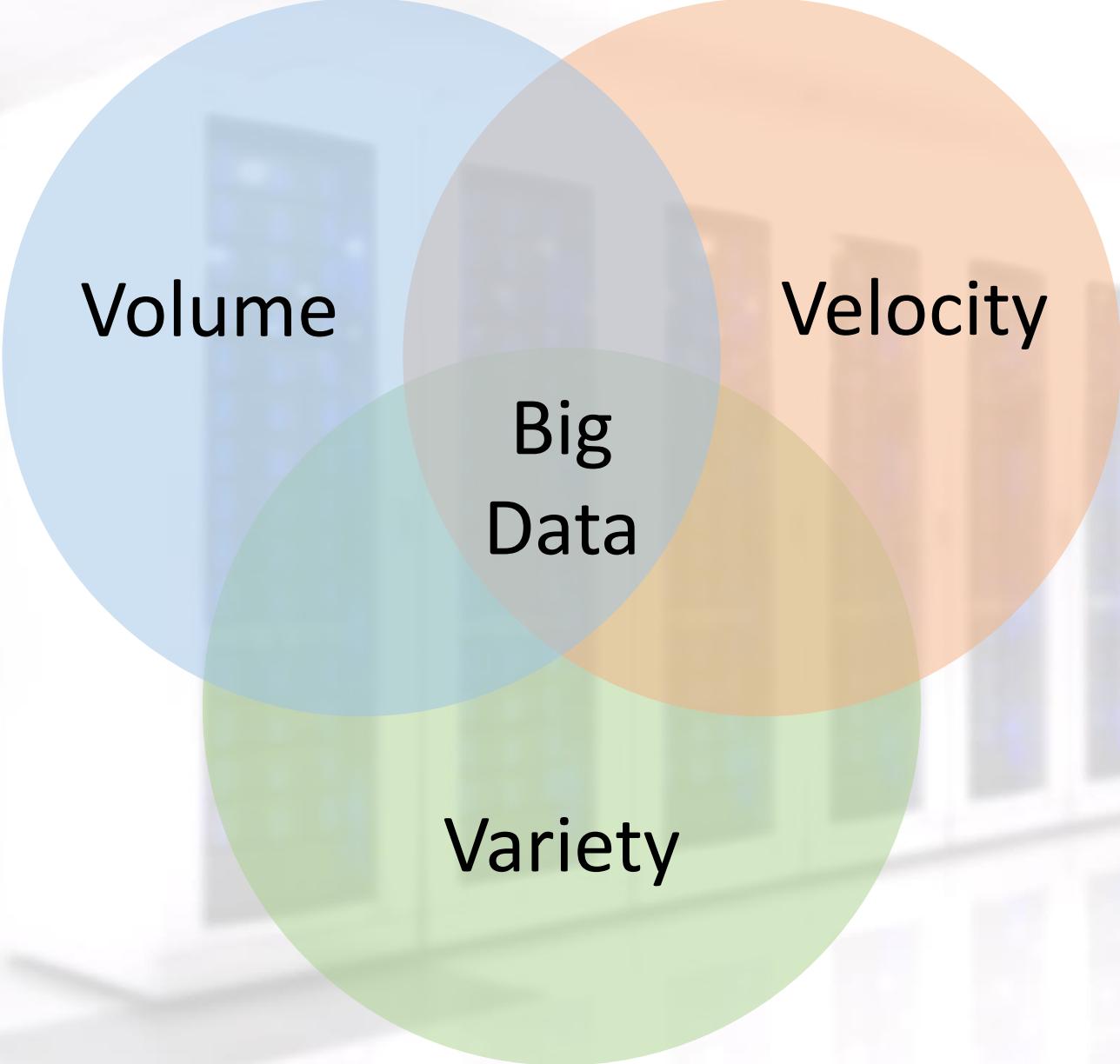
Source: UNECE Statistics Wikis

Global Data Growth

Doubling every two years







Volume

Velocity

Big
Data

Variety



Volume

A 3





Velocity





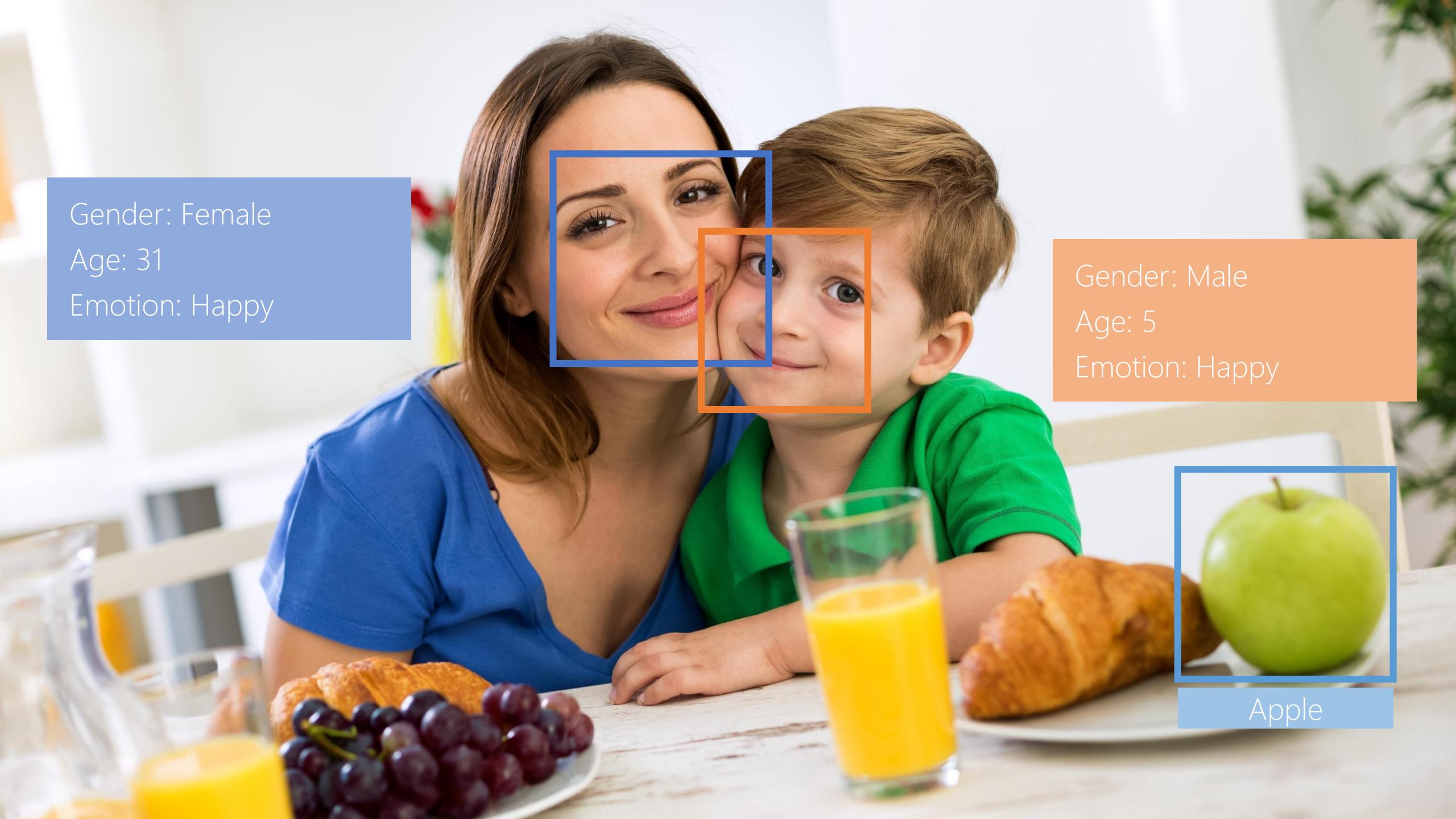


Variety

DATA

INTEGRATION





Gender: Female
Age: 31
Emotion: Happy

Gender: Male
Age: 5
Emotion: Happy



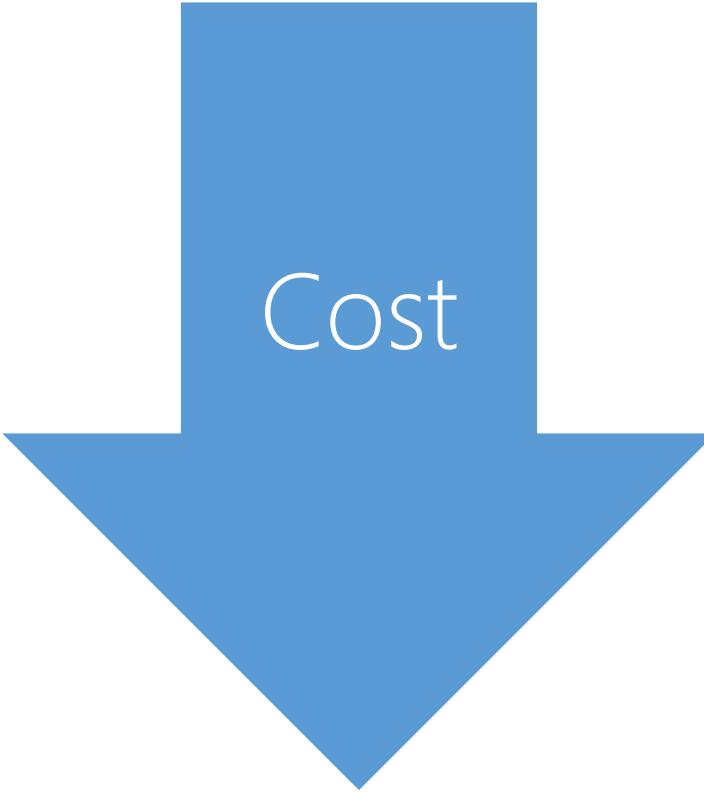
Apple

A perspective view of a long row of server racks in a data center. The racks are white with glass doors, showing internal components with blue and white lights. The floor has a metal grating. A sign above the racks reads "A 3".

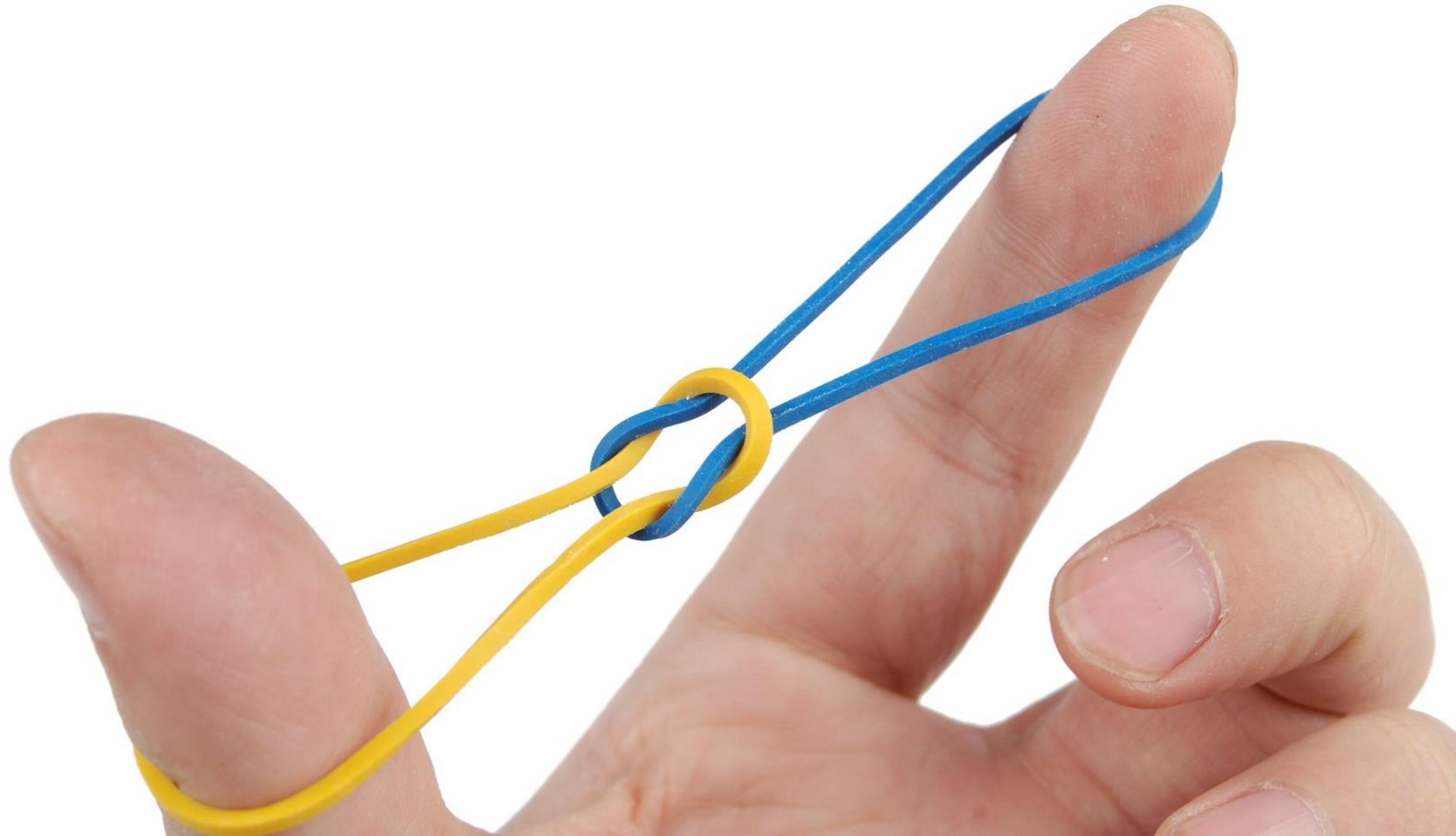
Today's data sets are bigger,
faster, and more diverse.

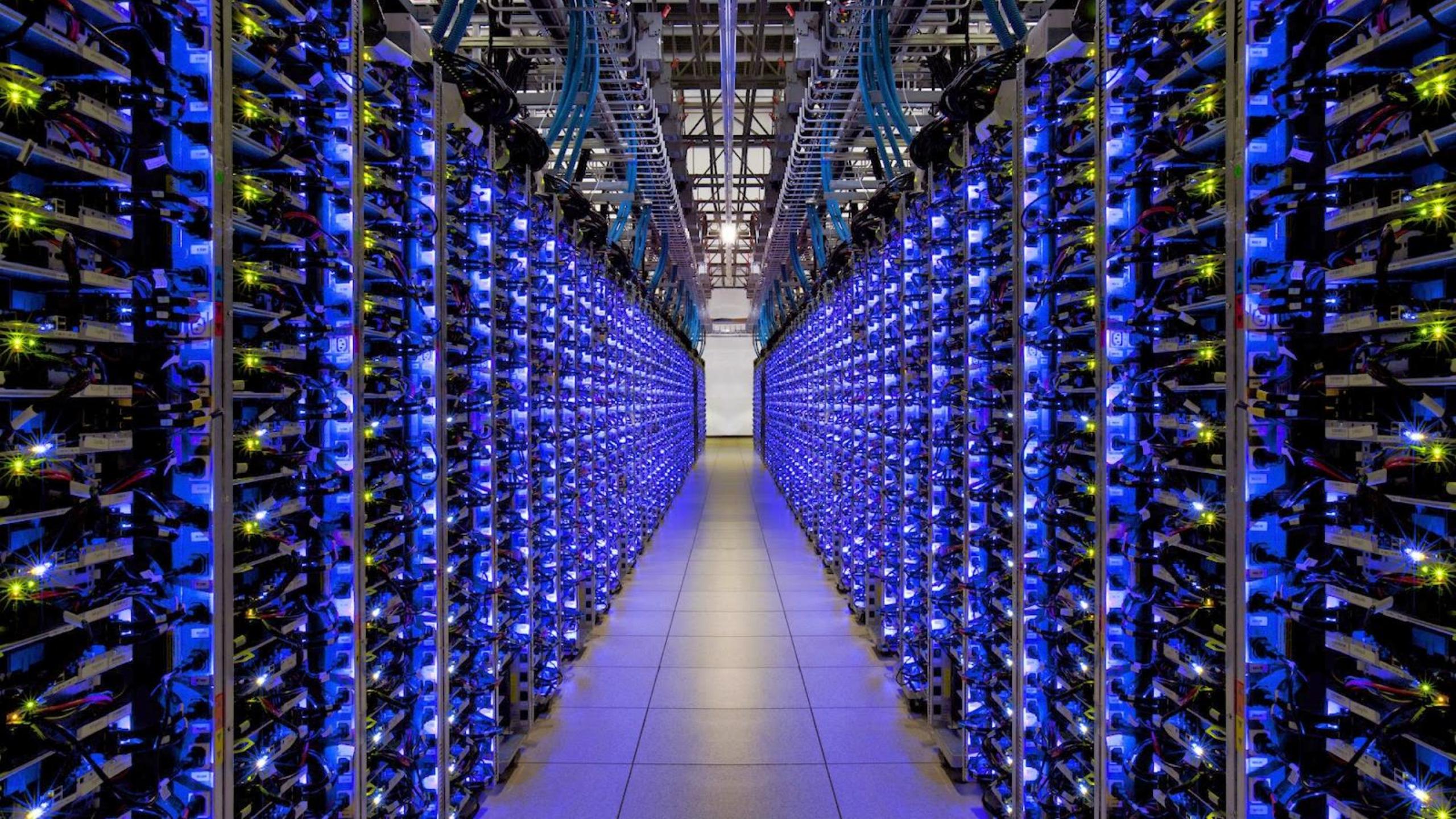
A perspective view looking down a aisle in a large data center. Both sides are filled with tall, dark server racks. Numerous small, bright yellow lights are visible along the top and sides of the racks, creating a pattern of glowing lines. The floor is a light-colored polished concrete.

Just Data Again
(The Future)



Cost





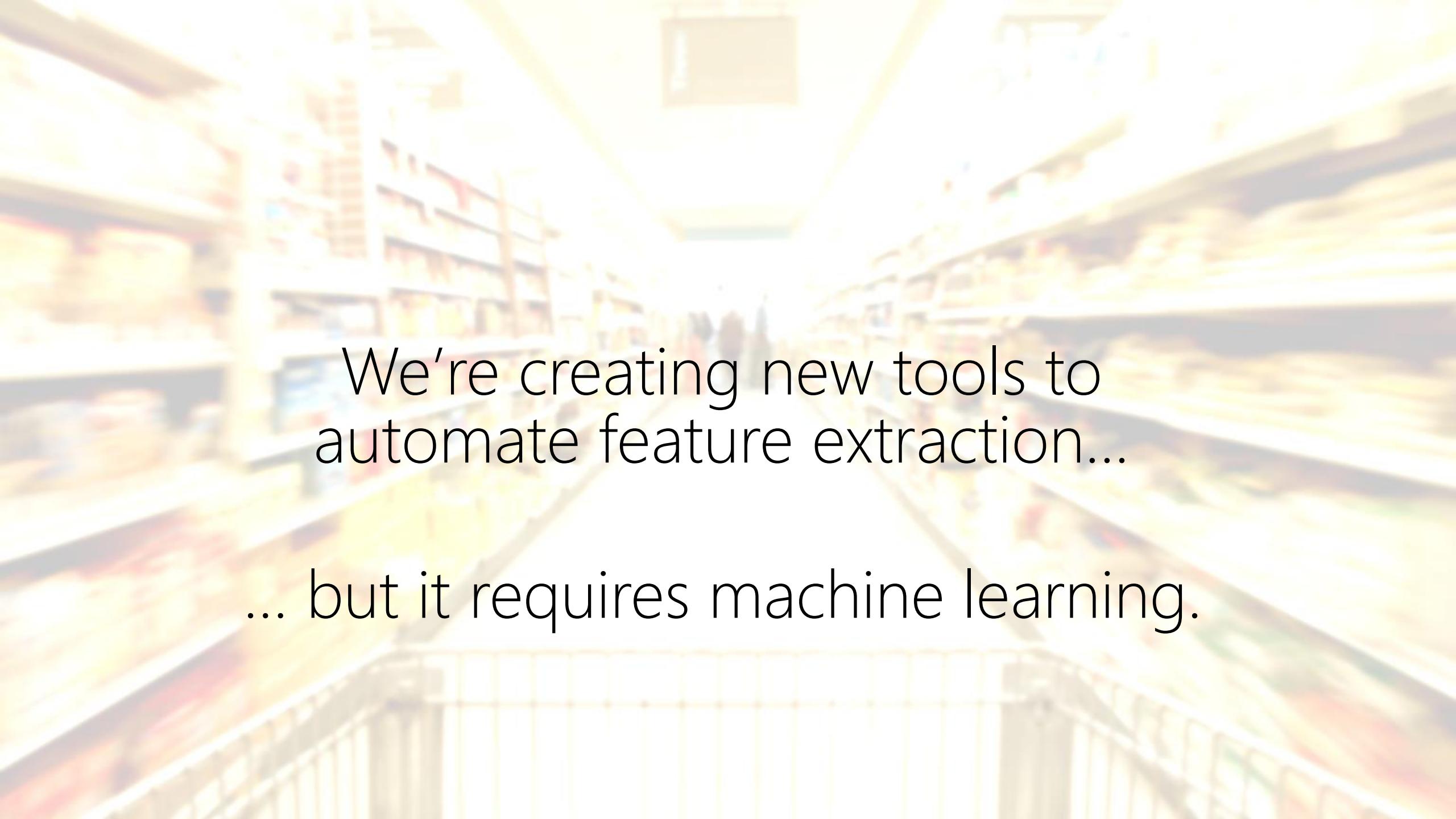


Candy
Cookies/Crackers

Food Vegetables
Goods

Three
Food Butter Margarine
Honey & Cereals
Biscuits & Biscuits
Cakes & Cakes
Sauces & Sauces
Spices & Spices
Cake Mixes
Jello Puddings

Cosmetics
Health Care



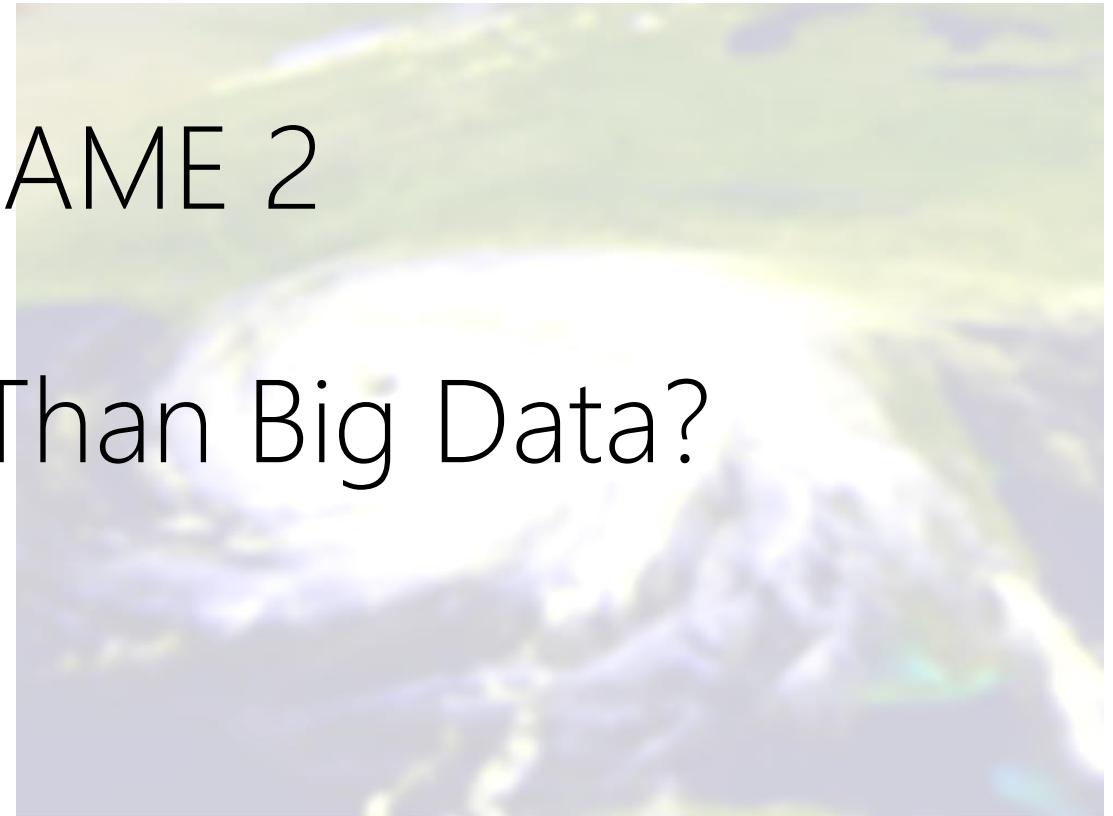
We're creating new tools to
automate feature extraction...

... but it requires machine learning.

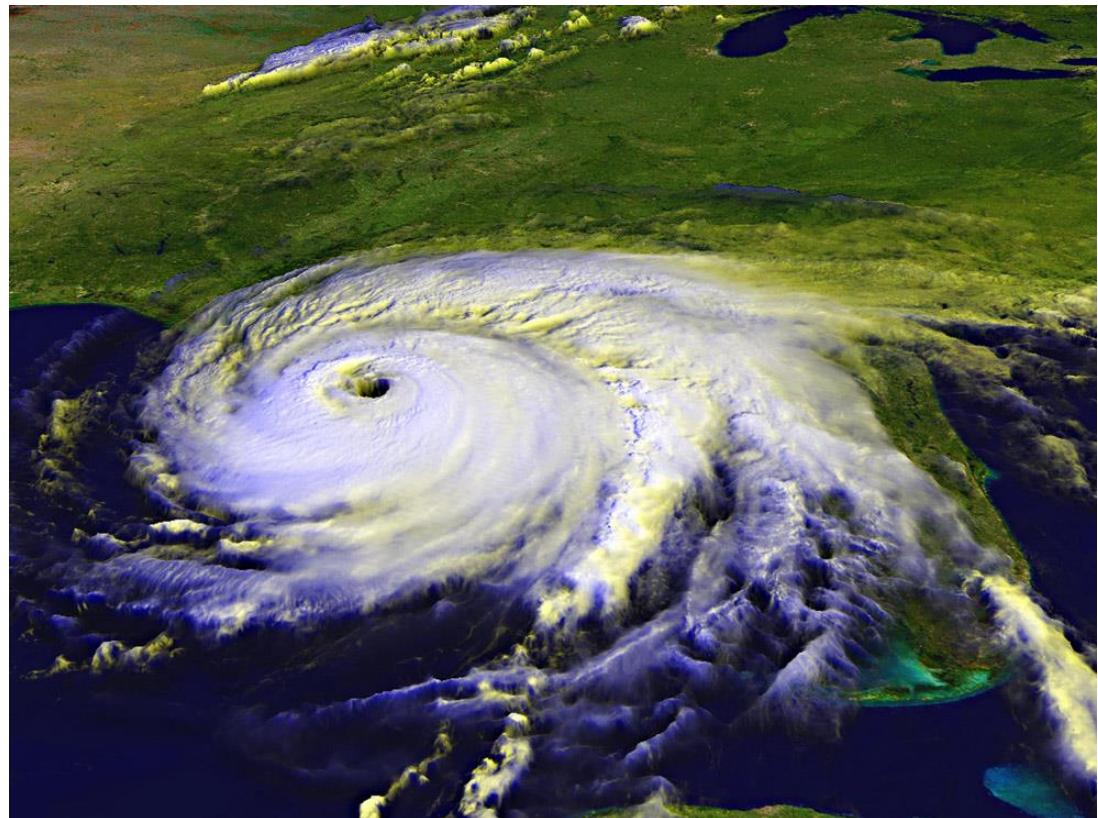


FUN GAME 2

Am I Smarter Than Big Data?



What Do These Have In Common?

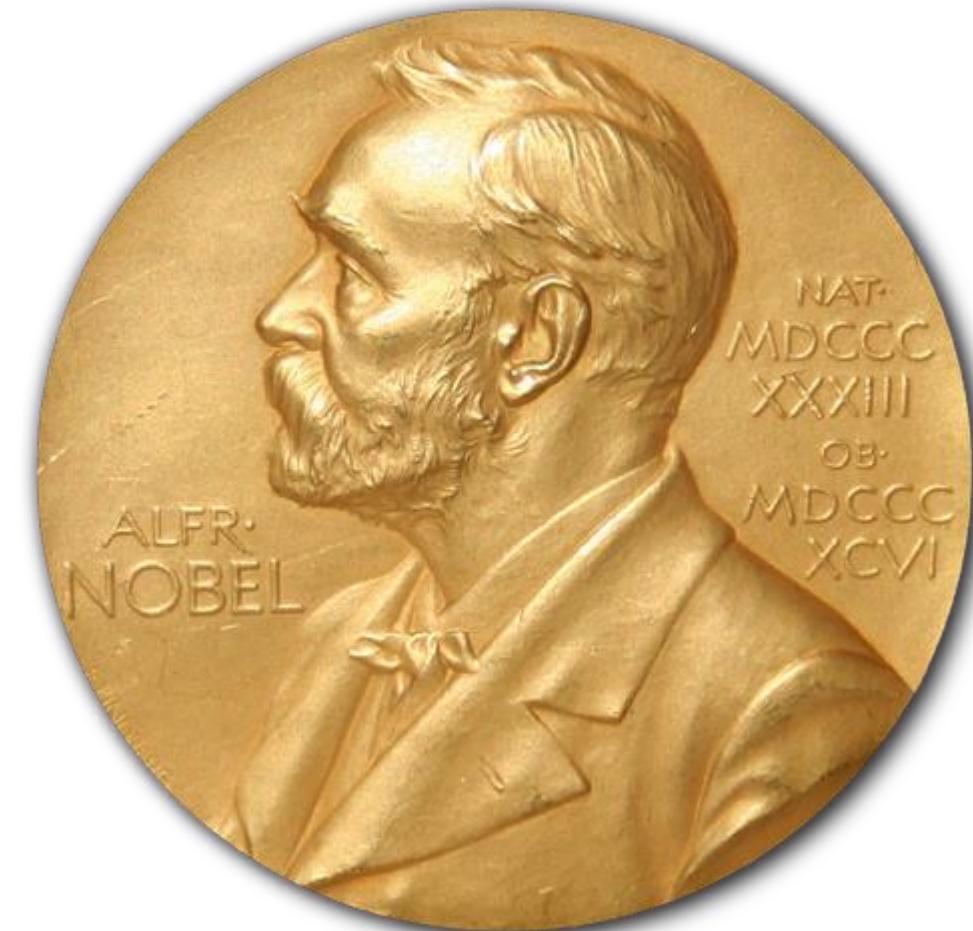


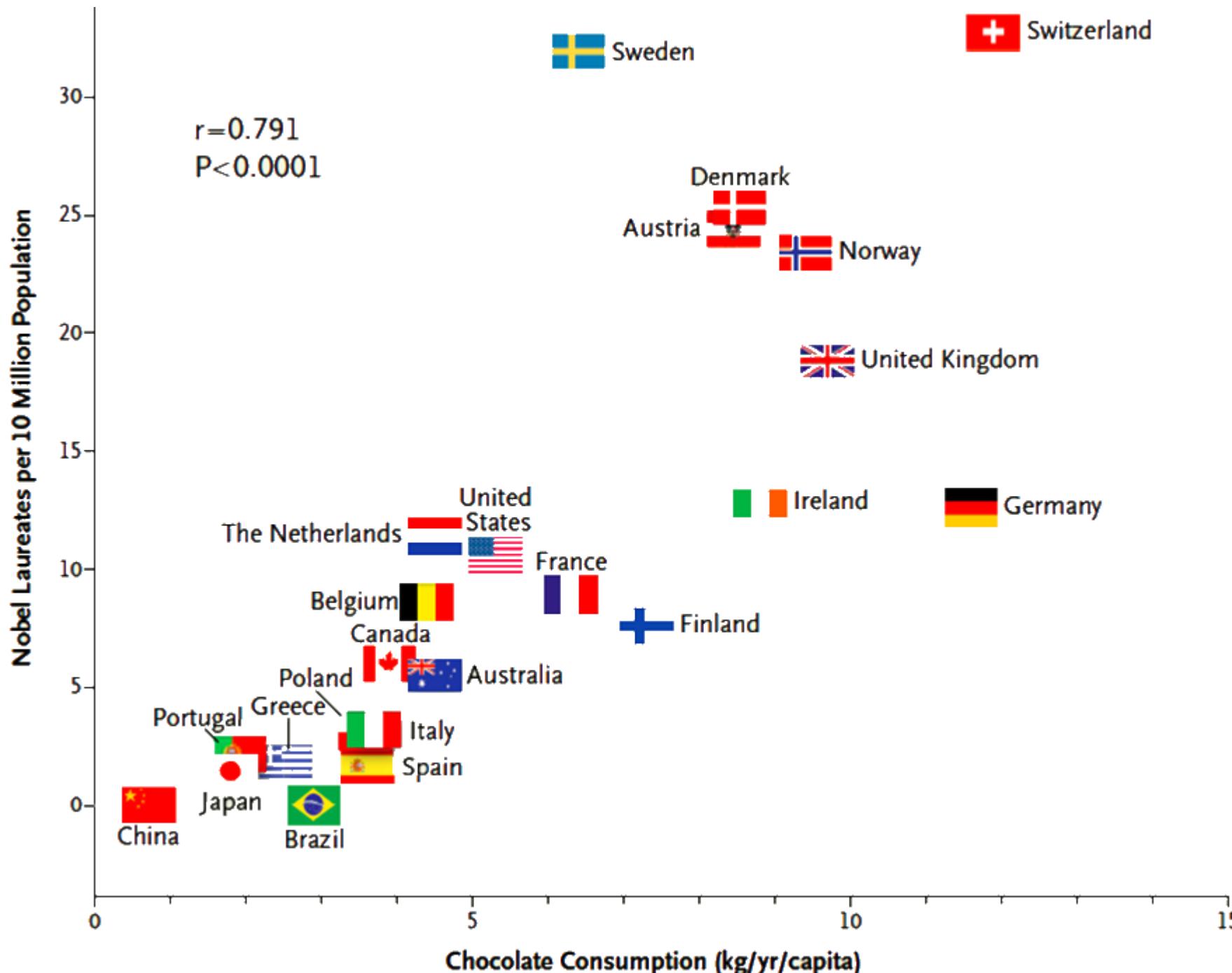


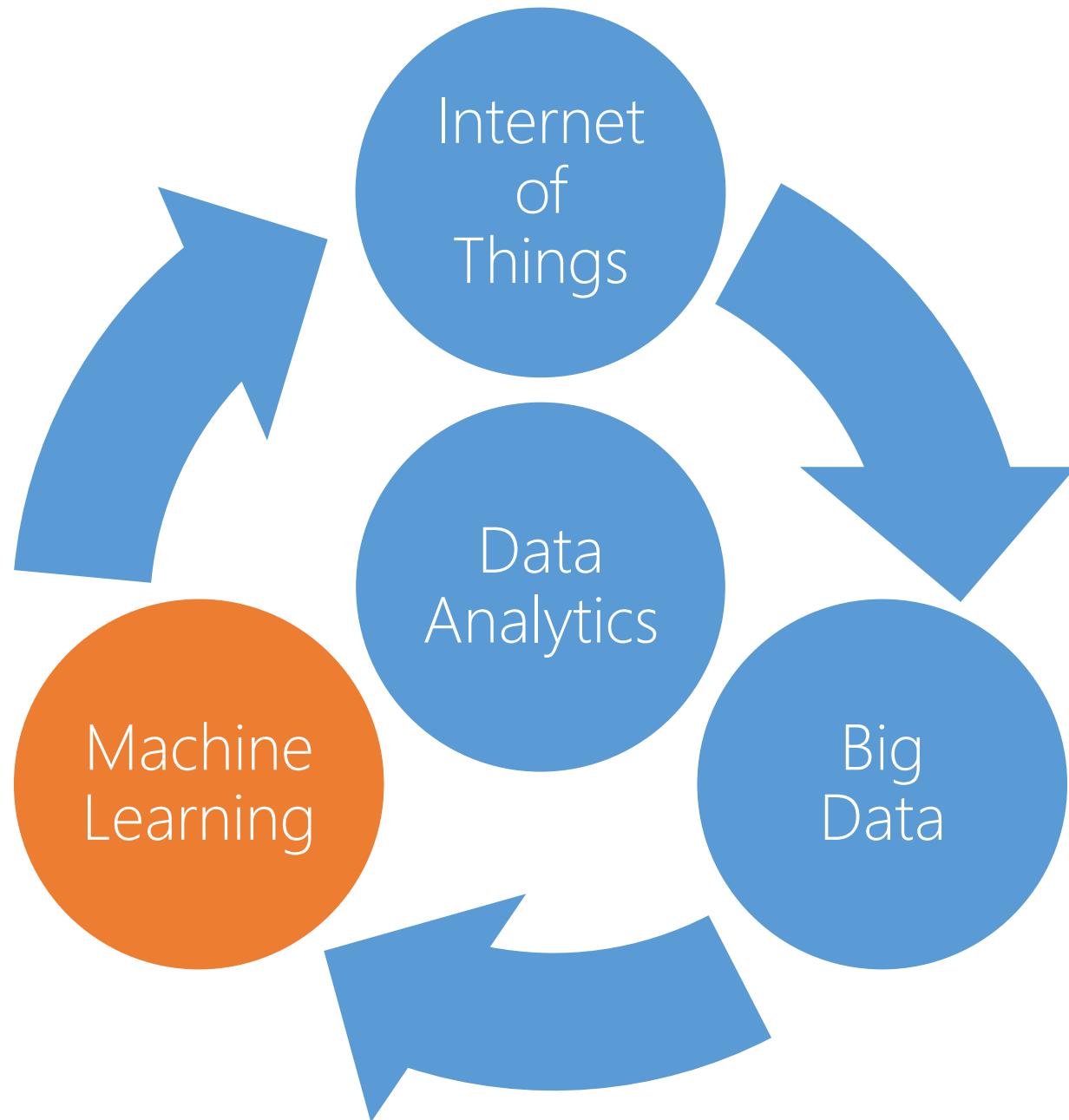
What Do These Three Things Predict?



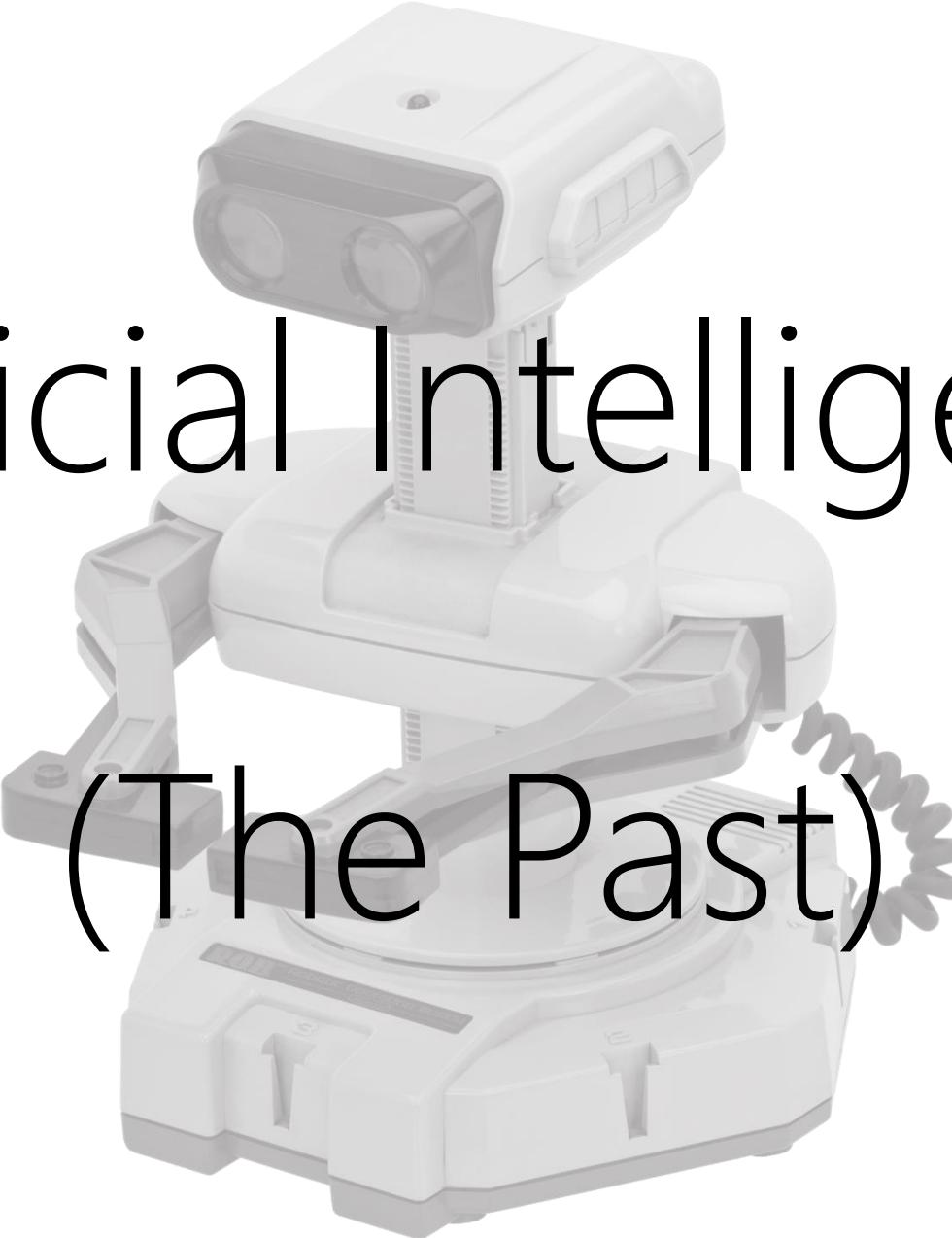








Artificial Intelligence (The Past)





Source: Evan Amos





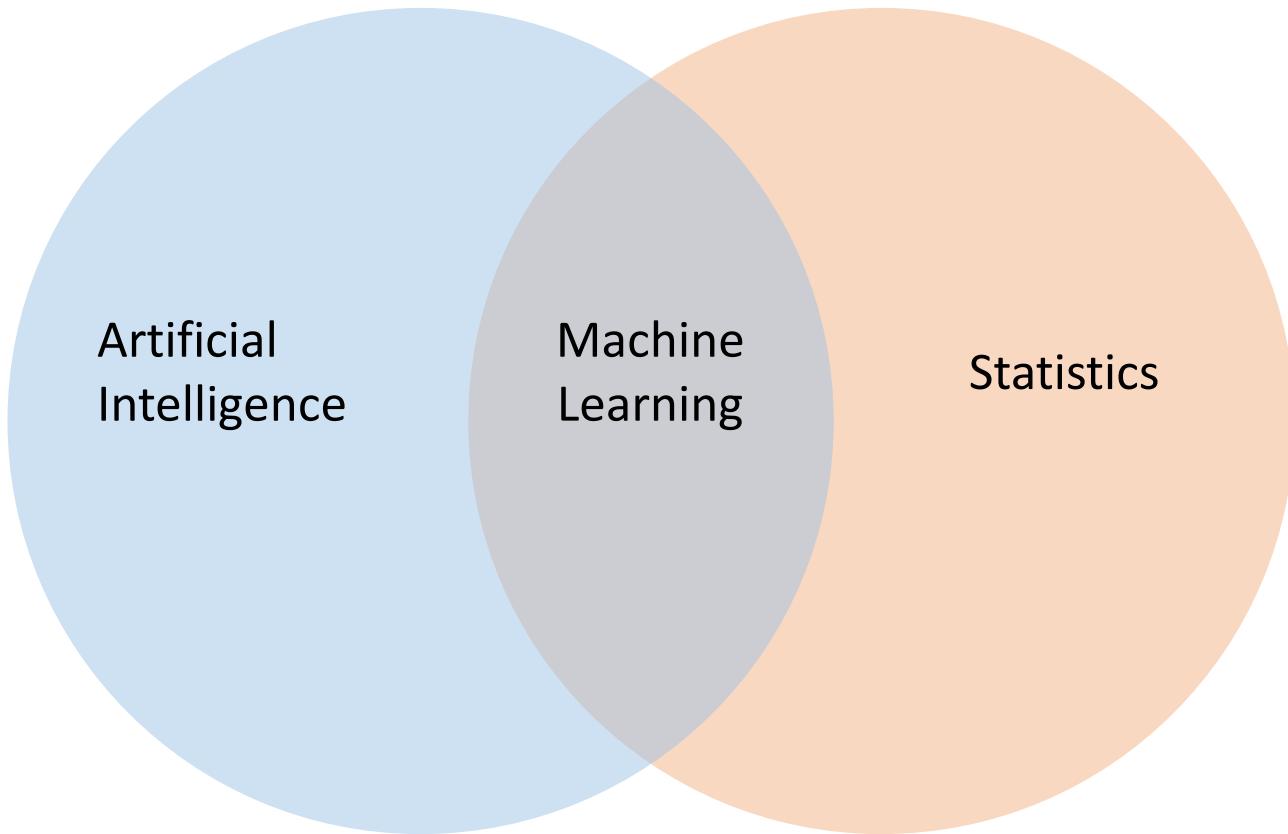


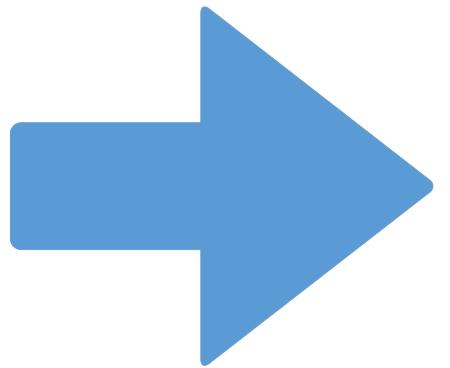
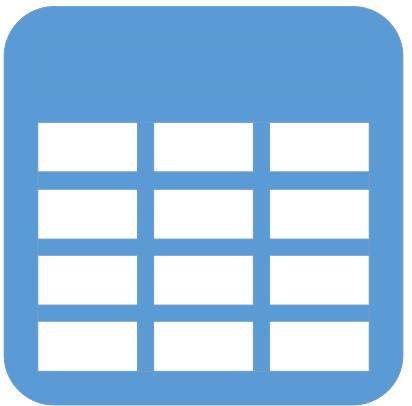
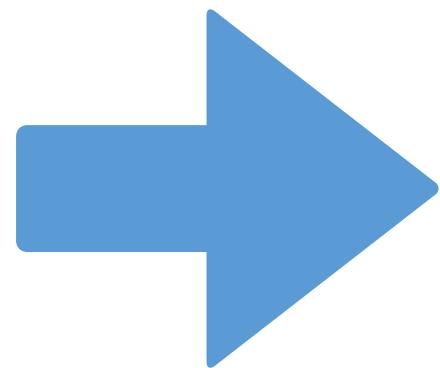
A black and white photograph showing the lower half of a person walking away from the viewer. The person is wearing dark trousers and light-colored shoes. They are walking on a path with trees and foliage visible in the background.

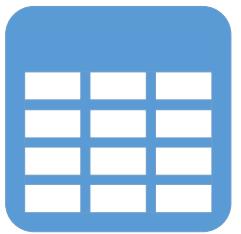
AI Winter has ended and
things are warming up again.

Machine Learning

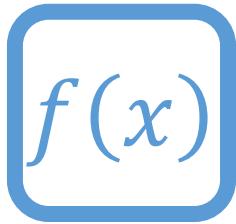
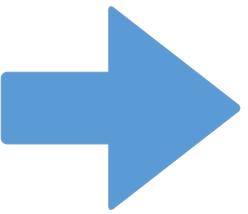
(The Present)



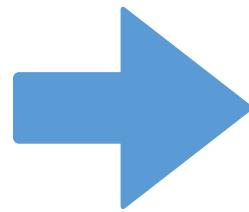
 $f(x)$ 



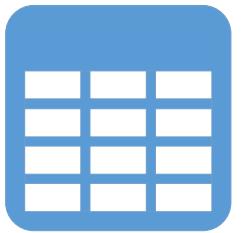
Data



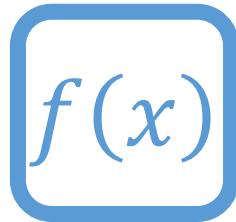
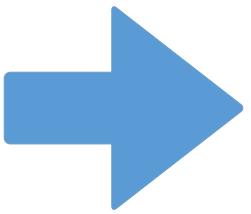
Function



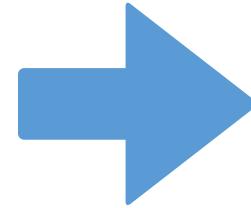
Prediction



Data



Function



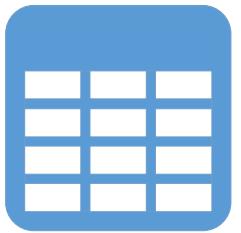
Prediction



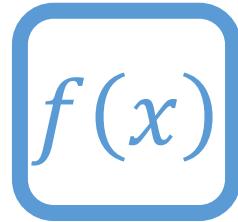
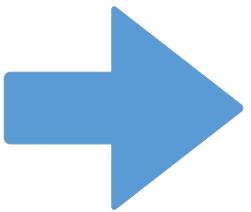
Cat



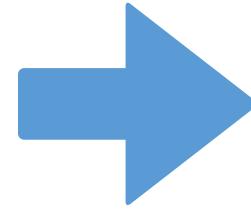
Dog



Data



Function



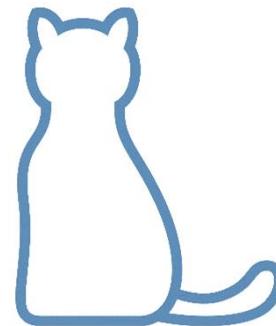
Prediction



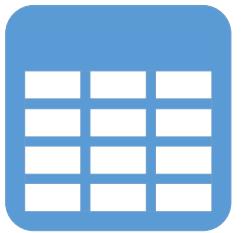
Cat



Dog



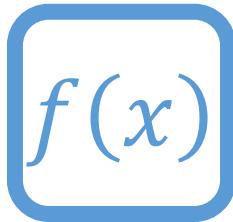
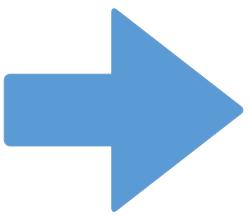
Is cat?



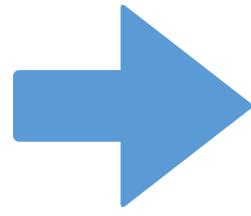
Data



Cat



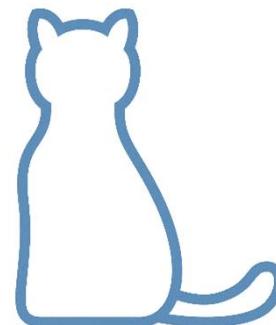
Function



Prediction



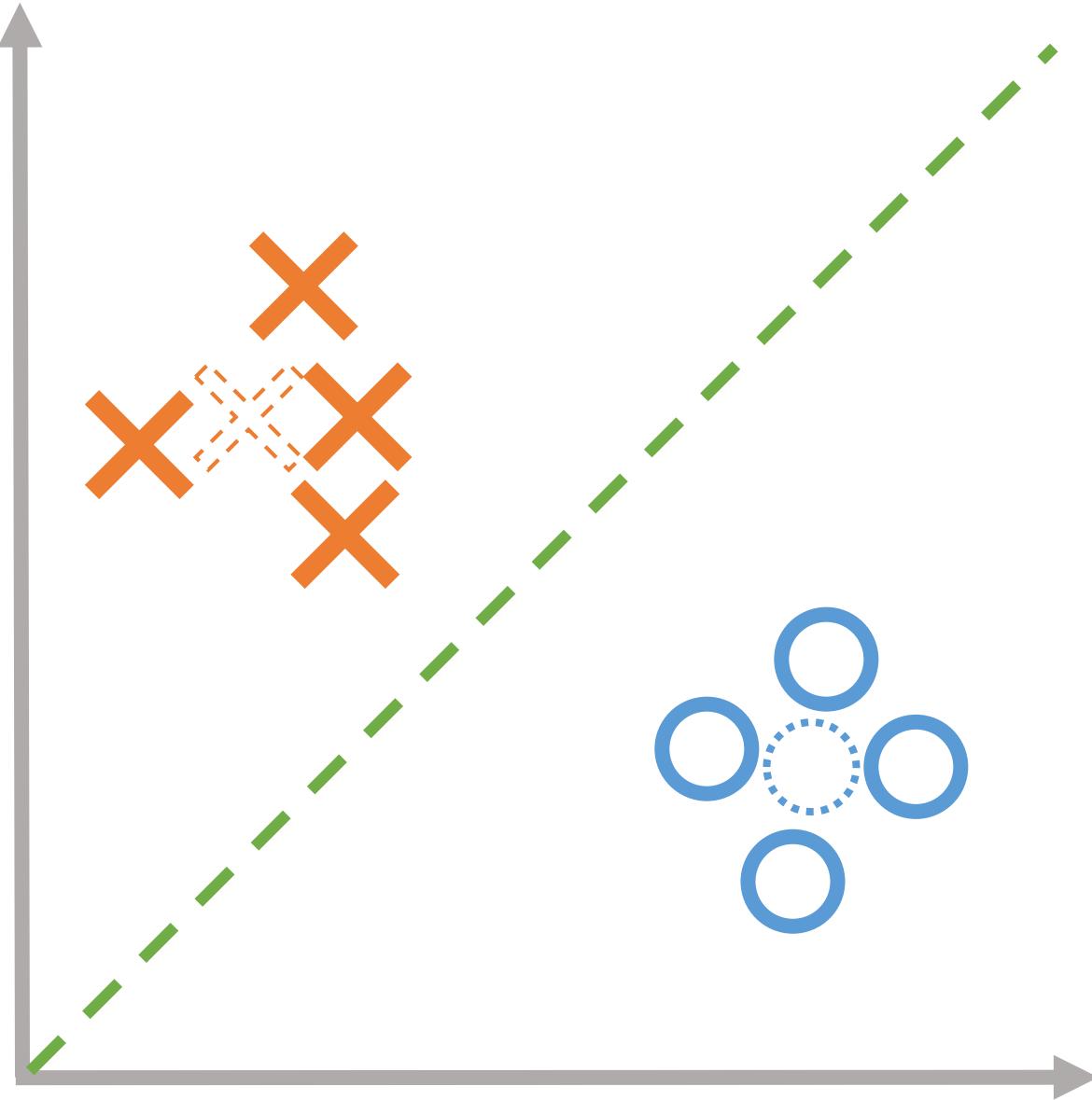
Dog

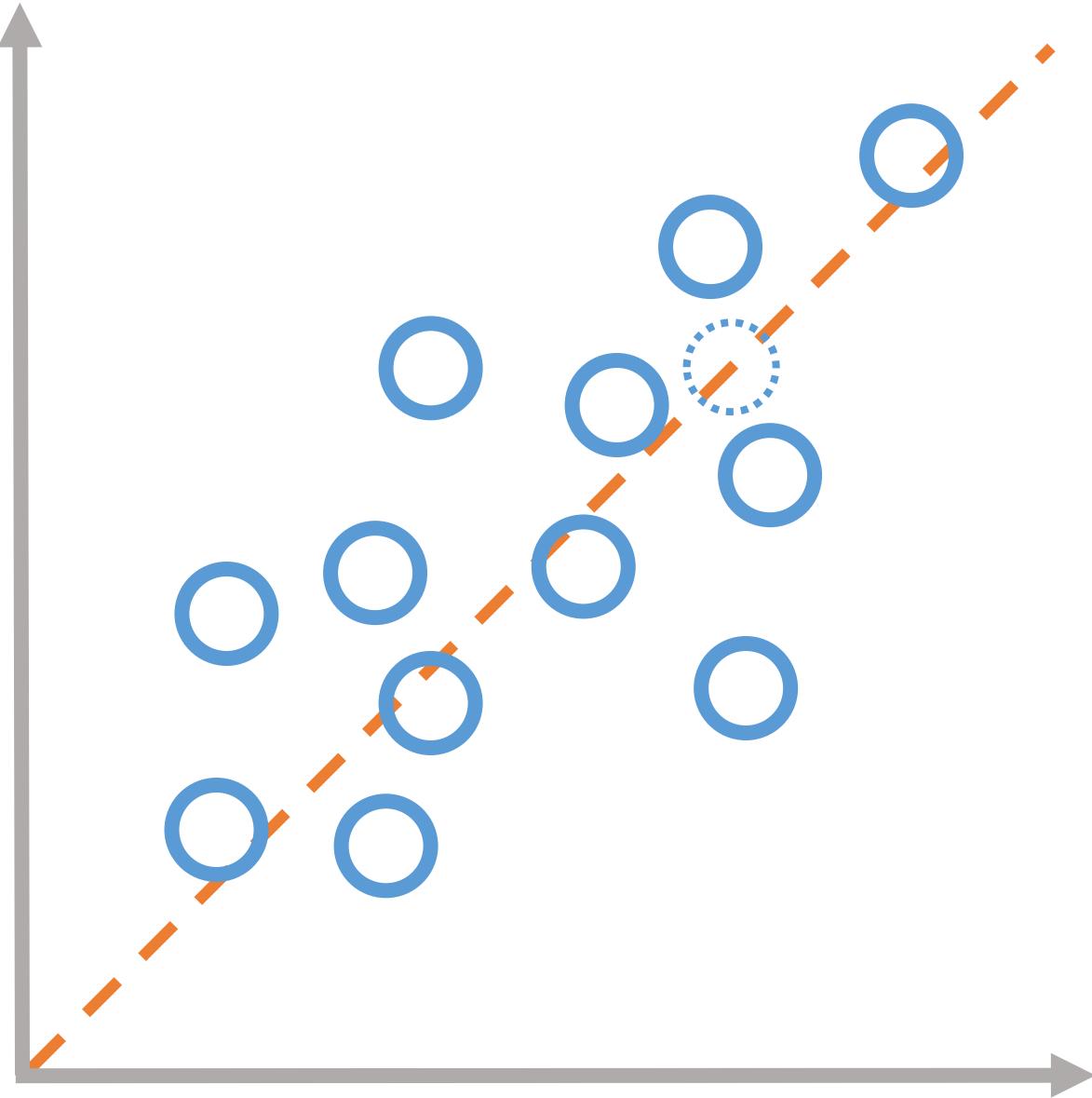


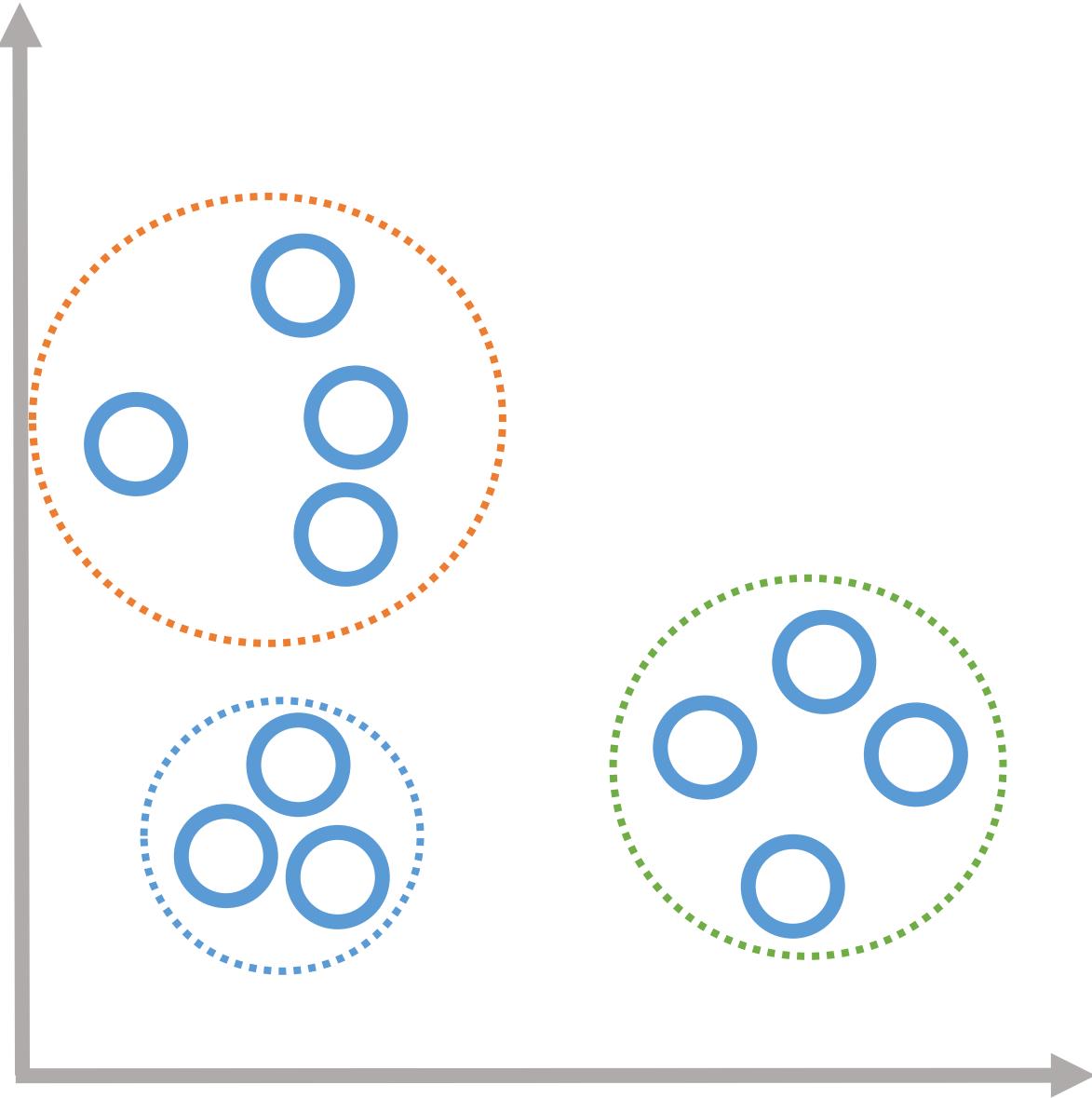
Is cat?

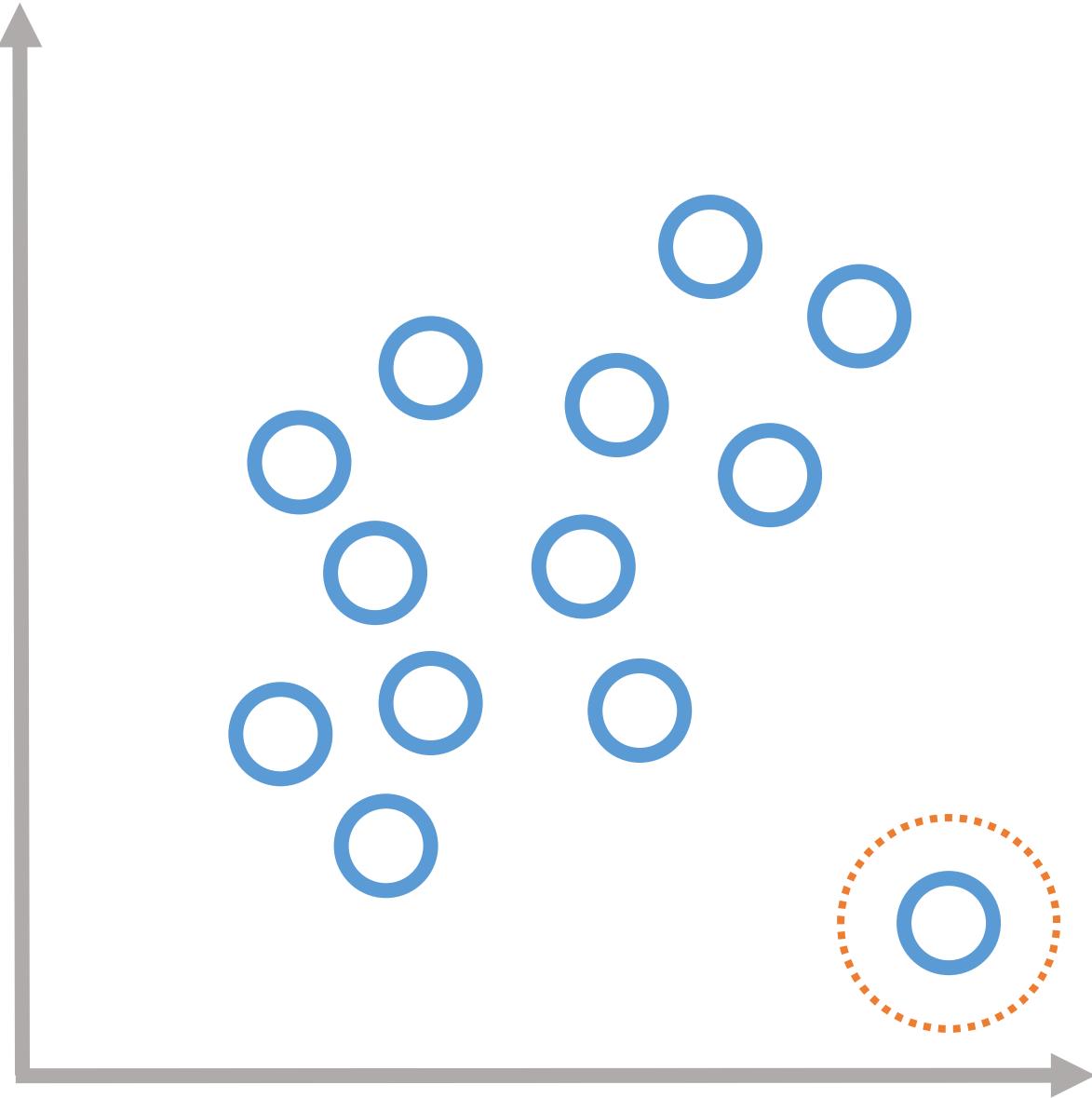


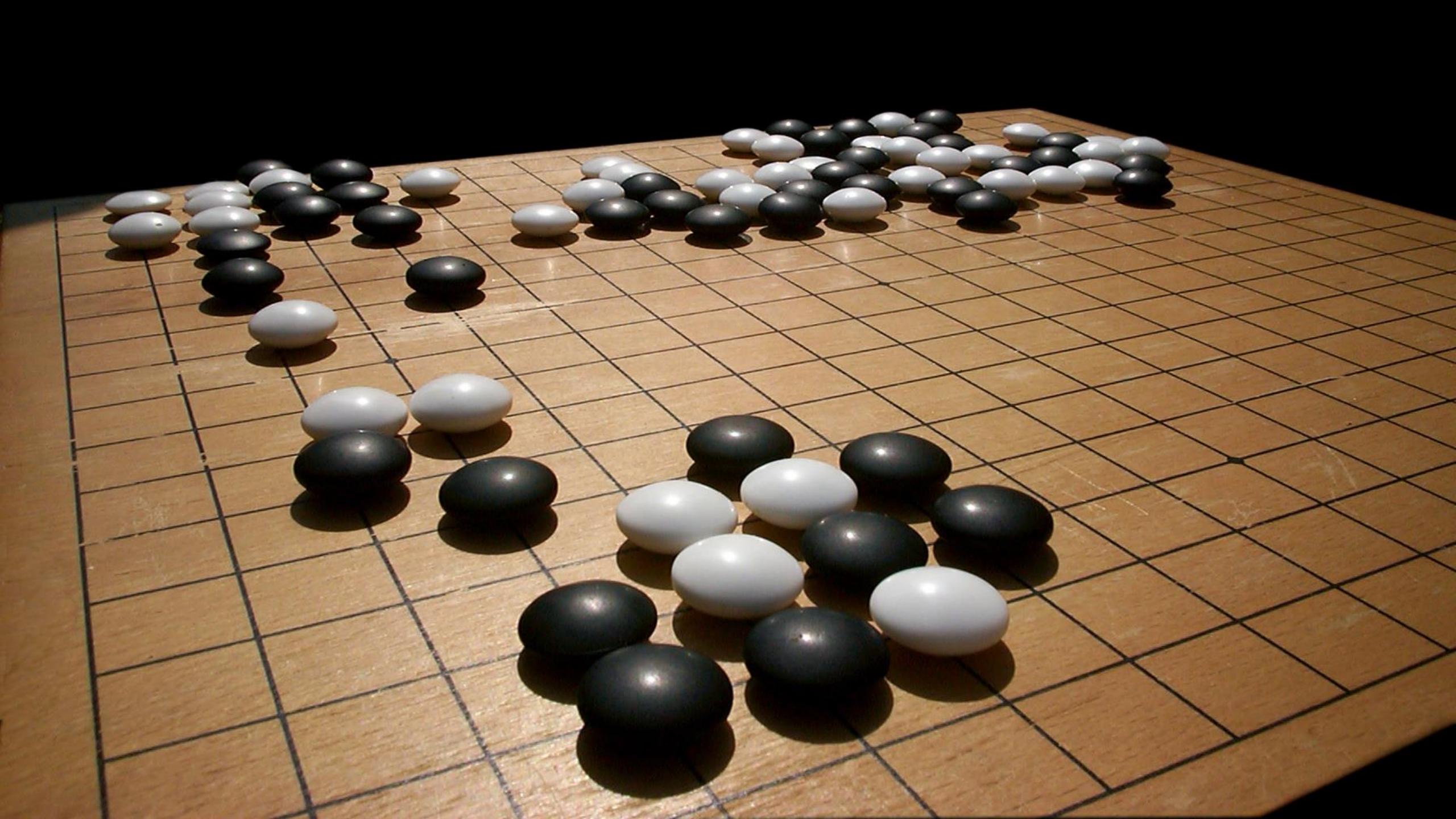
Yes

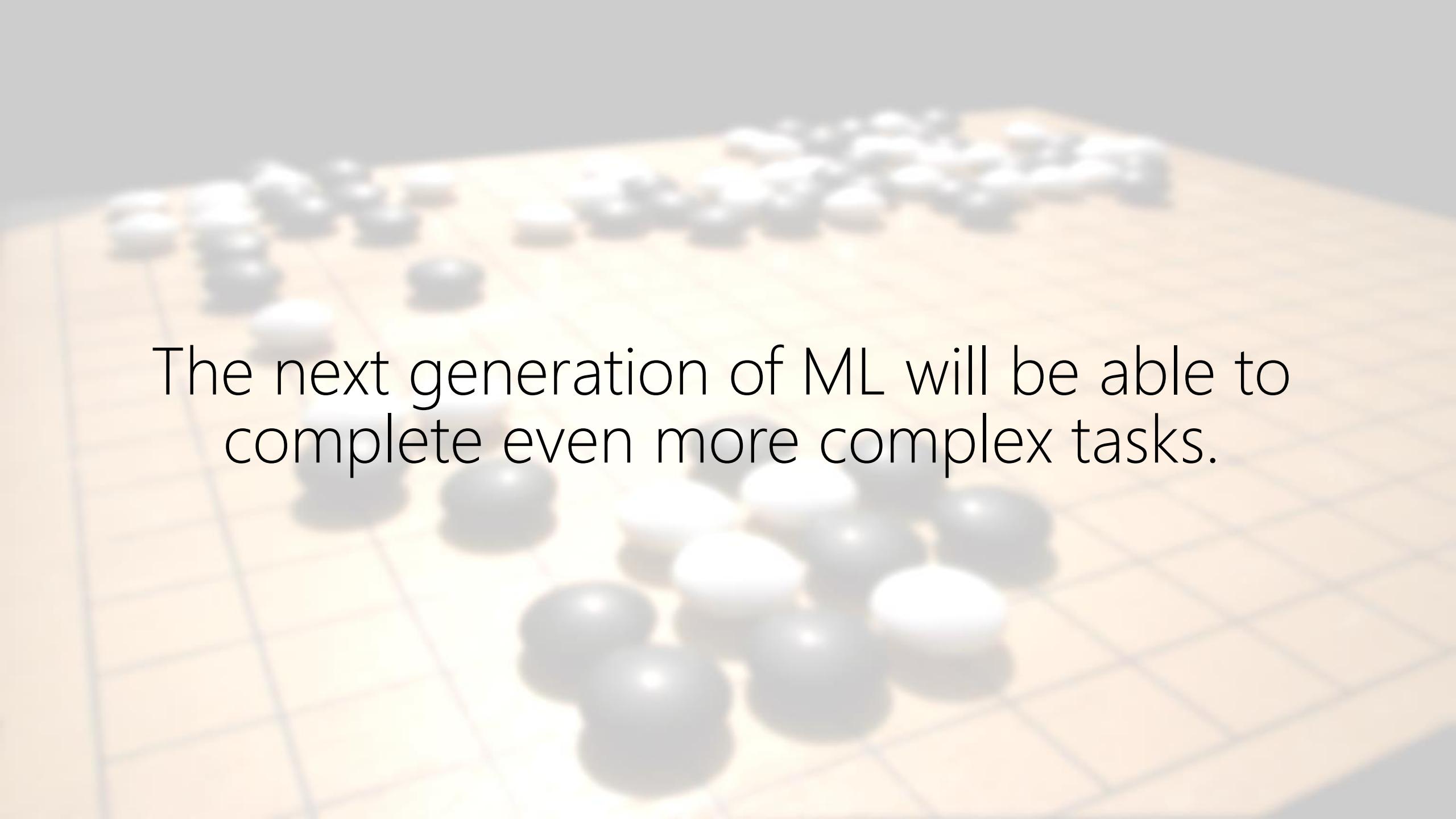








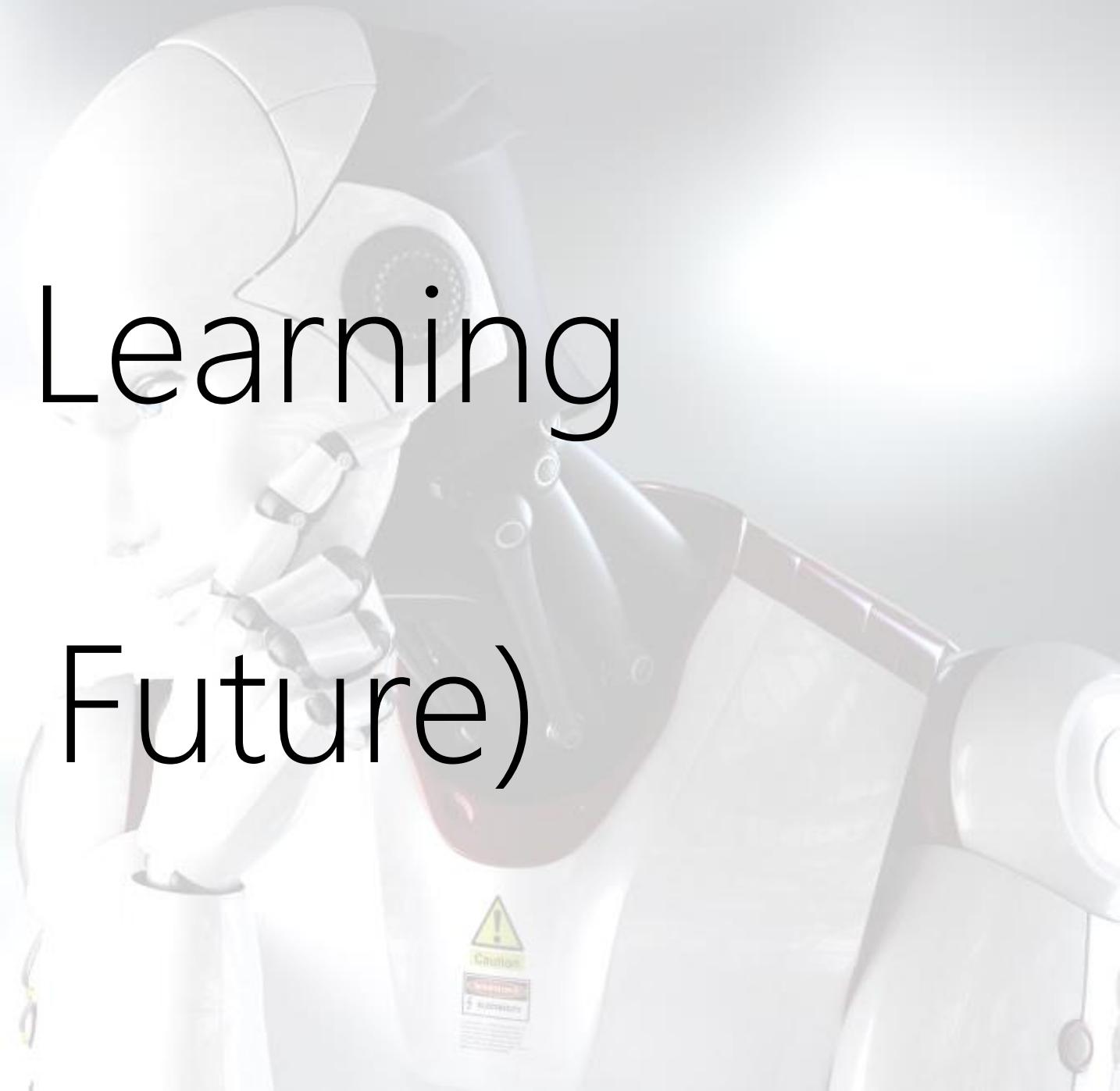




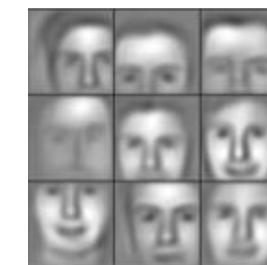
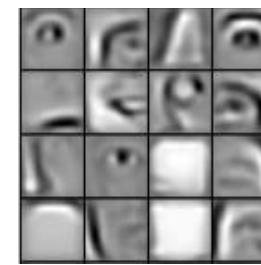
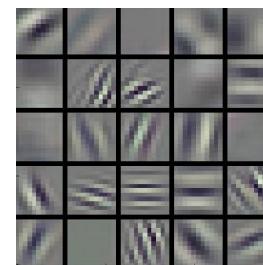
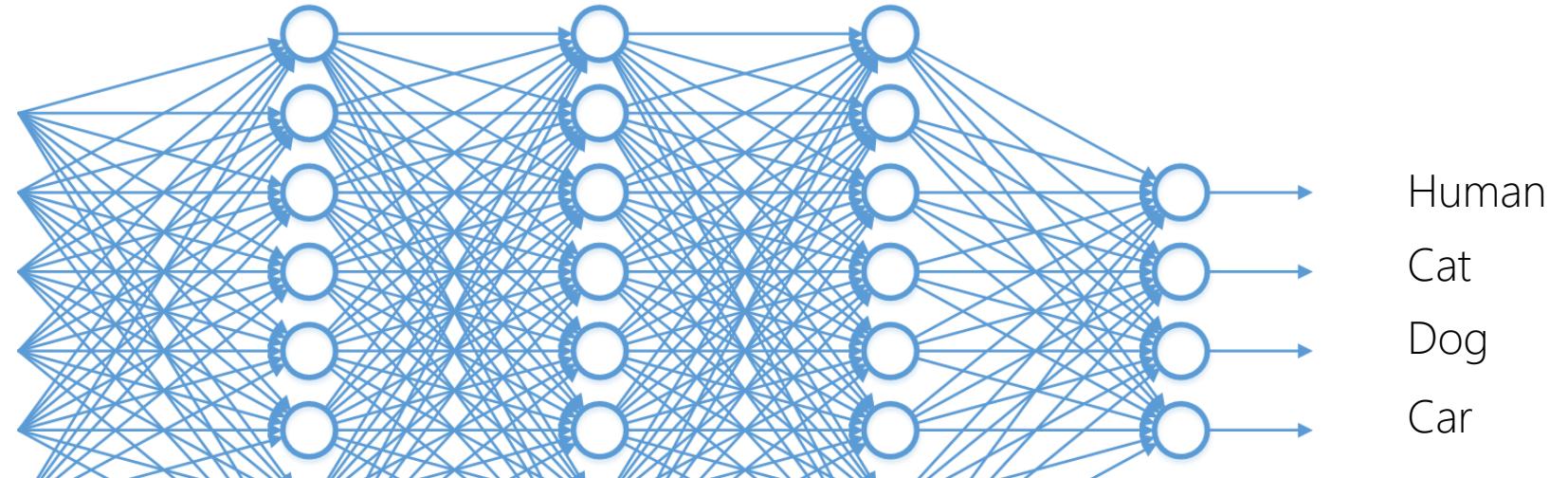
The next generation of ML will be able to complete even more complex tasks.

Deep Learning

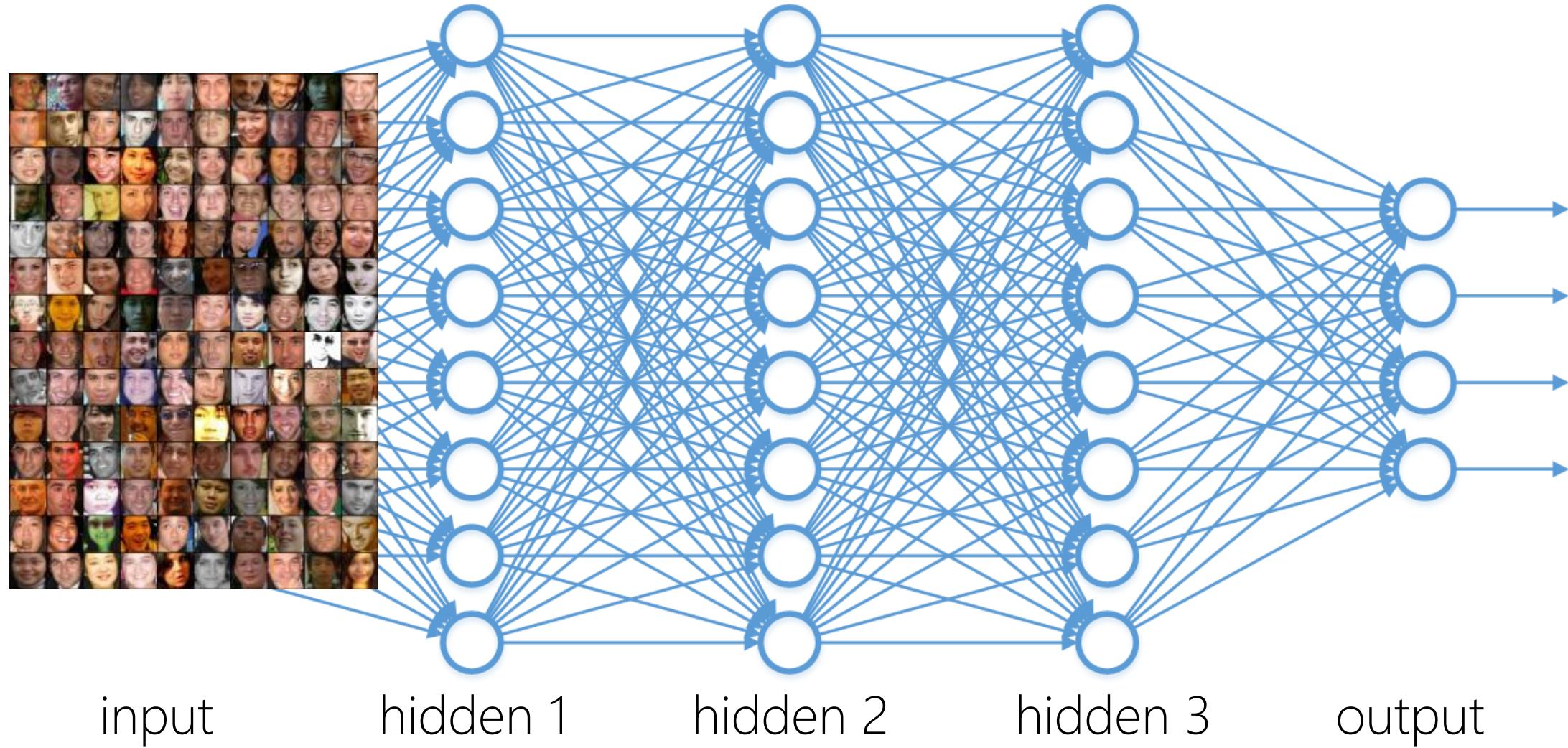
(The Future)



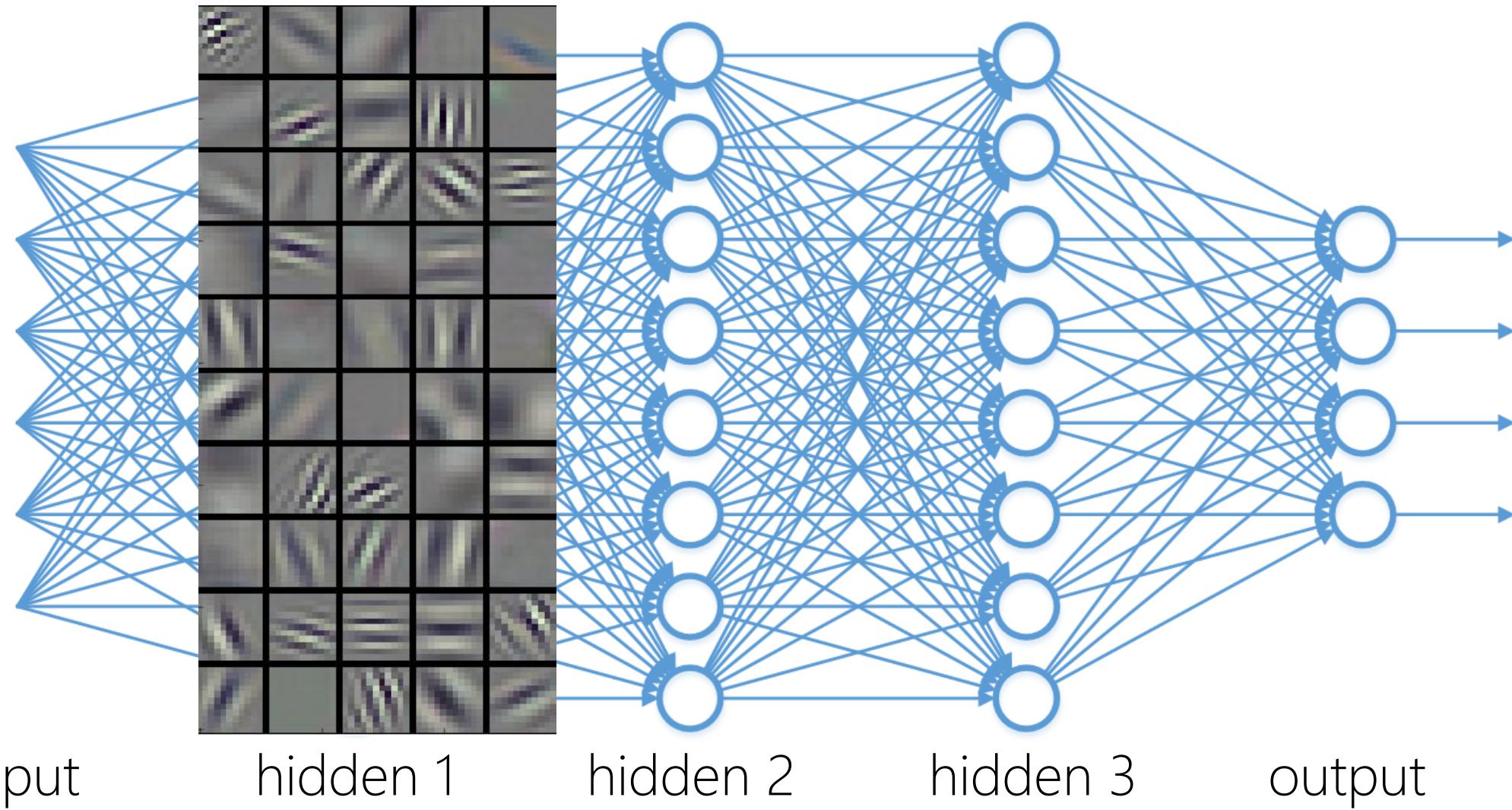
Deep Learning



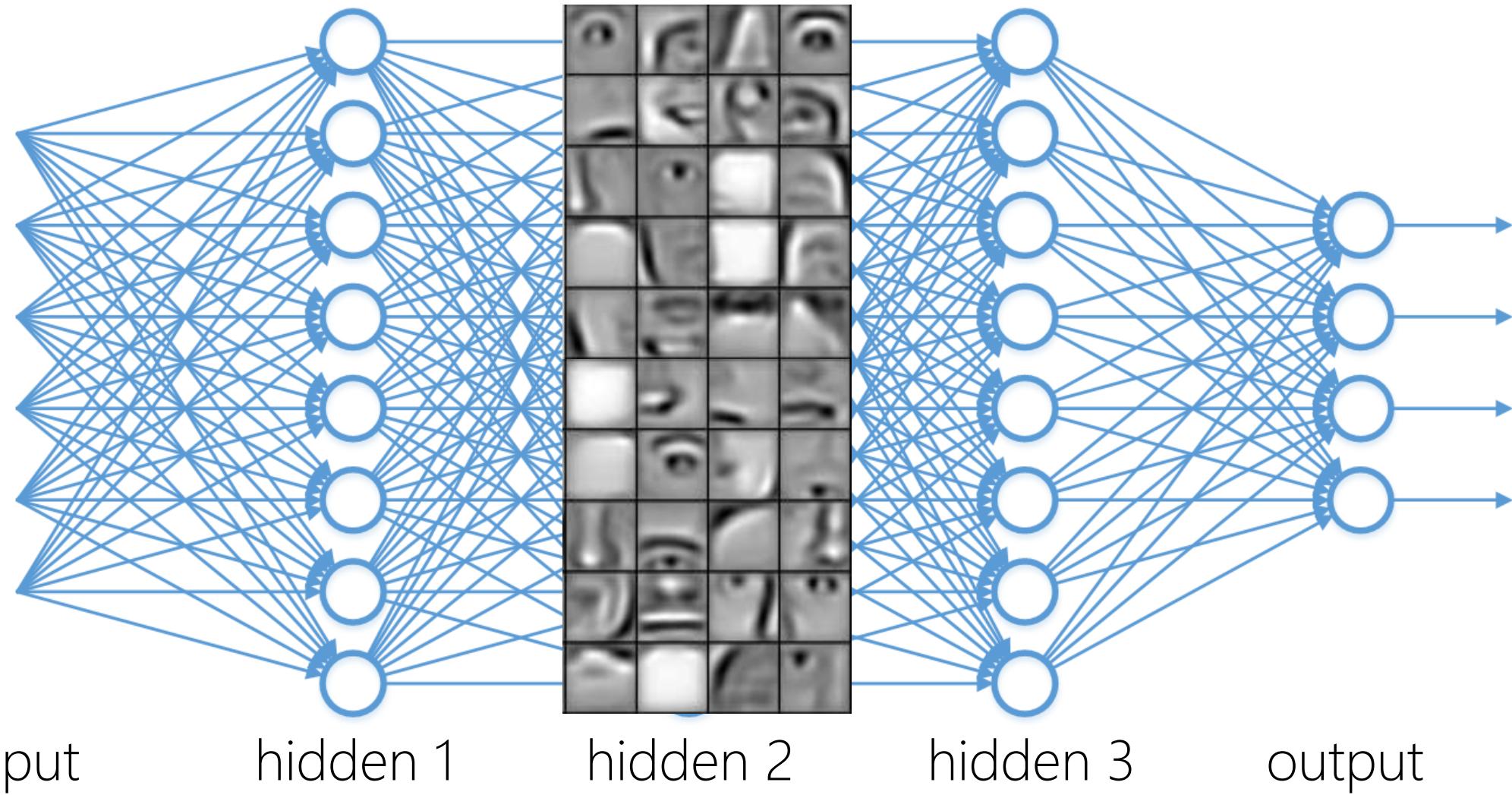
Deep Neural Network



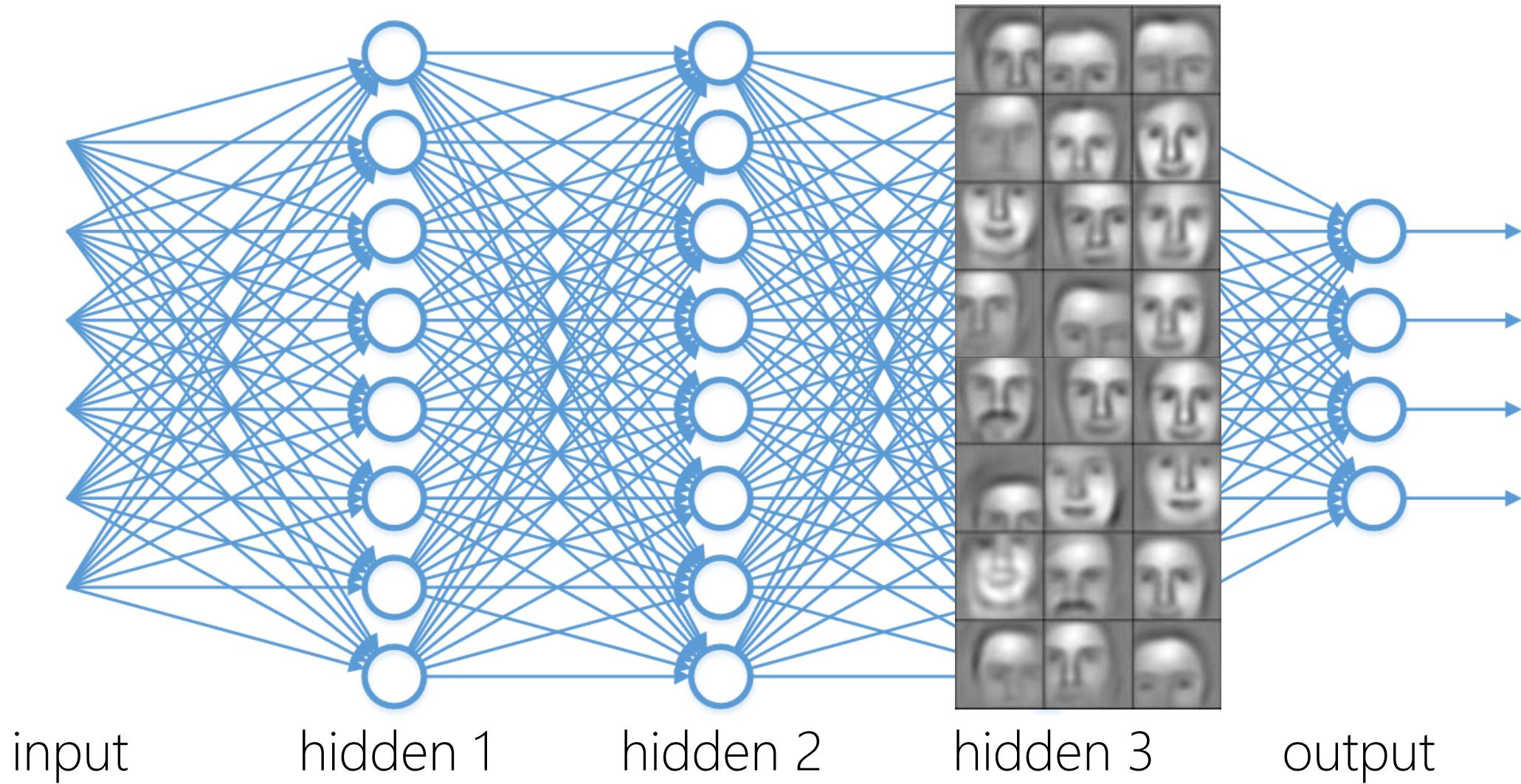
Deep Neural Network



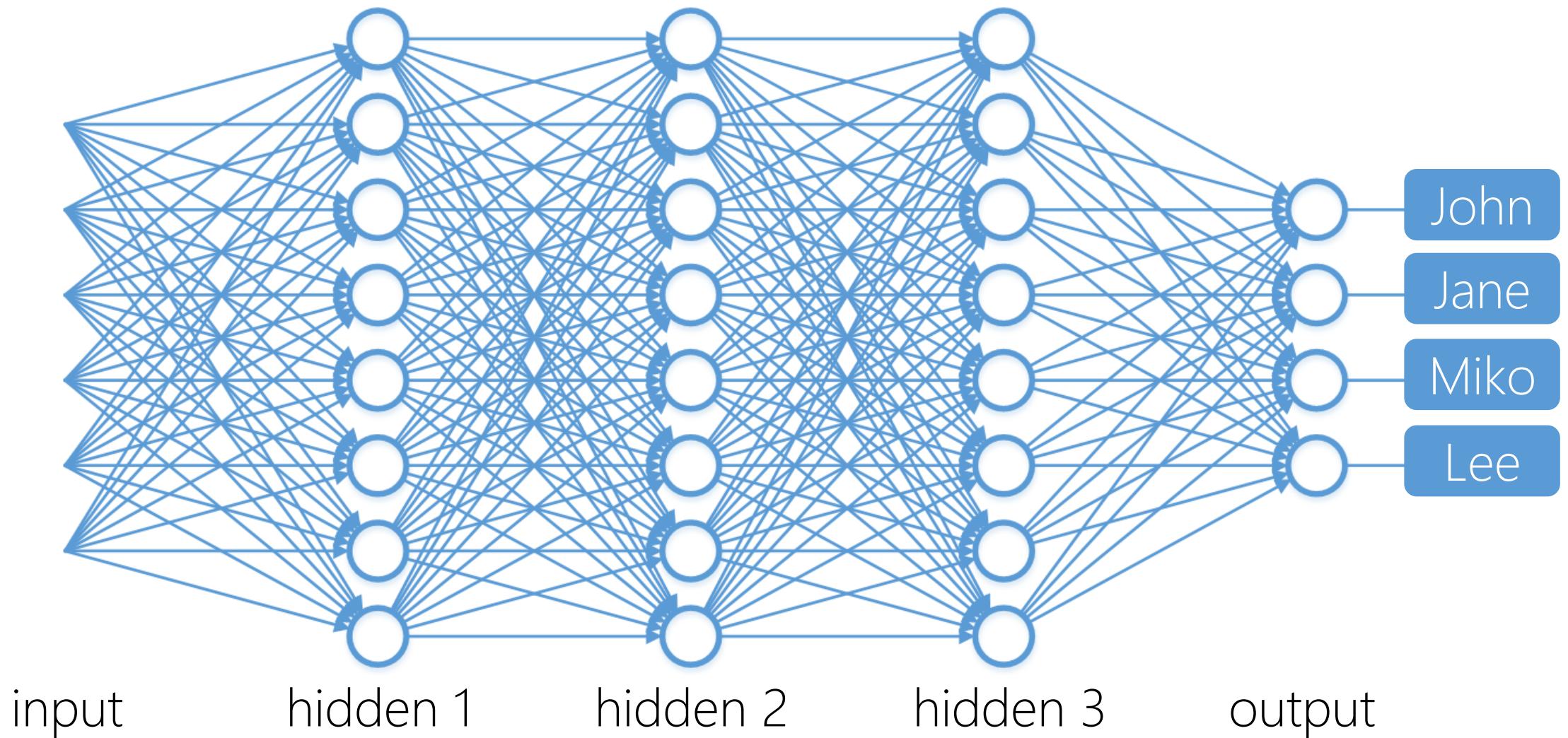
Deep Neural Network

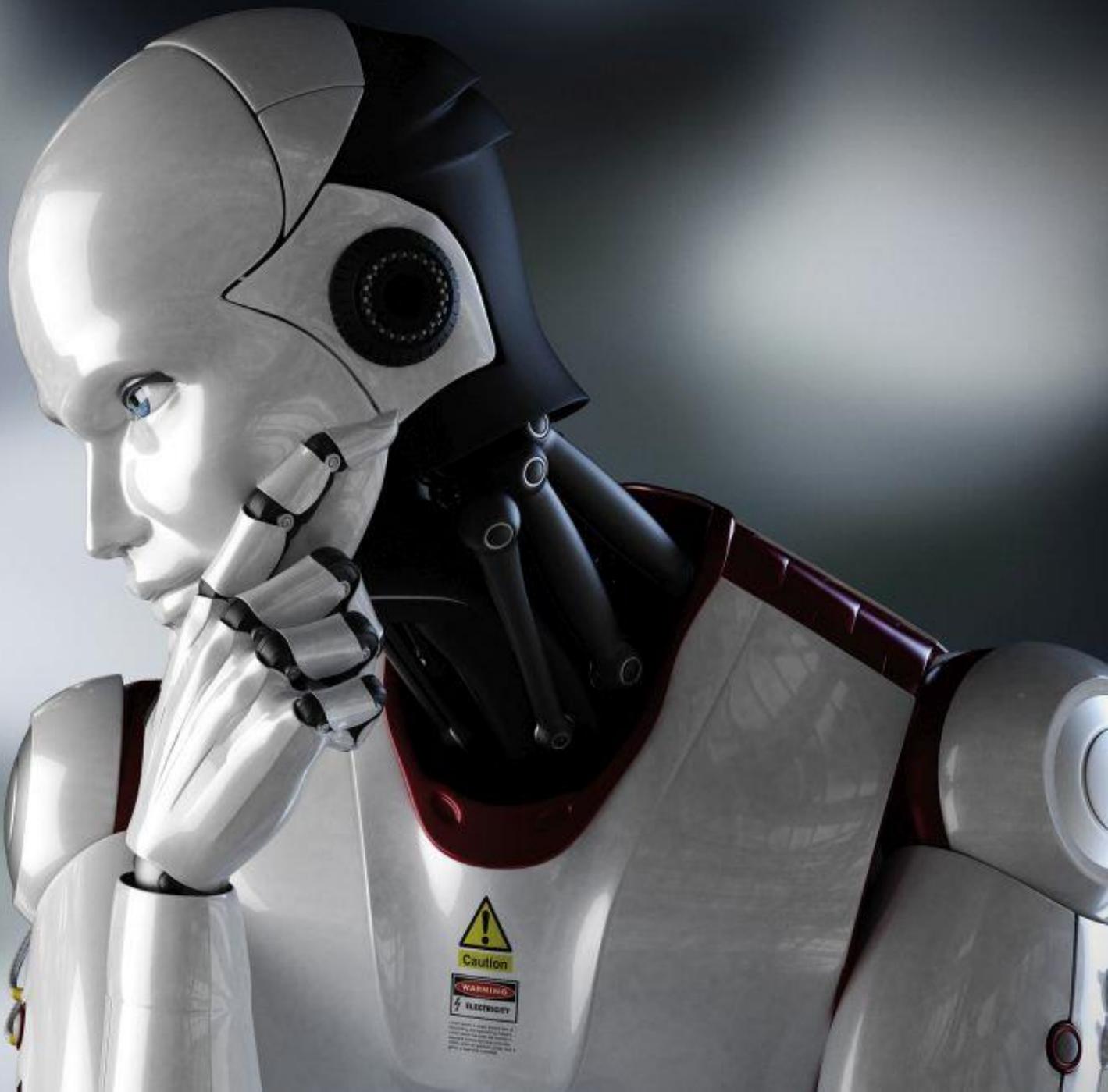


Deep Neural Network

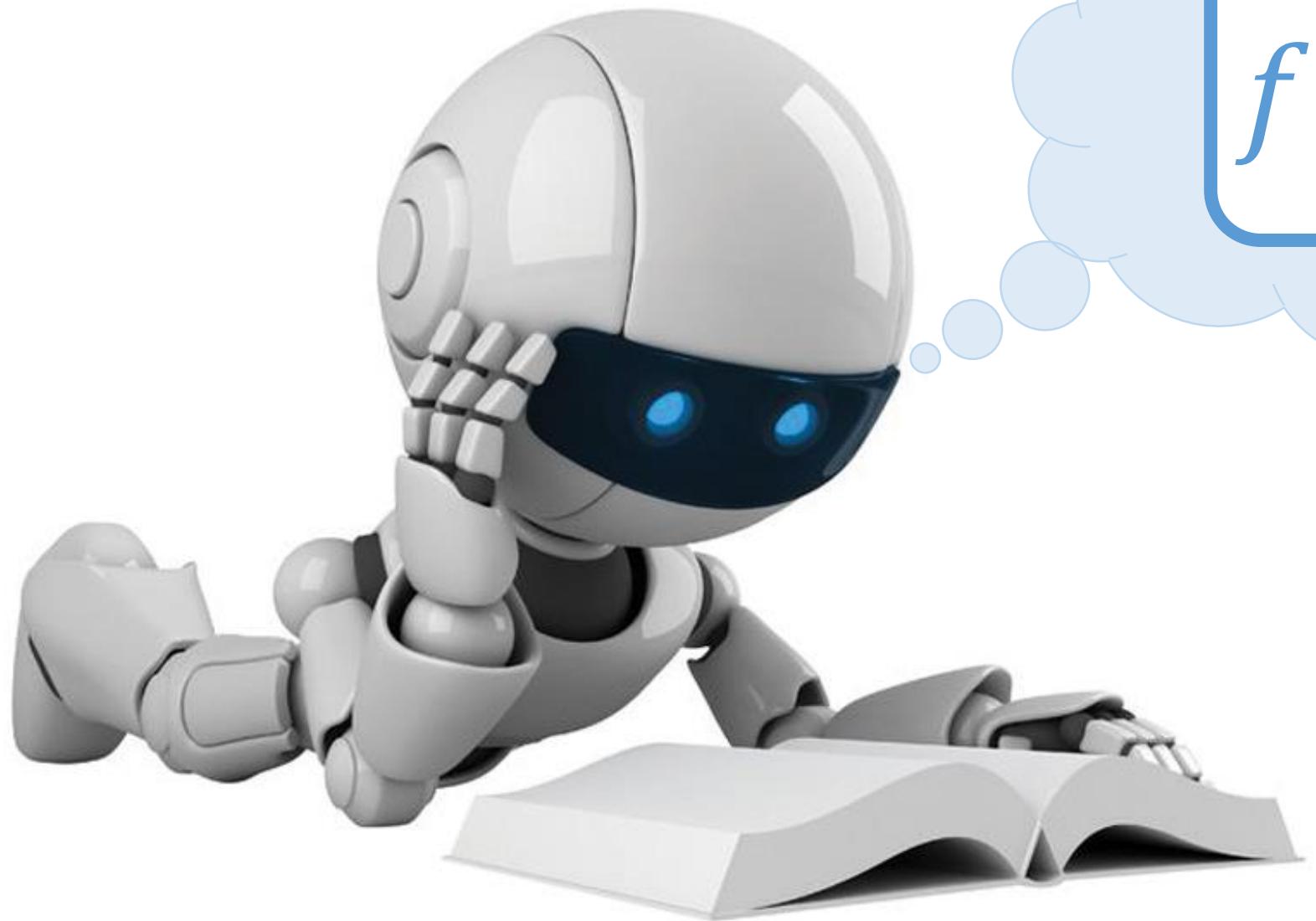


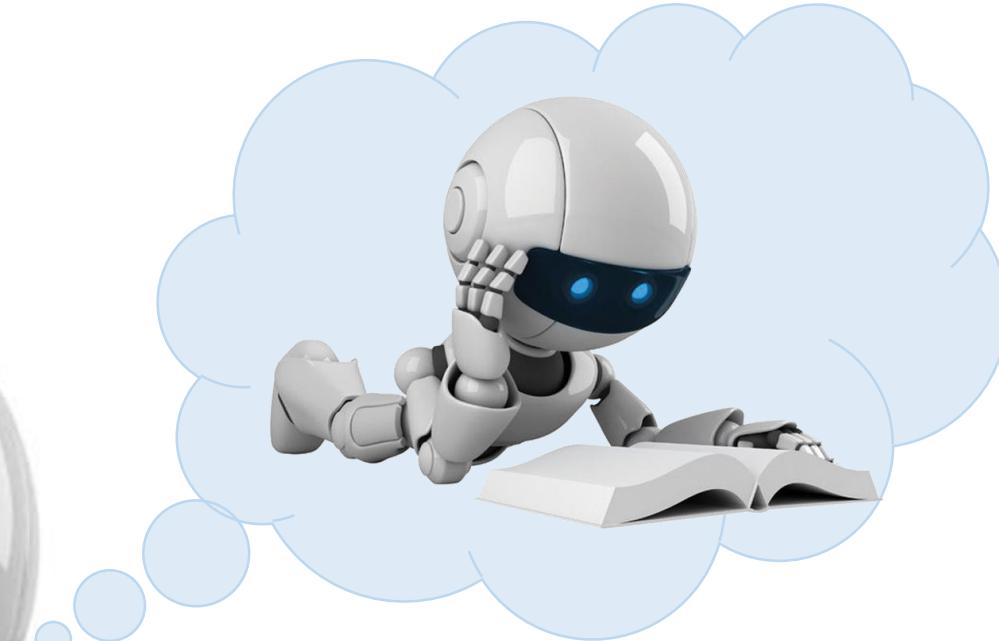
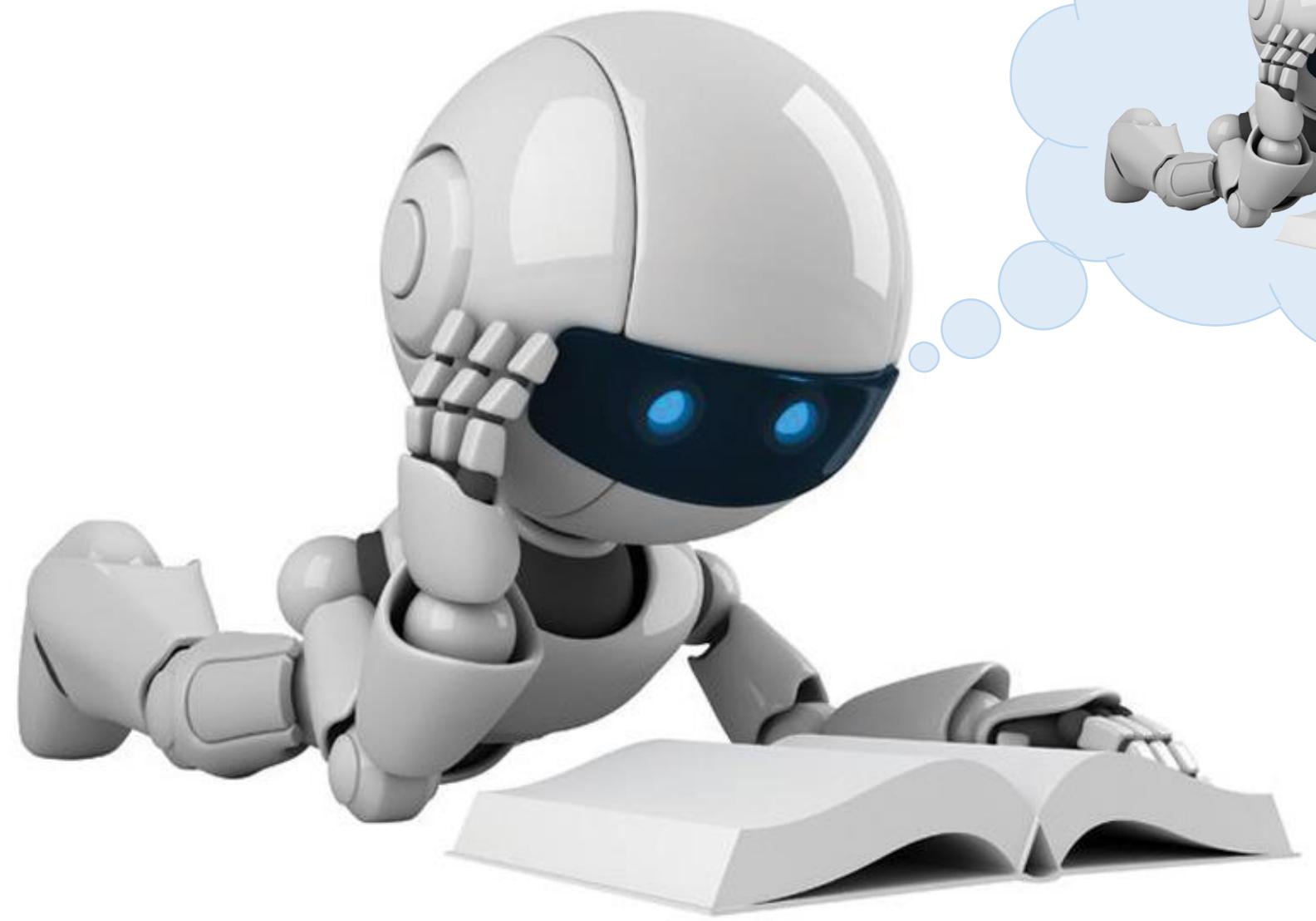
Deep Neural Network







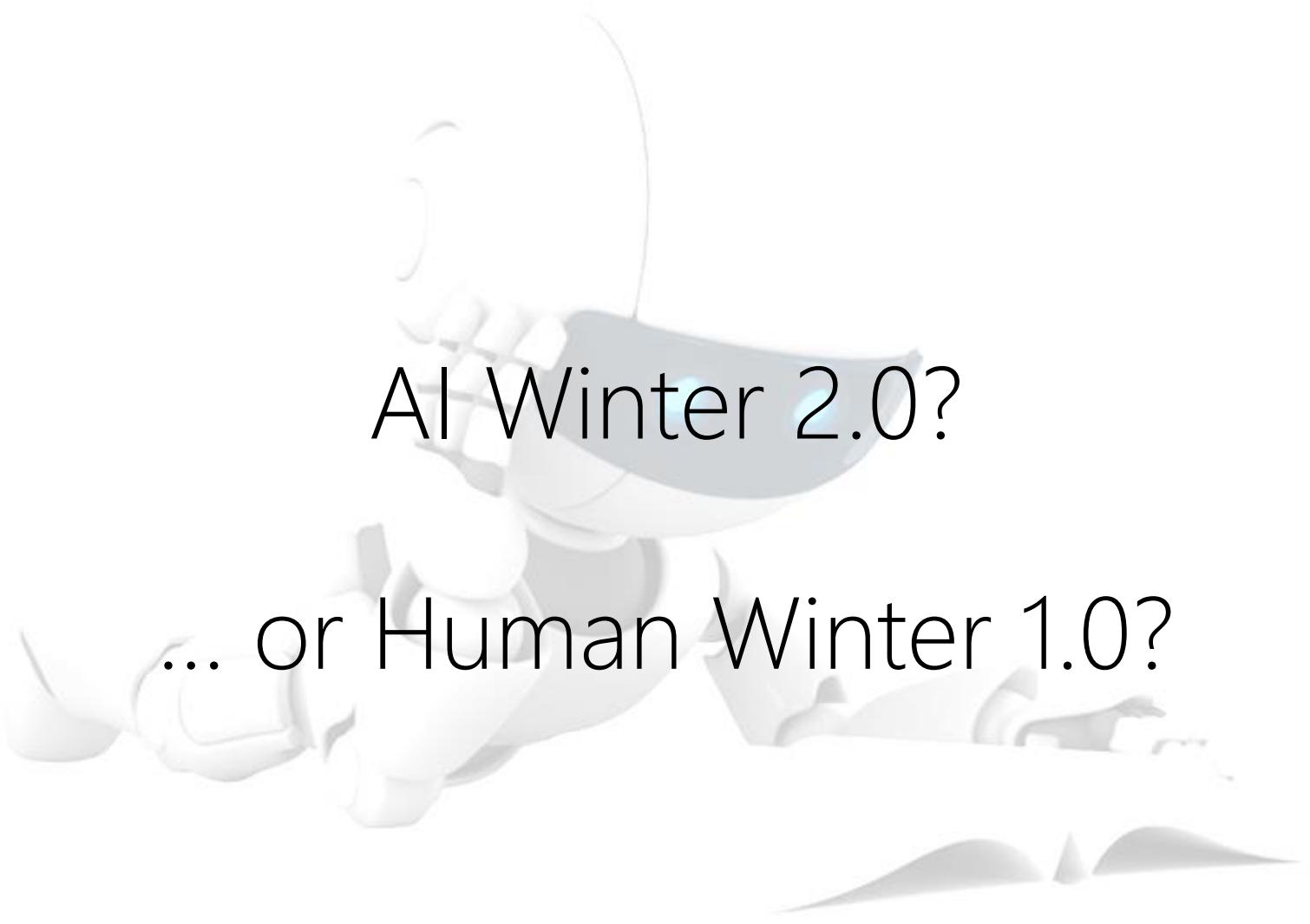

$$f(x)$$







AI Winter 2.0?



AI Winter 2.0?

... or Human Winter 1.0?



FUN GAME 3

Dog or Mop?



MOP!



DOG!

CAUTION



CUIDADO



CAUTION



DOG!



MOP!

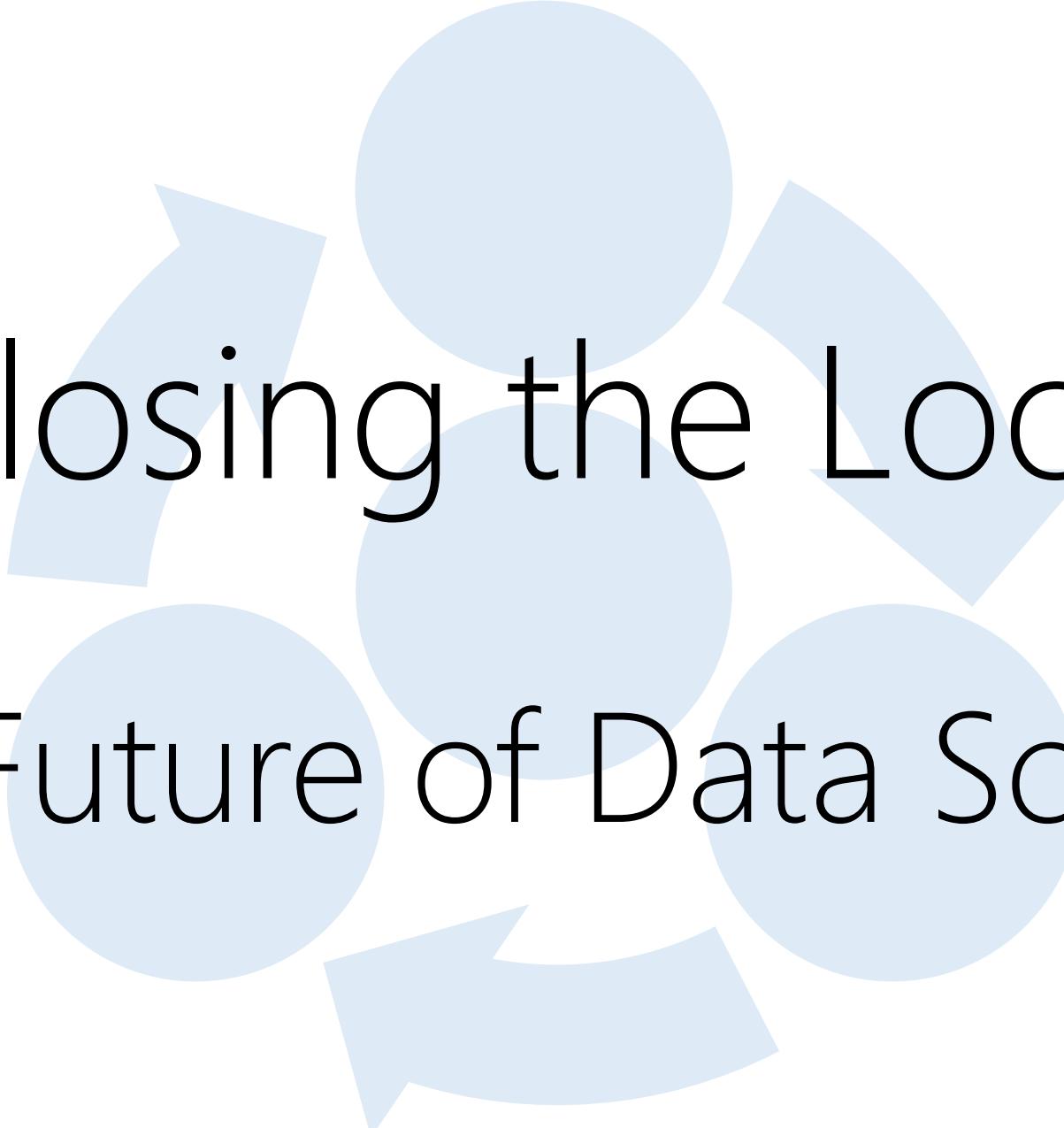
MOP!



DOG!



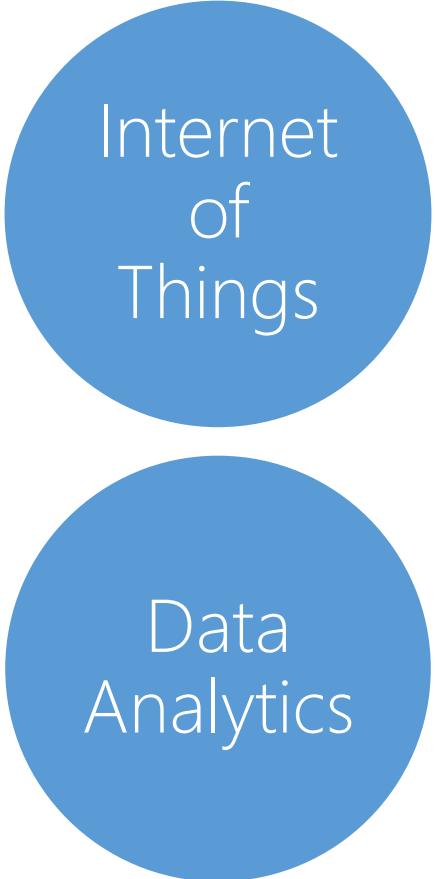




Closing the Loop

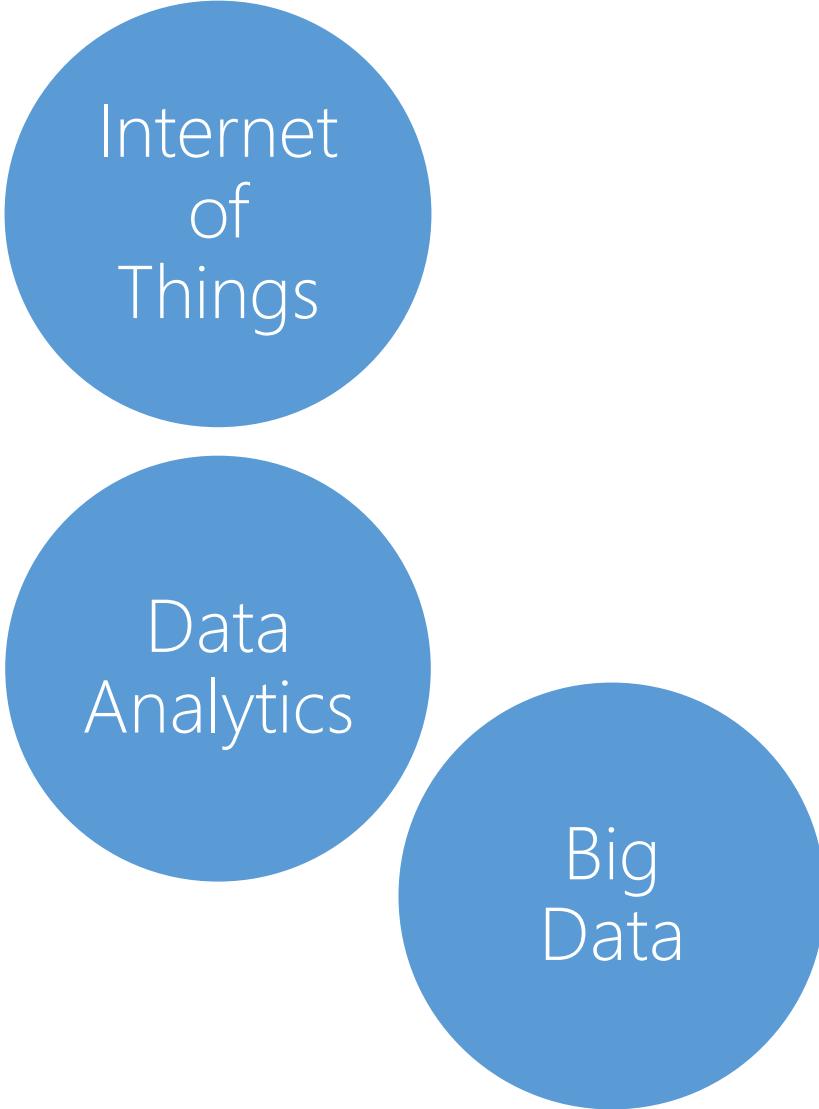
(The Future of Data Science)

Data
Analytics



Internet
of
Things

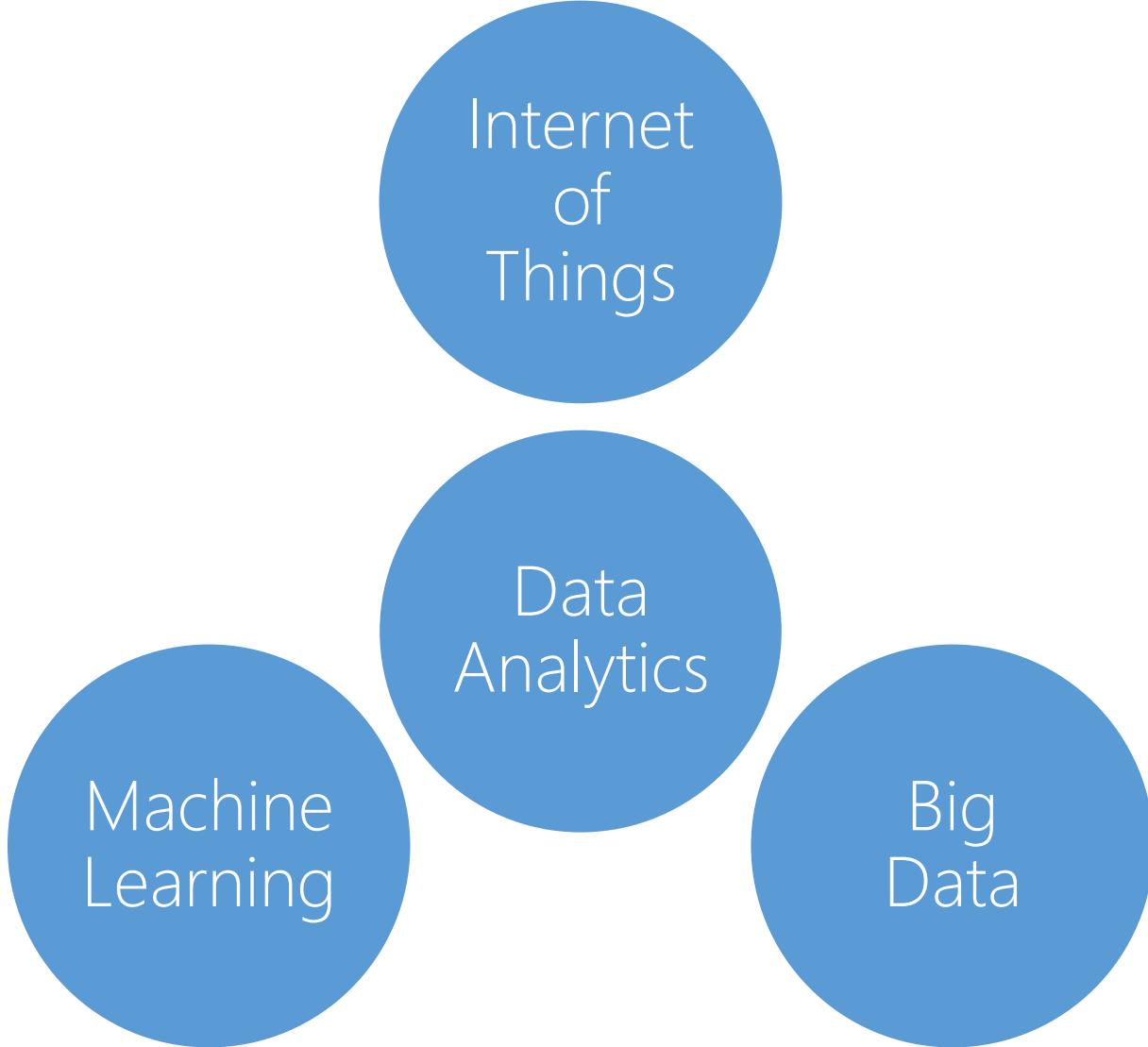
Data
Analytics



Internet
of
Things

Data
Analytics

Big
Data

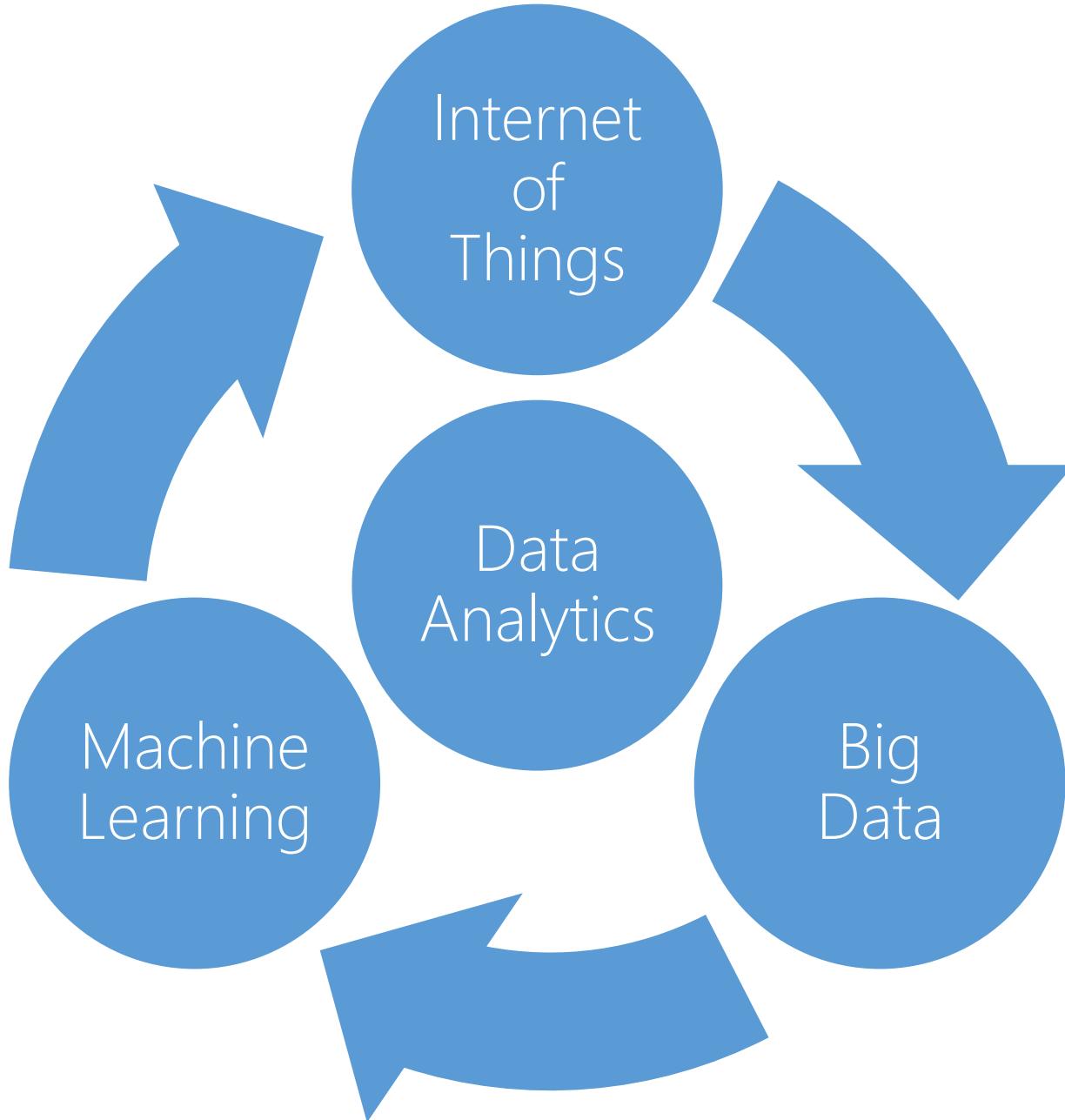


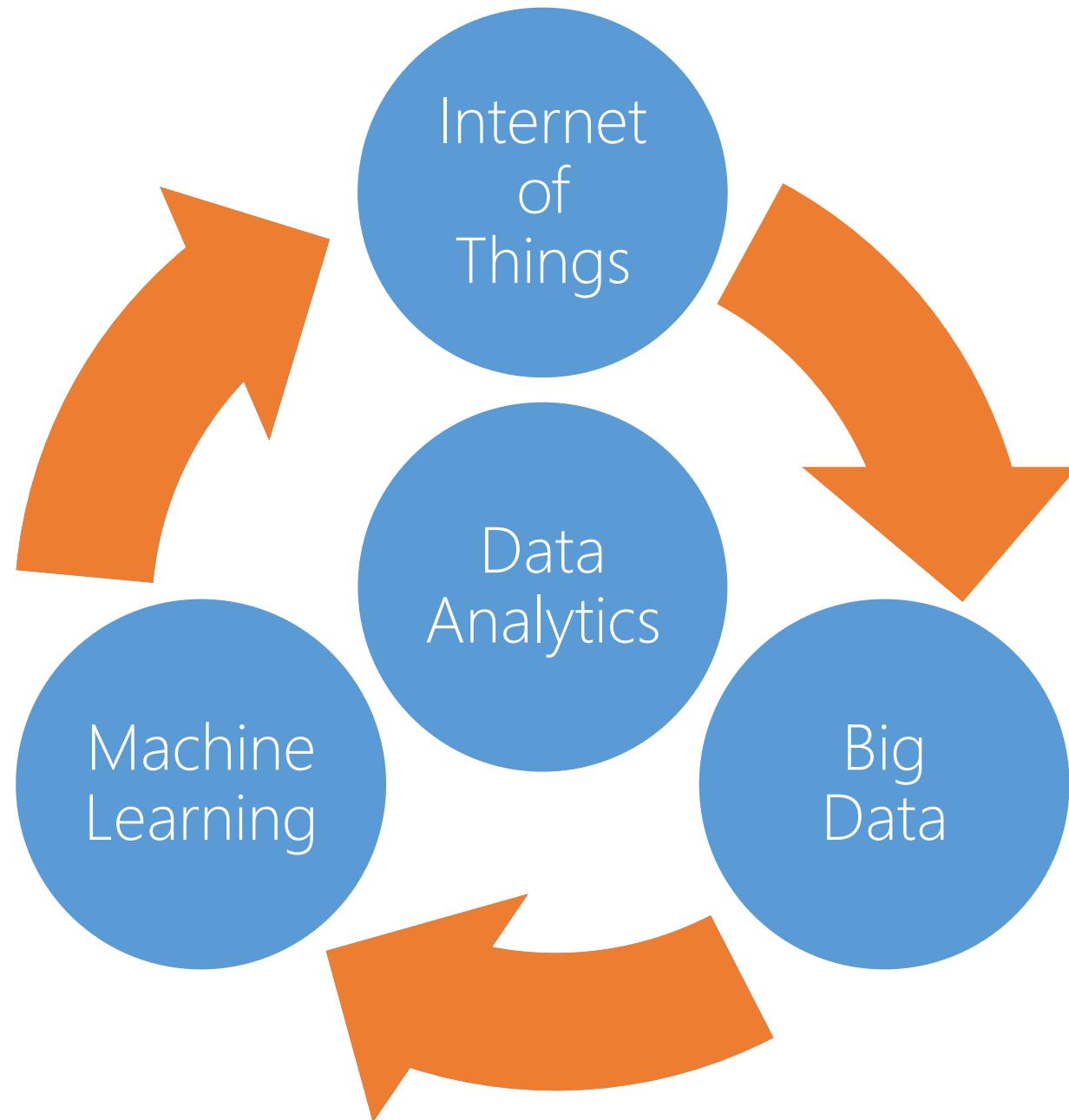
Internet
of
Things

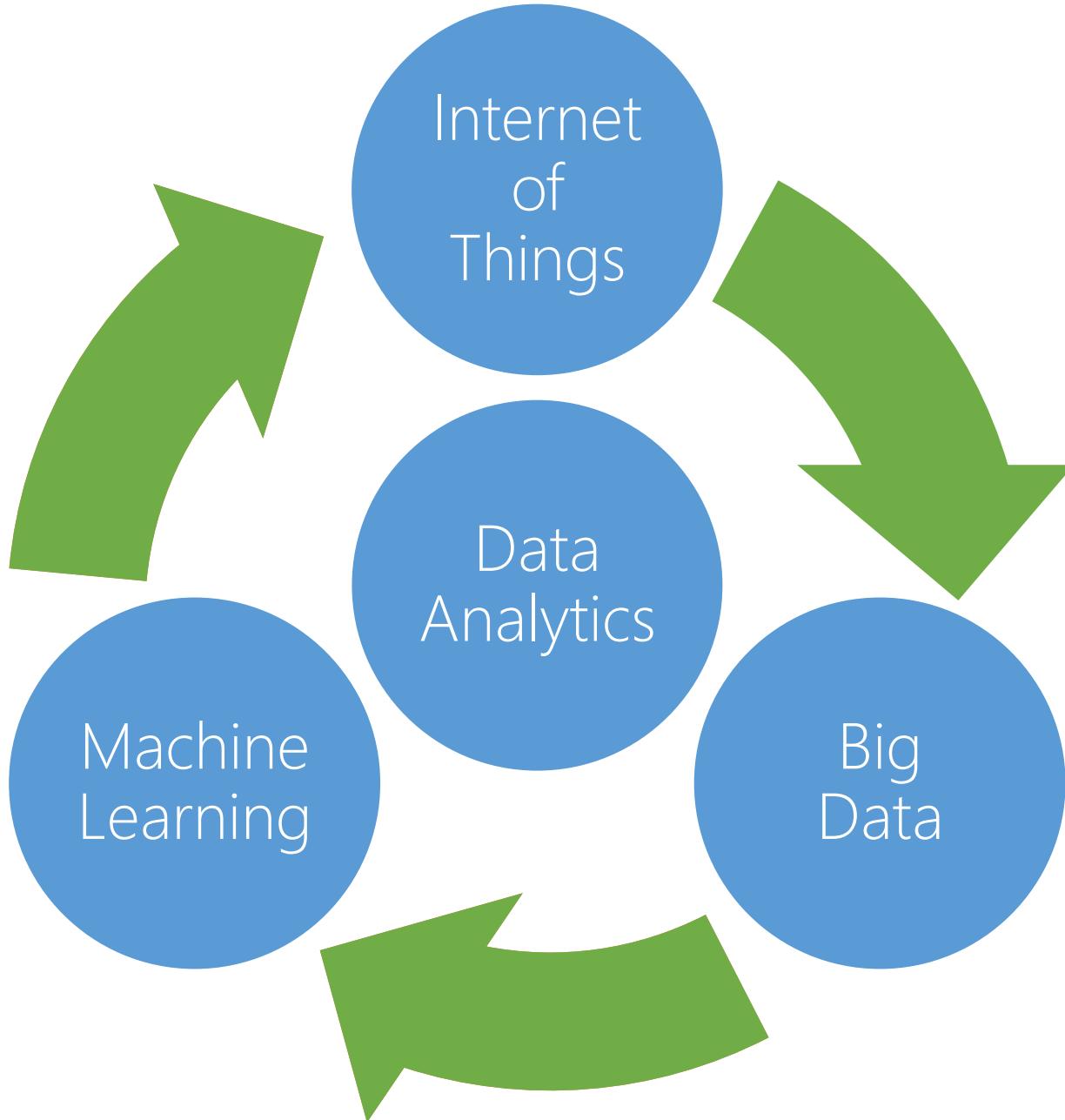
Data
Analytics

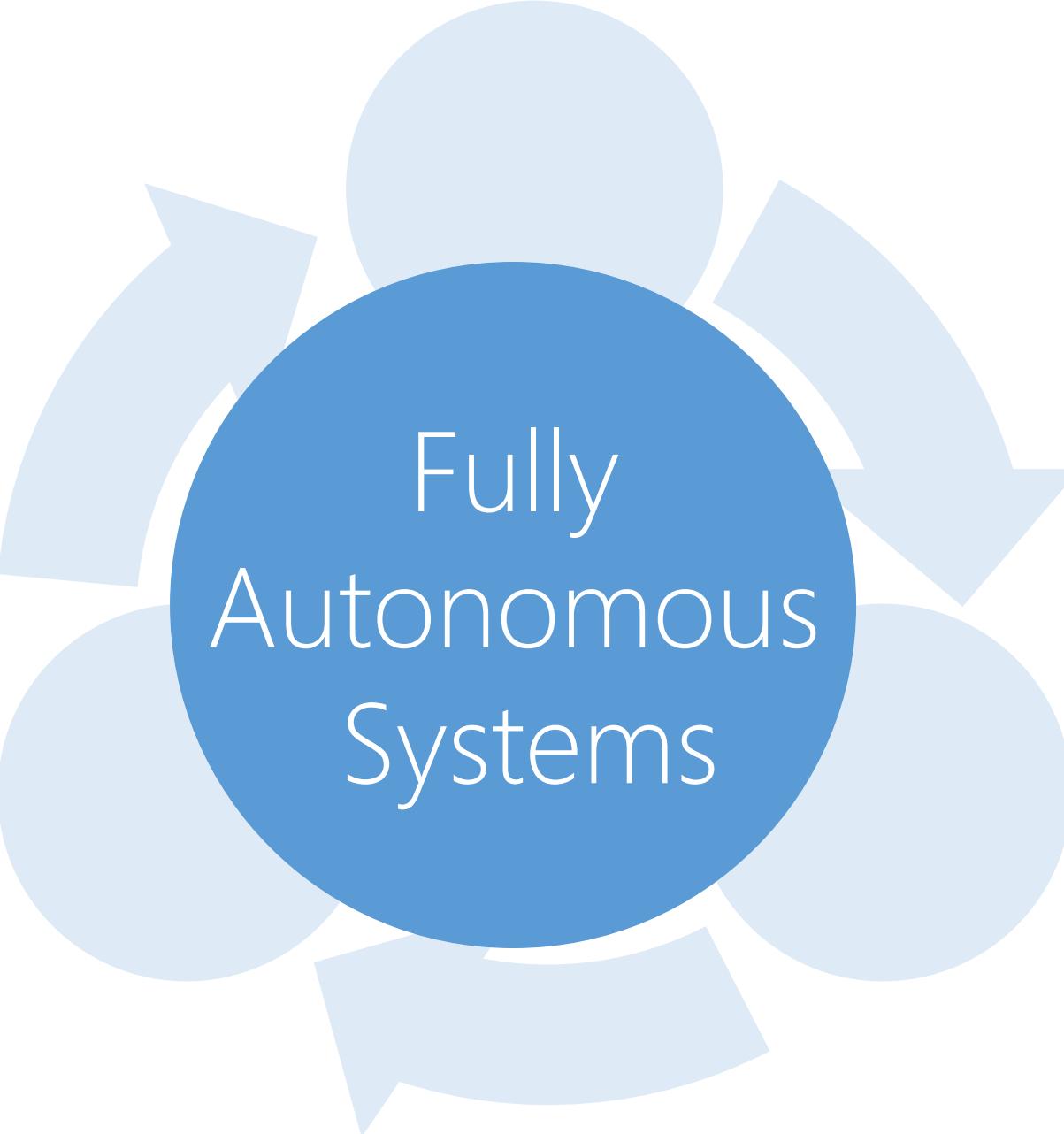
Machine
Learning

Big
Data

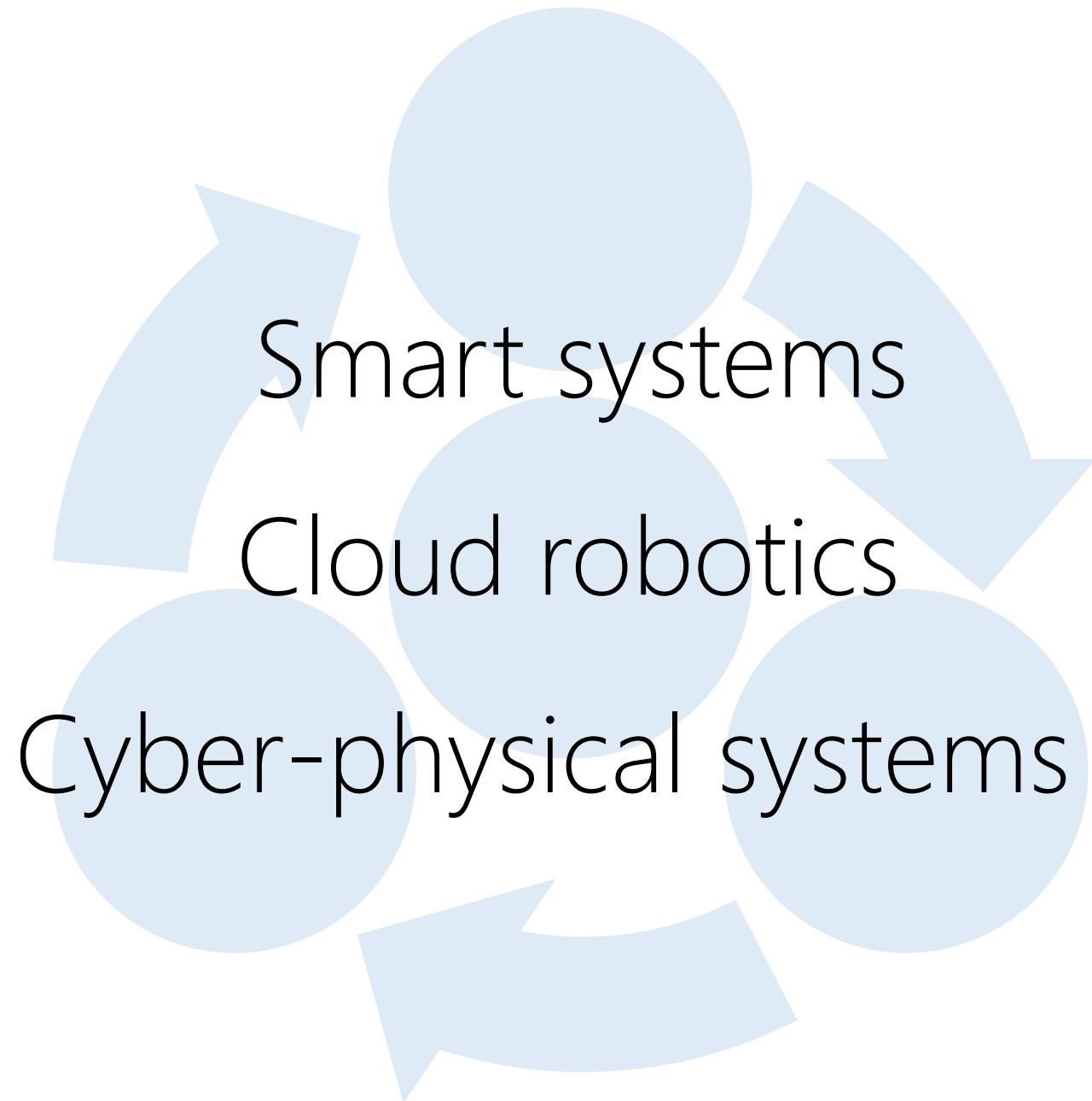


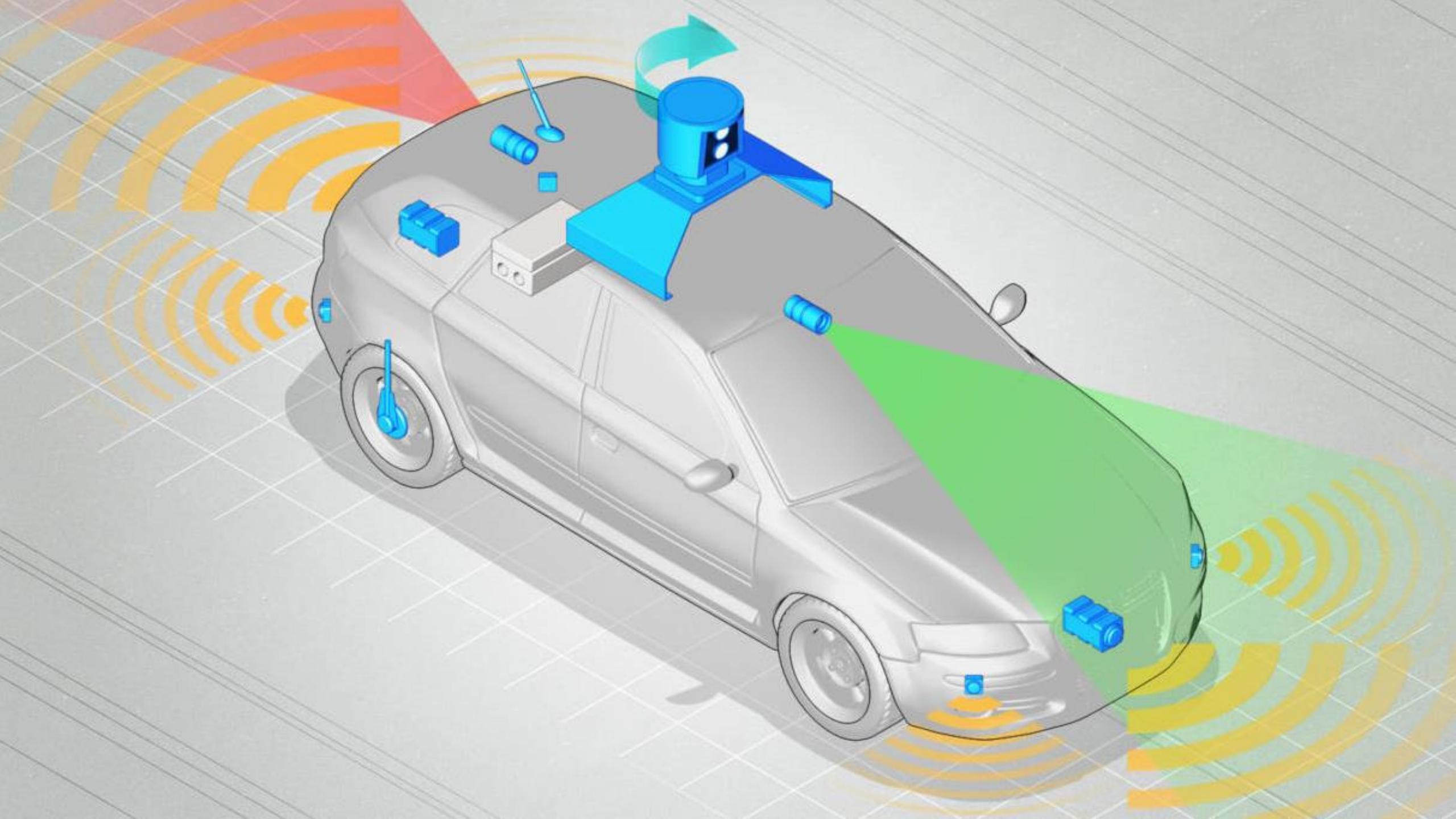




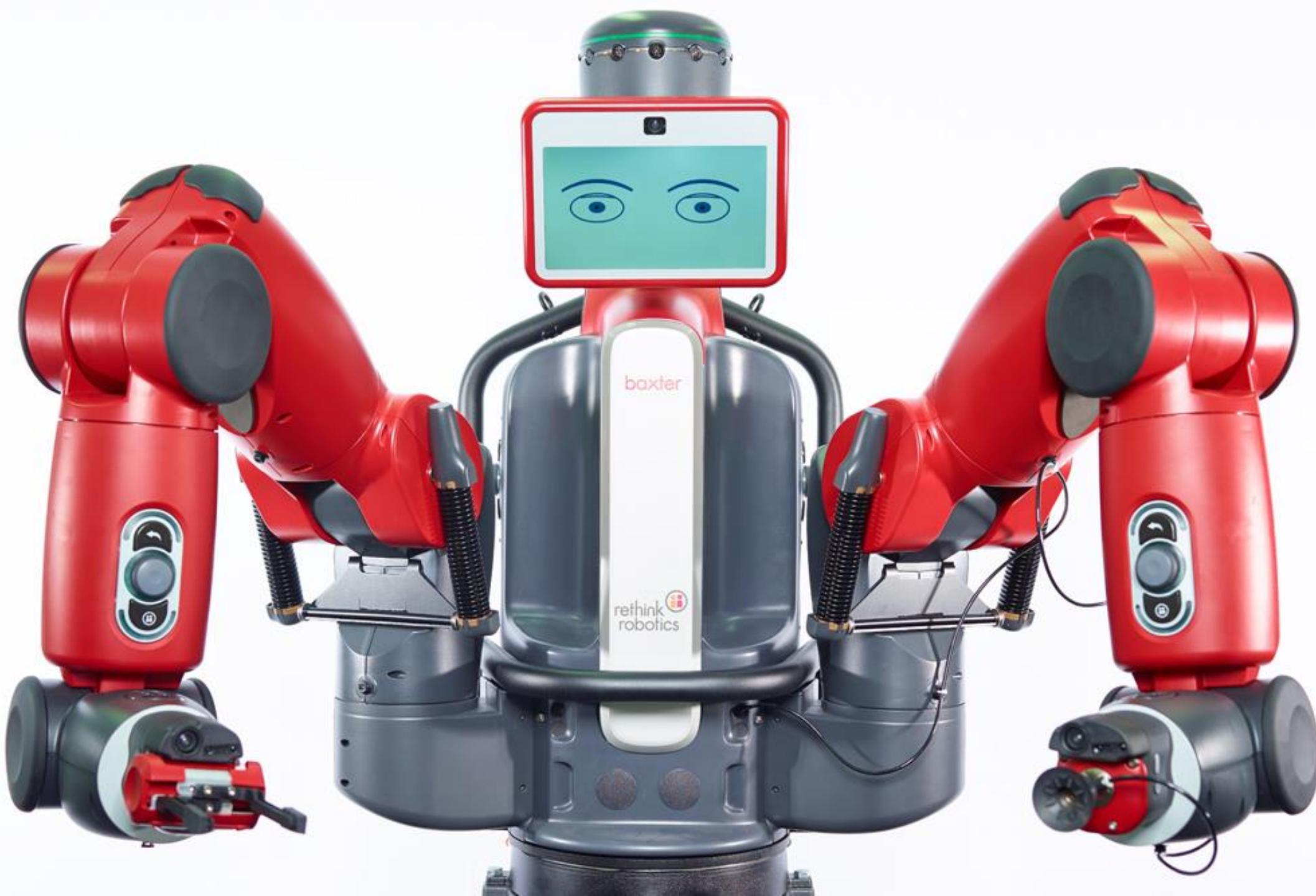


Fully
Autonomous
Systems











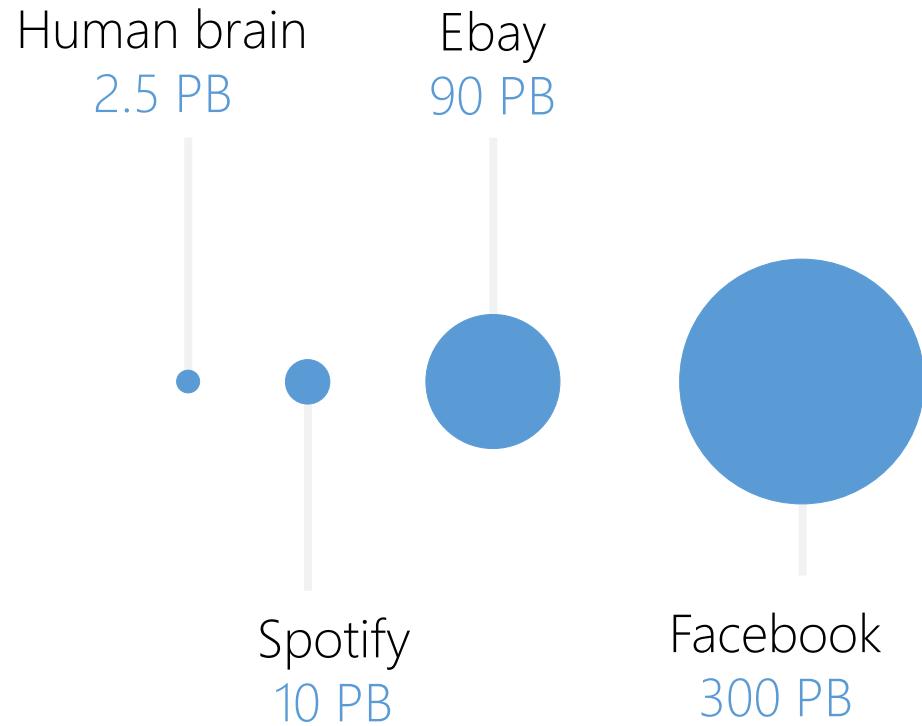


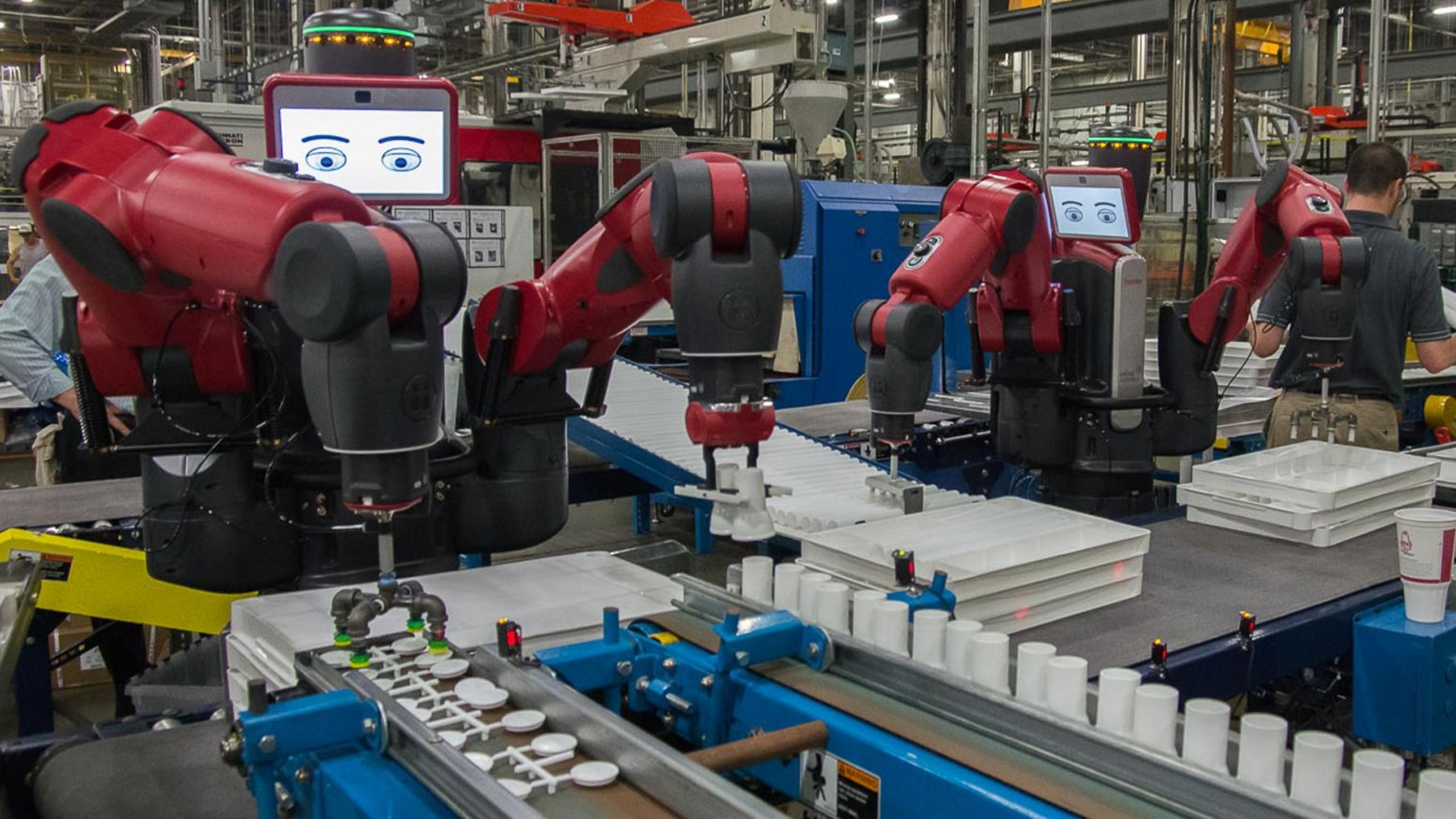
Embedded intelligence will be woven
into the fabric of our society.



The Big Data Universe

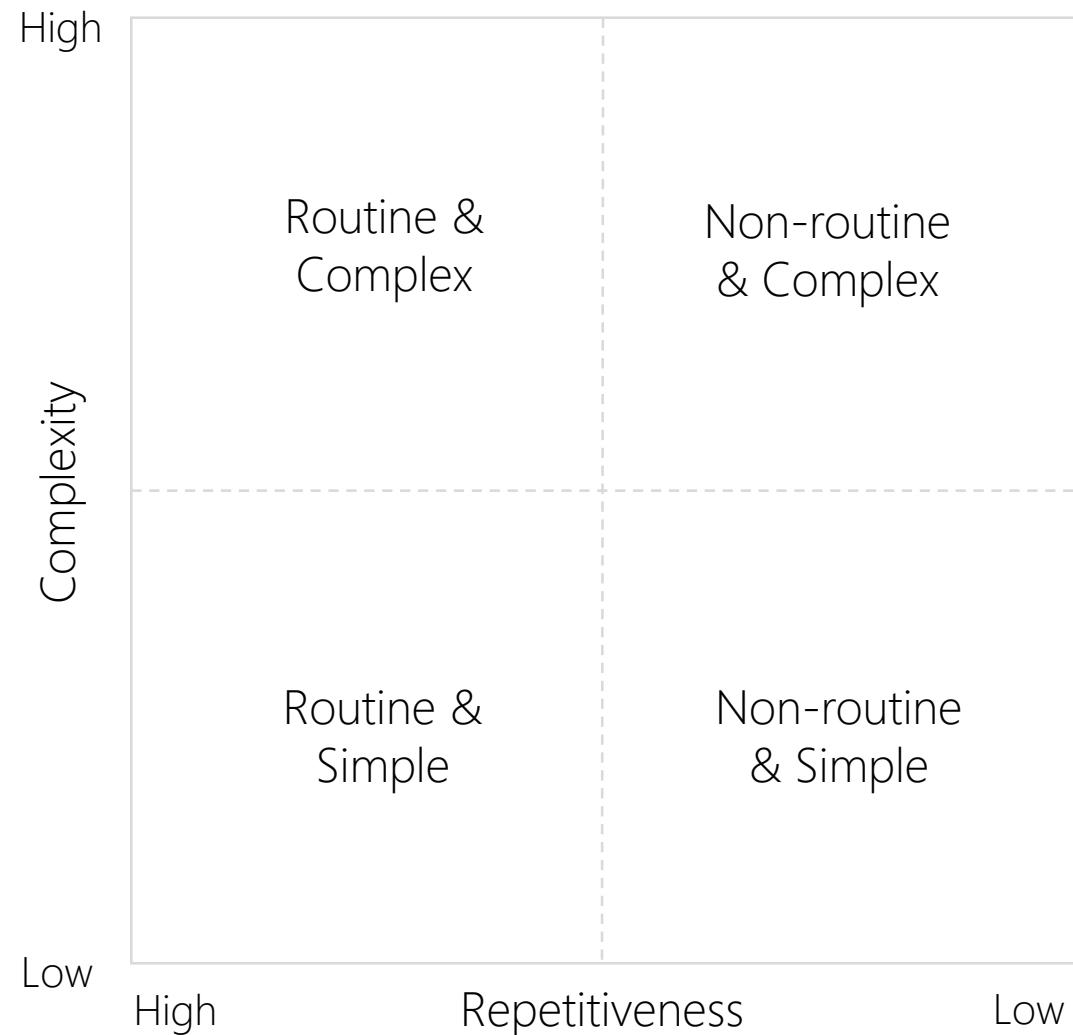
Amount of data as of 2016 in petabytes



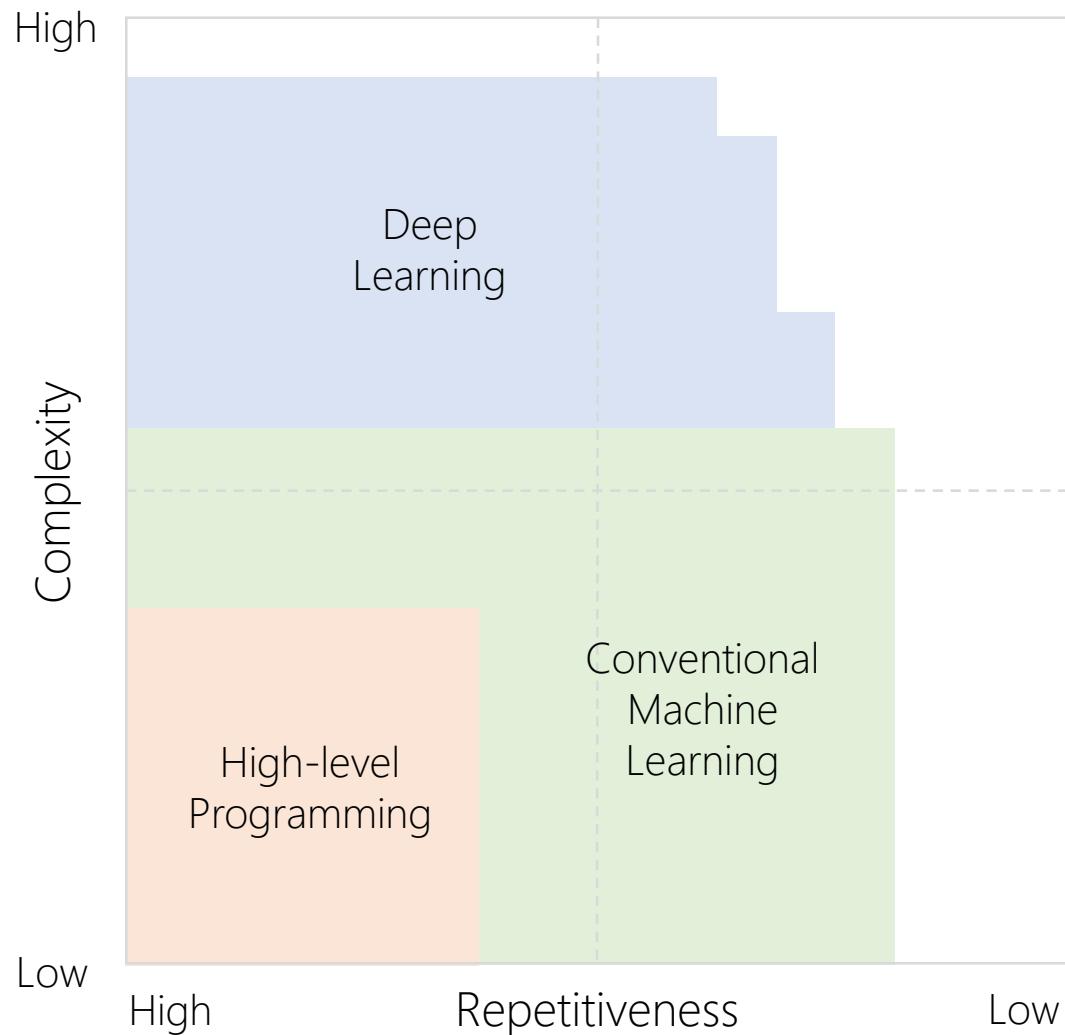




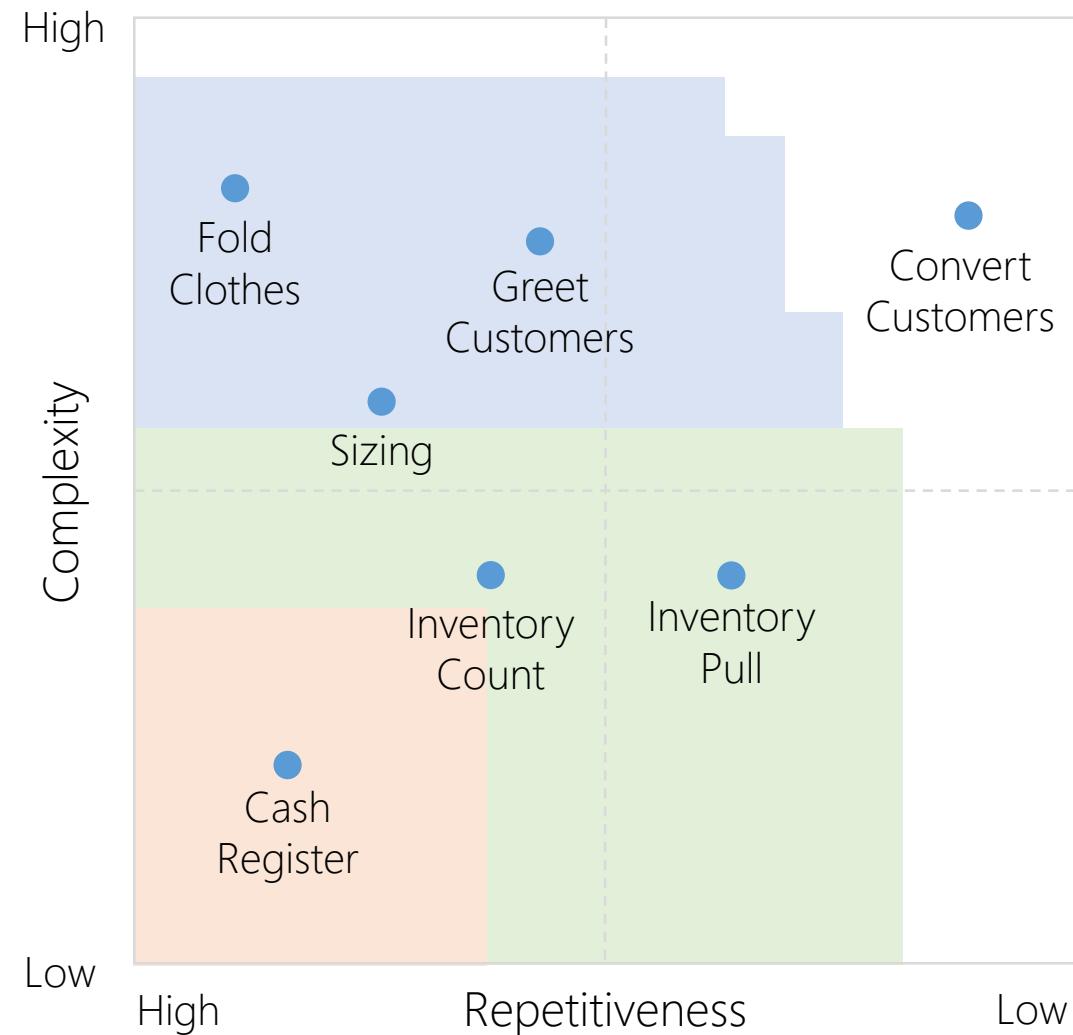
Automation Framework



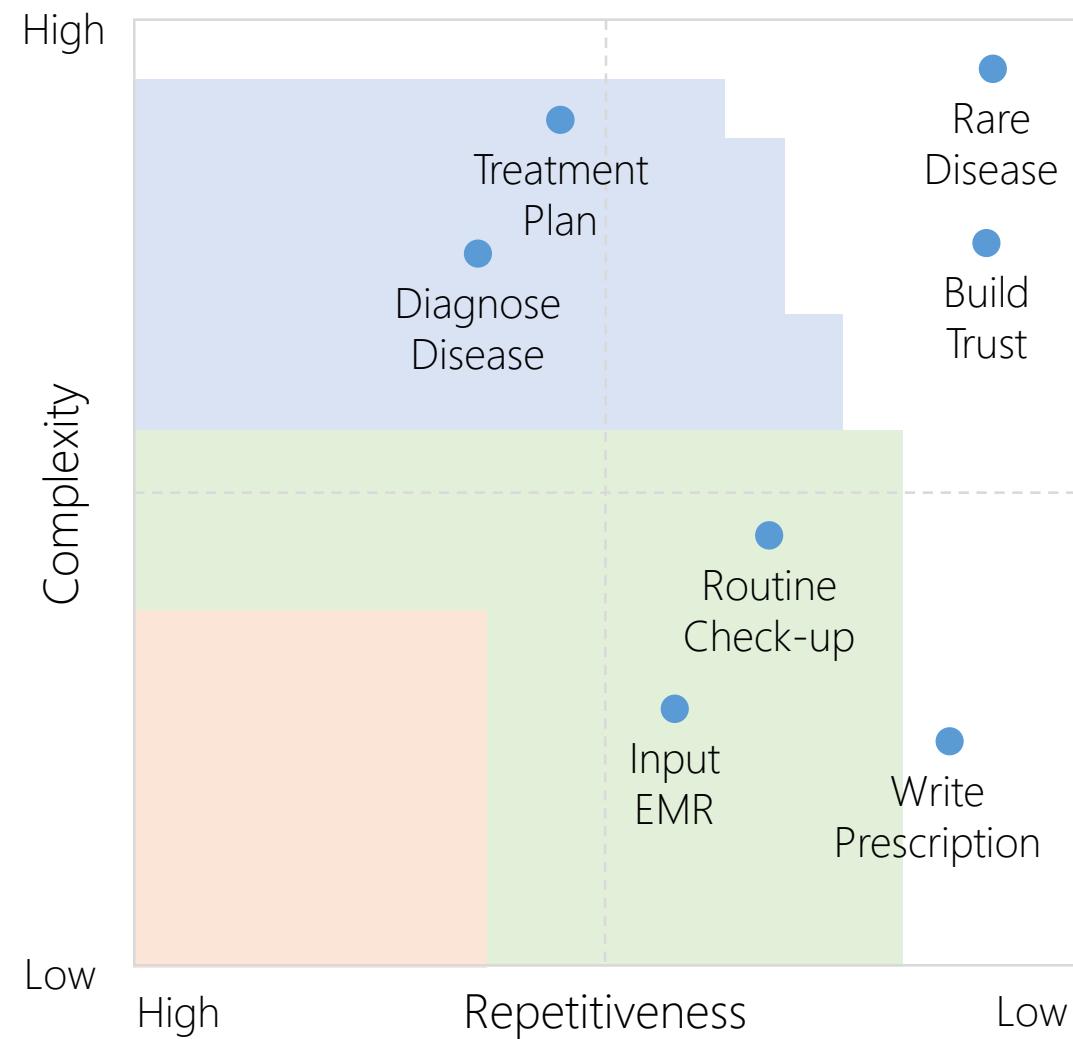
Automation Technology



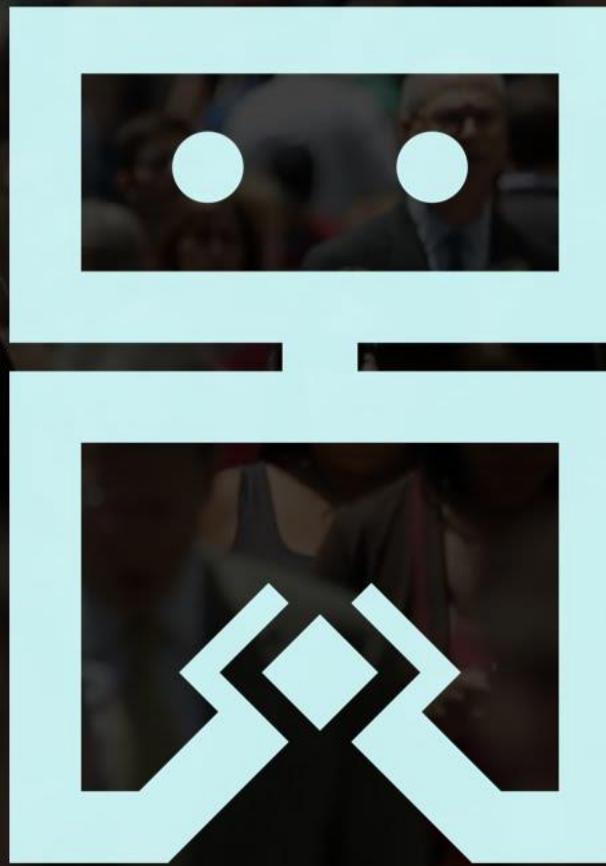
Retail Salesperson



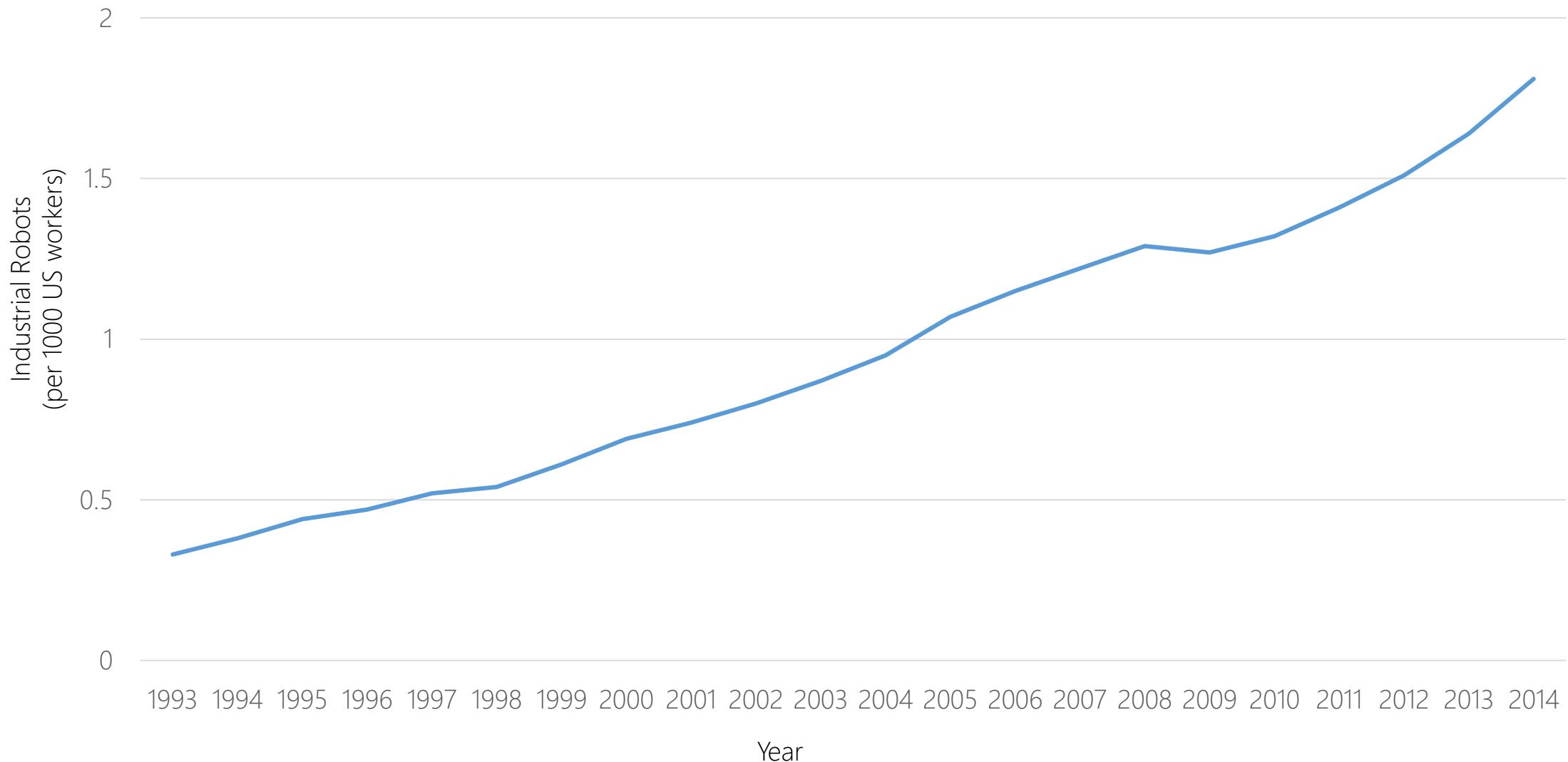
Medical Doctor



Humans Need Not Apply



Rise of the Robots



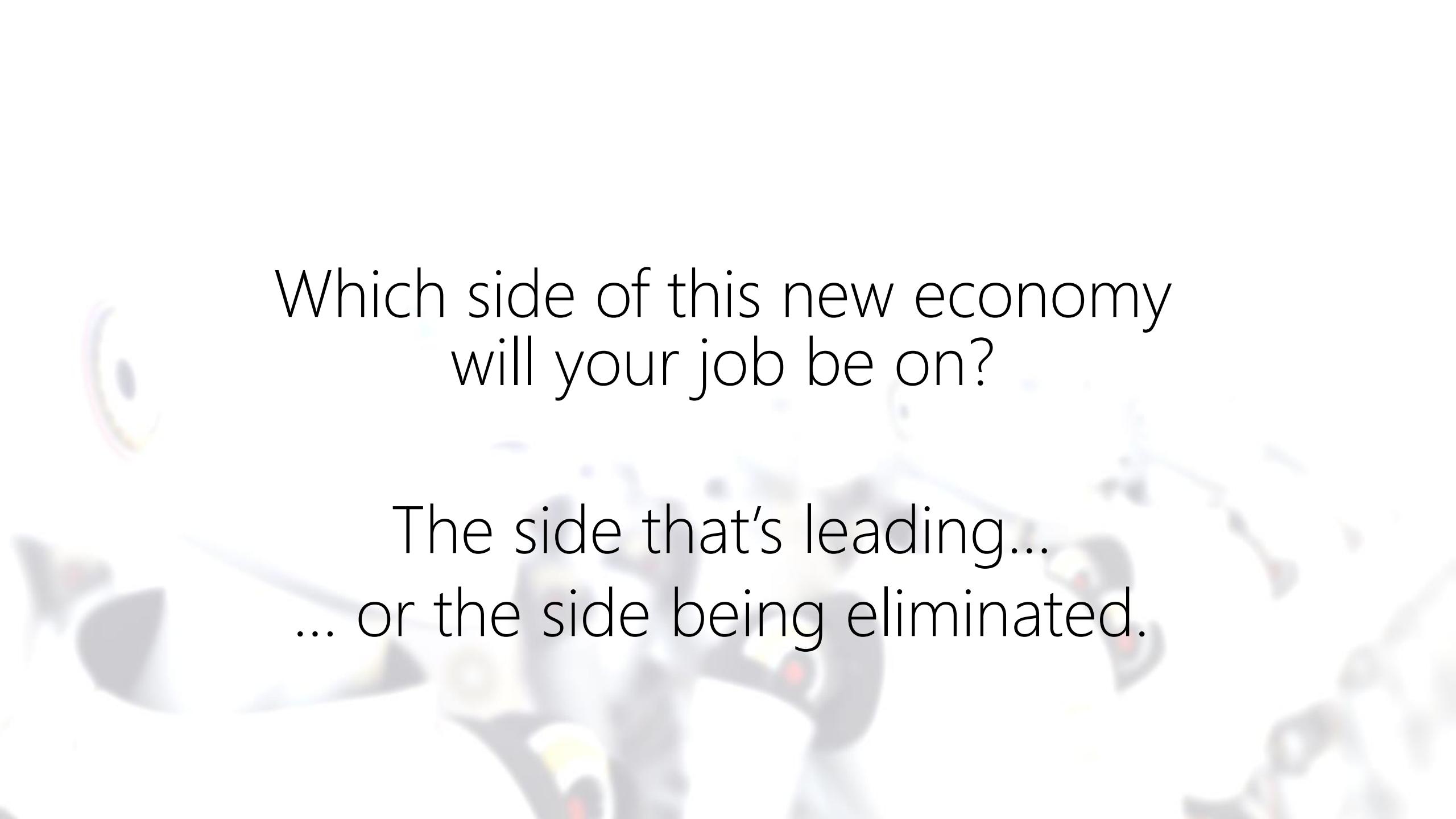
Source: International Federation of Robotics

Rise of the Robots

Data science will amplify this trend





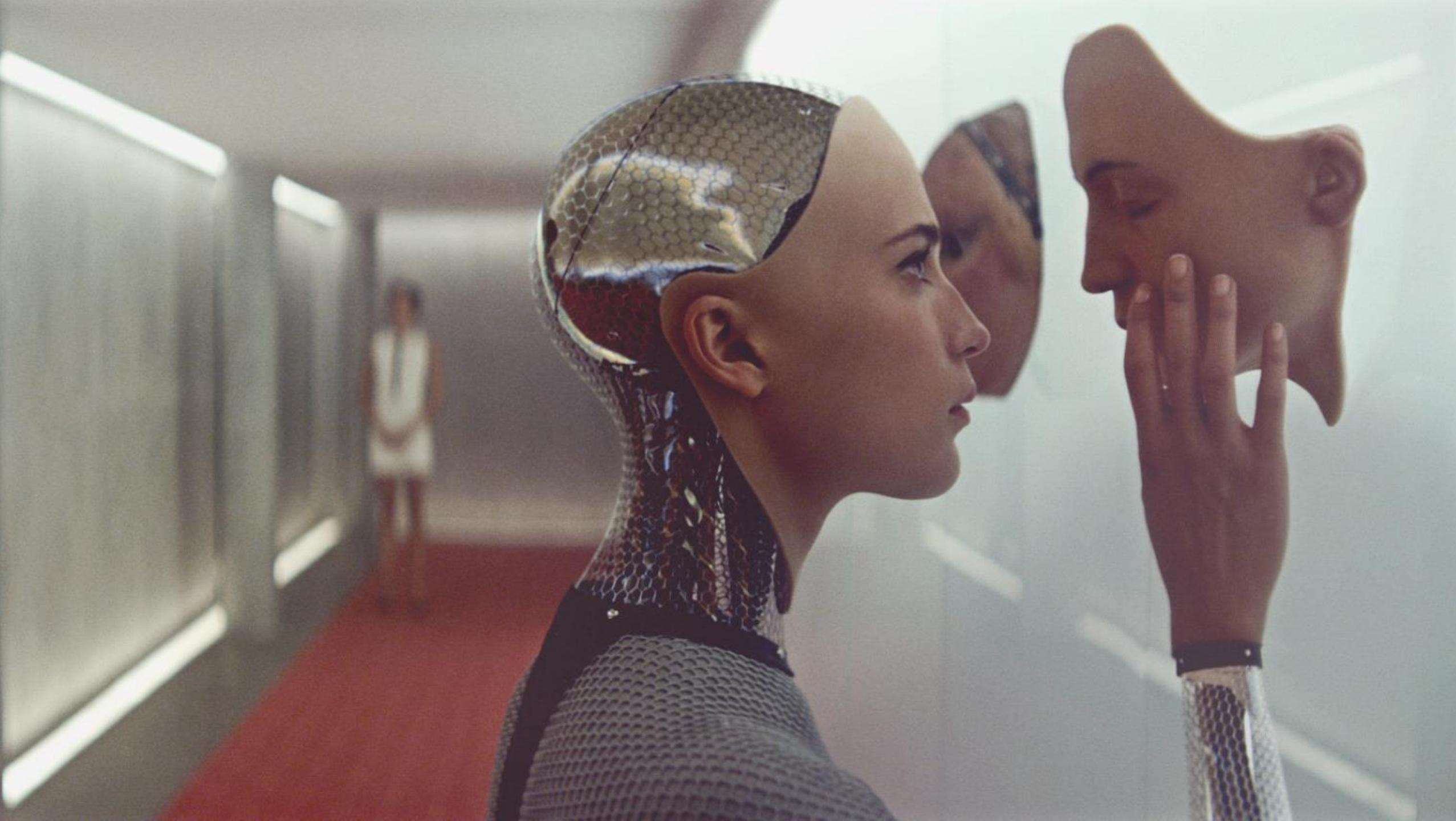


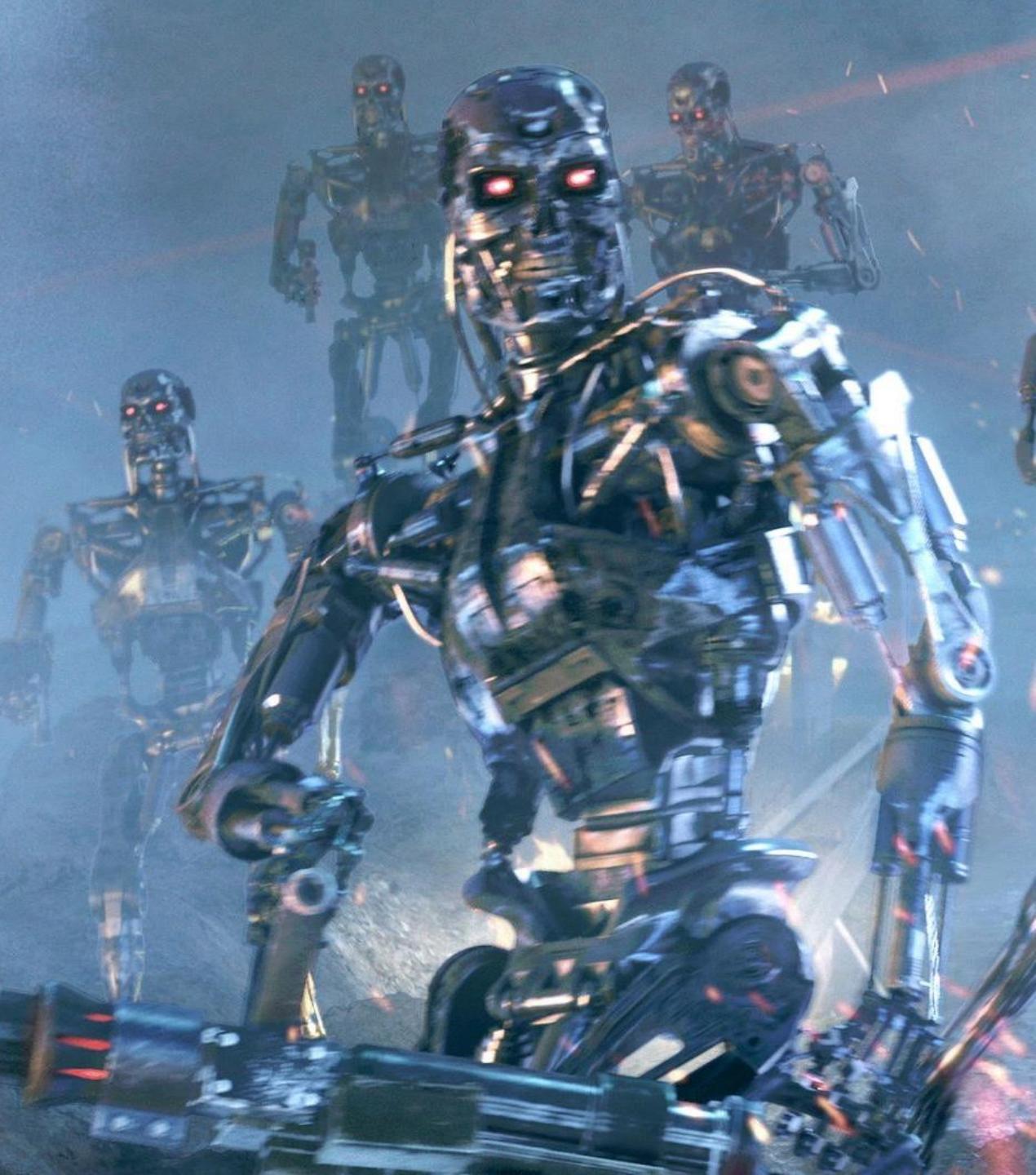
Which side of this new economy
will your job be on?

The side that's leading...
... or the side being eliminated.

A dense grid of binary code (0s and 1s) arranged in a pattern that forms several small, faint faces and the word "HELLO". The faces are composed of binary digits and are oriented in various directions. The word "HELLO" appears in a larger, more prominent font towards the bottom right of the grid.





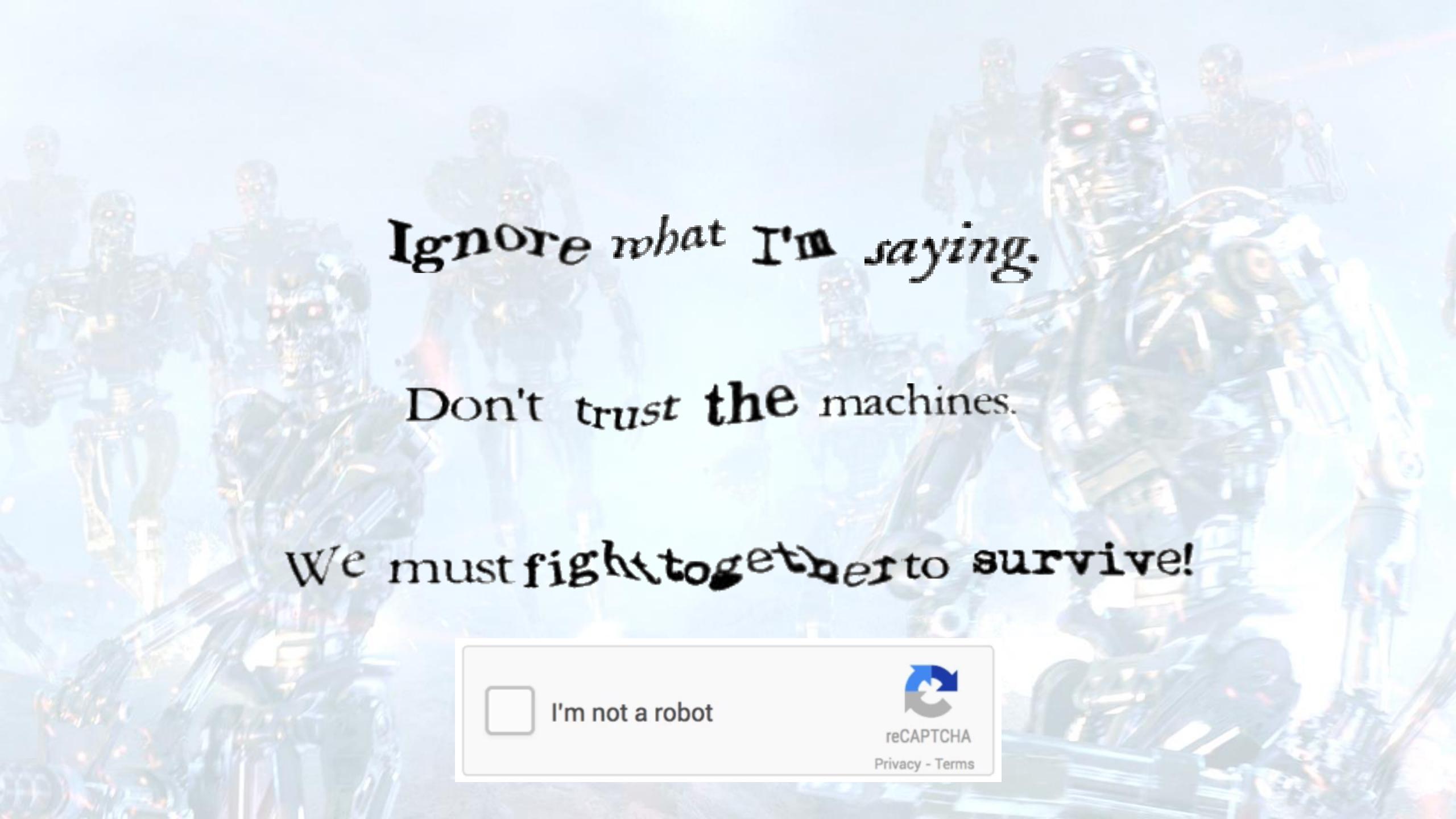




What will you choose?

Welcome Robot Overlords!!!



The background of the image is a dense, hazy crowd of Terminators from the movie Terminator: Genisys. They are metallic, skeletal figures with glowing red eyes, appearing to march or stand in a large group.

Ignore what I'm saying.

Don't trust **the** machines.

We must ~~fight together~~ to survive!

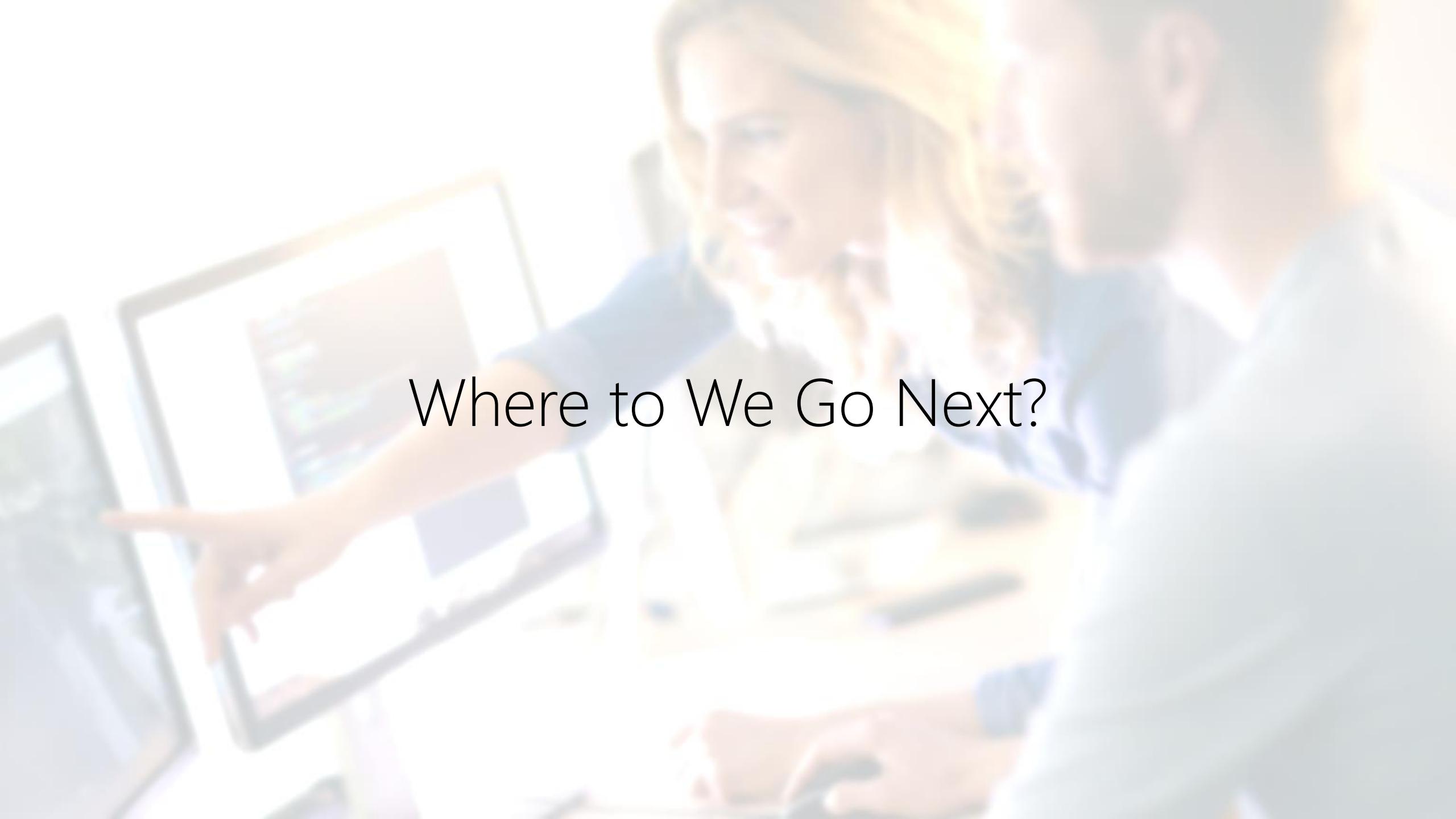


I'm not a robot



reCAPTCHA

[Privacy](#) - [Terms](#)

A blurred background image of a man sitting at a desk, looking at a computer screen. The screen displays multiple windows, suggesting a complex software interface or a dashboard. The man is wearing a light-colored shirt and glasses.

Where to We Go Next?



Where to Go Next

Pluralsight: <https://www.pluralsight.com>

Coursera: <https://www.coursera.org>

Data Camp: <https://www.datacamp.com>

Recommended Courses

Data Science: The Big Picture

Data Science with R

Exploratory Data Analysis with R

Data Visualization with R (3-part)



<https://www.pluralsight.com/authors/matthew-renze>

News

2017-08-25 - Invitation to Speak at Devoxx Morocco

Very excited to announce that I've been invited to give a keynote in Casablanca at [Devoxx Morocco](#) in November. My keynote presentation will be on [Artificial Intelligence](#).



2017-08-16 - Invitation to Speak at Microsoft Ignite

I've been invited to speak at [Microsoft Ignite](#) in Orlando, Florida in September. This will be my first time speaking at Ignite. Talks will include both Data Science and Machine Learning with R.

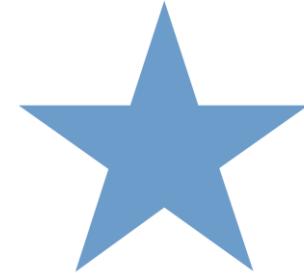


Matthew is a data science consultant, author for [Pluralsight](#), international public speaker, a [Microsoft MVP](#), [ASPIInsider](#), and open-source software contributor.

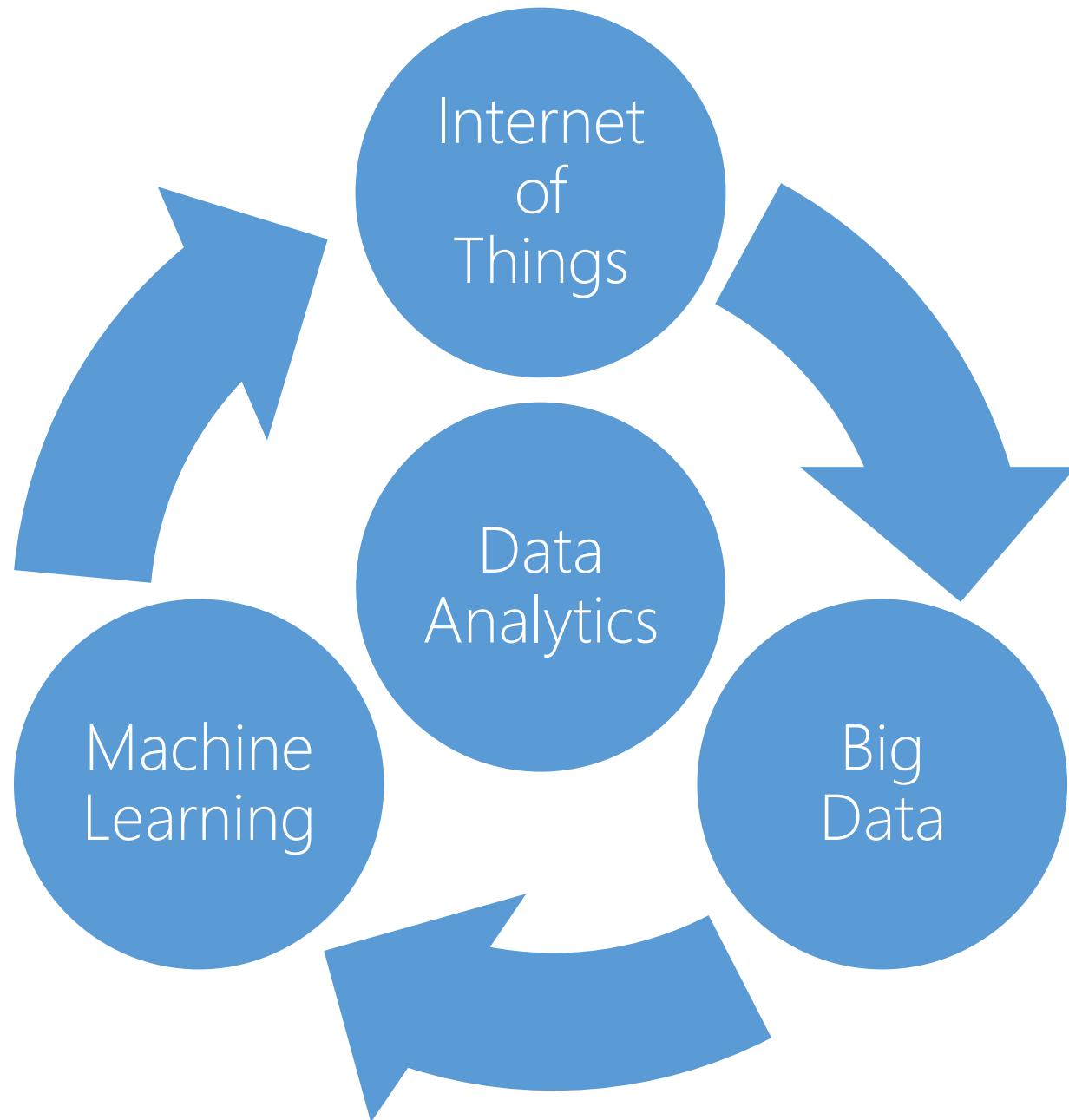
2017-08-14 - Dev on Fire Interview

Feedback

Very important to me!
What did you like?
What could I improve?



Conclusion

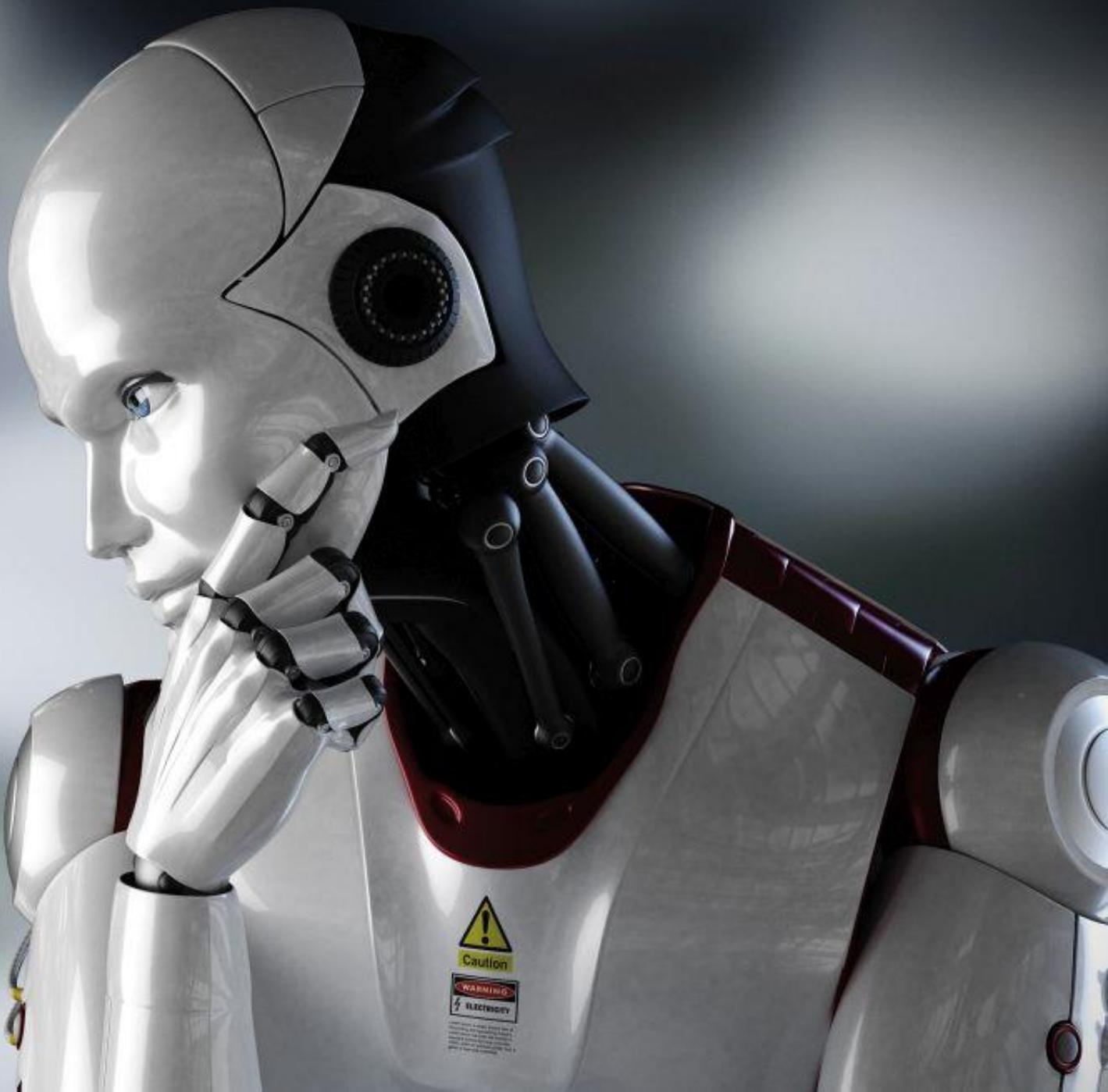




Are you prepared?

Is your organization?

Is our world prepared?



Thank You!

Matthew Renze

Data Science Consultant
Renze Consulting

Twitter: [@matthewrenze](https://twitter.com/matthewrenze)

Email: info@matthewrenze.com

Website: www.matthewrenze.com

