

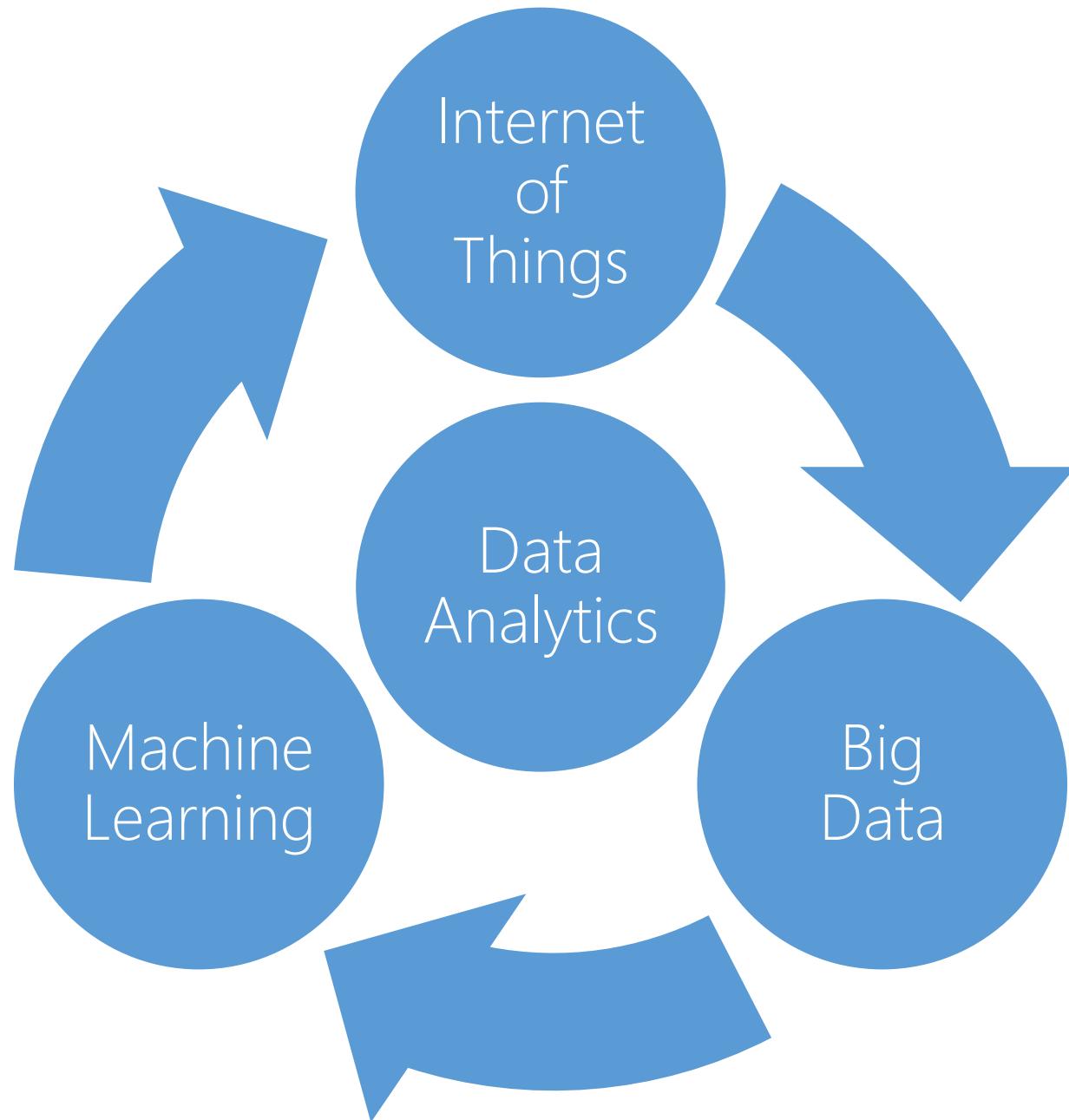
Data Science: The Big Picture

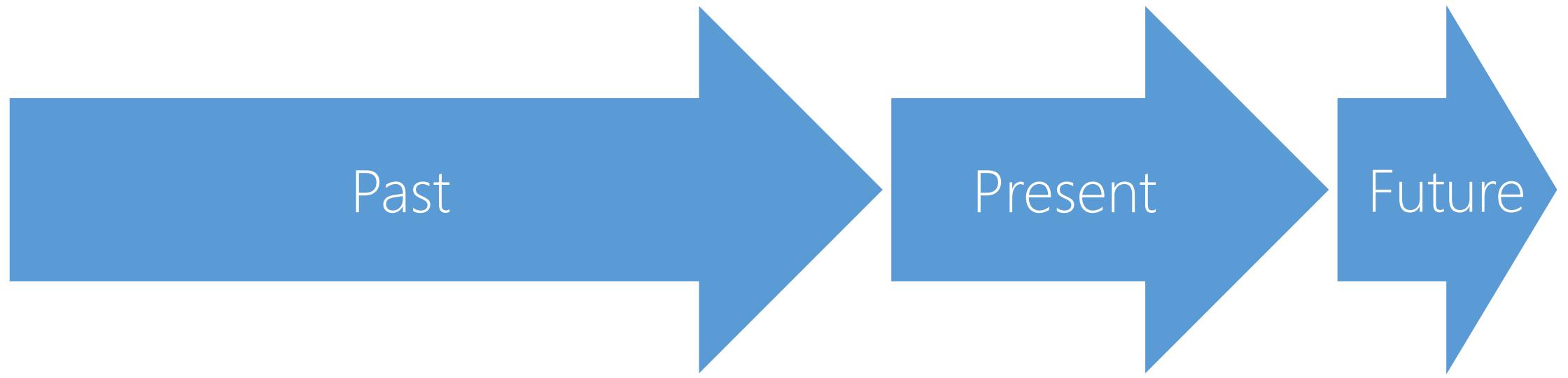
@MatthewRenze
#DetroitCode

What is data science?

Why is it important?

Where is this all going?





1978

2017

2037







The New York Times

For Today's Graduate, Just One Word: Statistics

By STEVE LOMK

Published: August 5, 2009

[TWITTER](#)

[LINKEDIN](#)

[COMMENTS
\(58\)](#)

[SIGN IN TO E-
MAIL](#)

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.

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

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

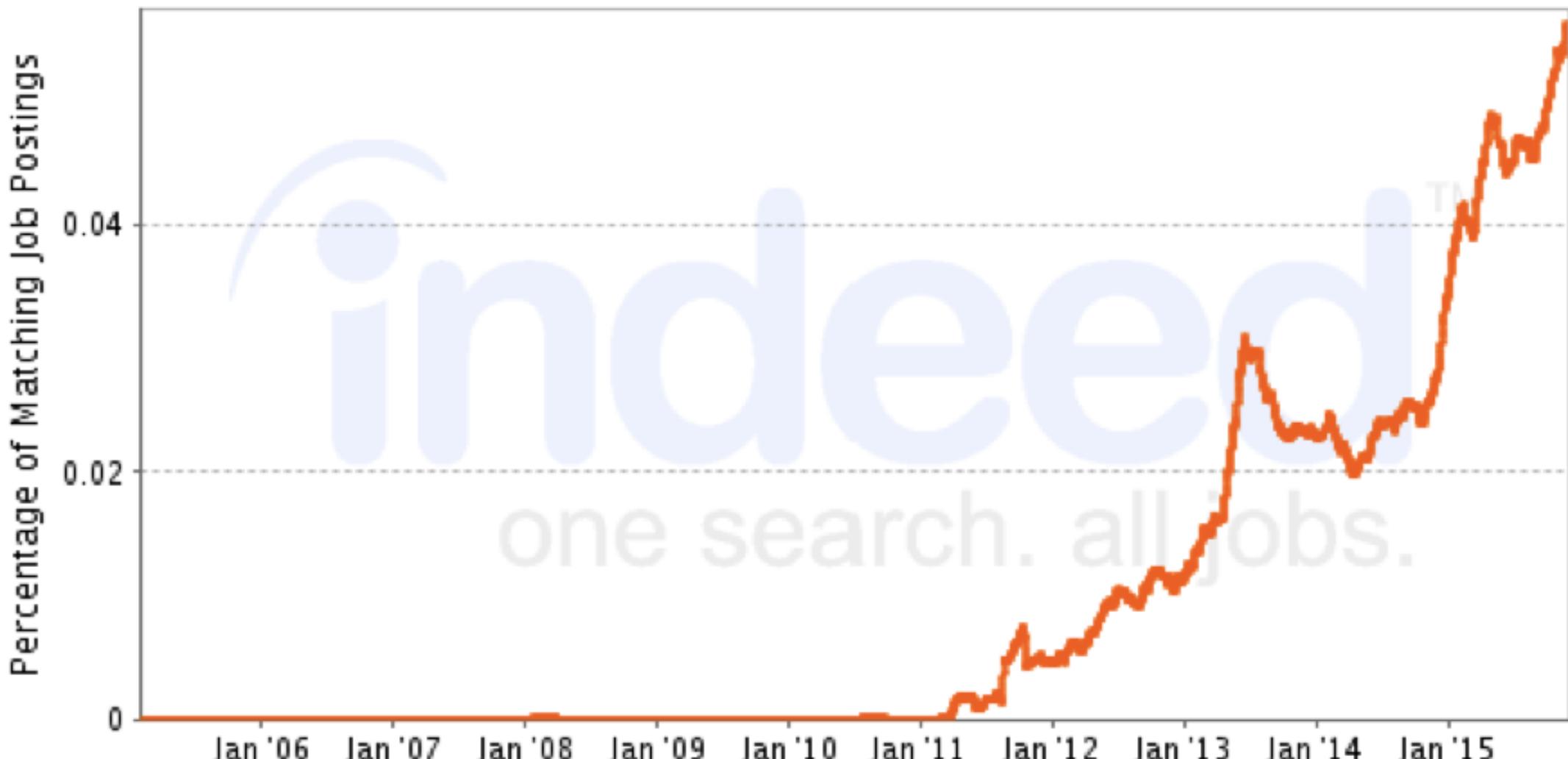
**Meet the people who
can coax treasure out of
messy, unstructured data.**

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 Trends from Indeed.com

— "Data Scientist"

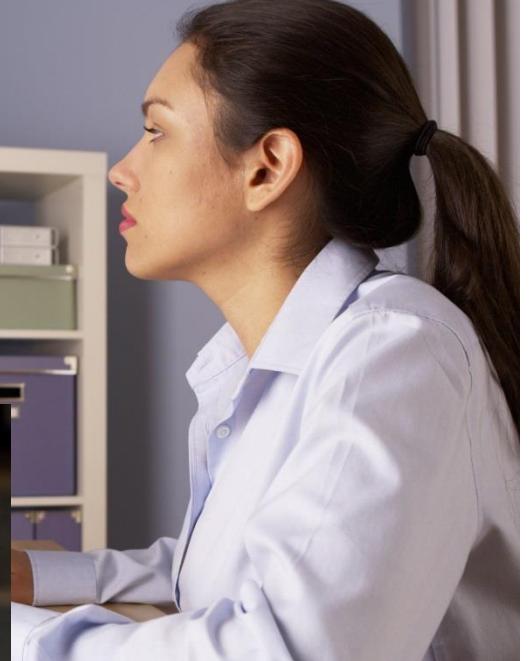
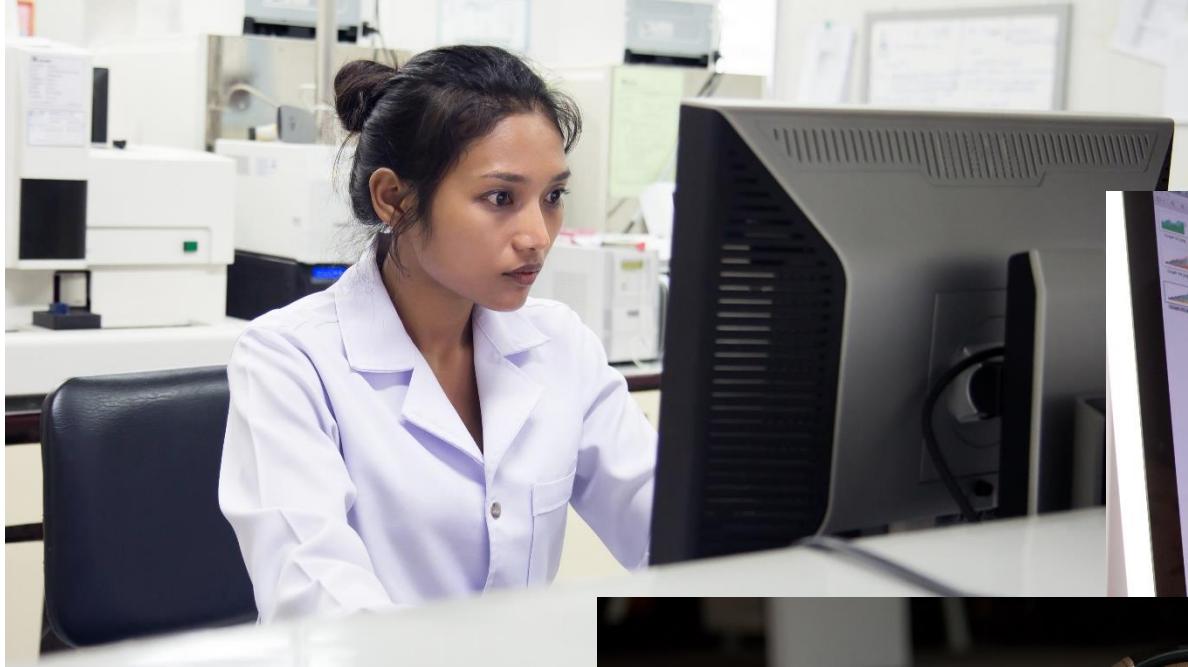


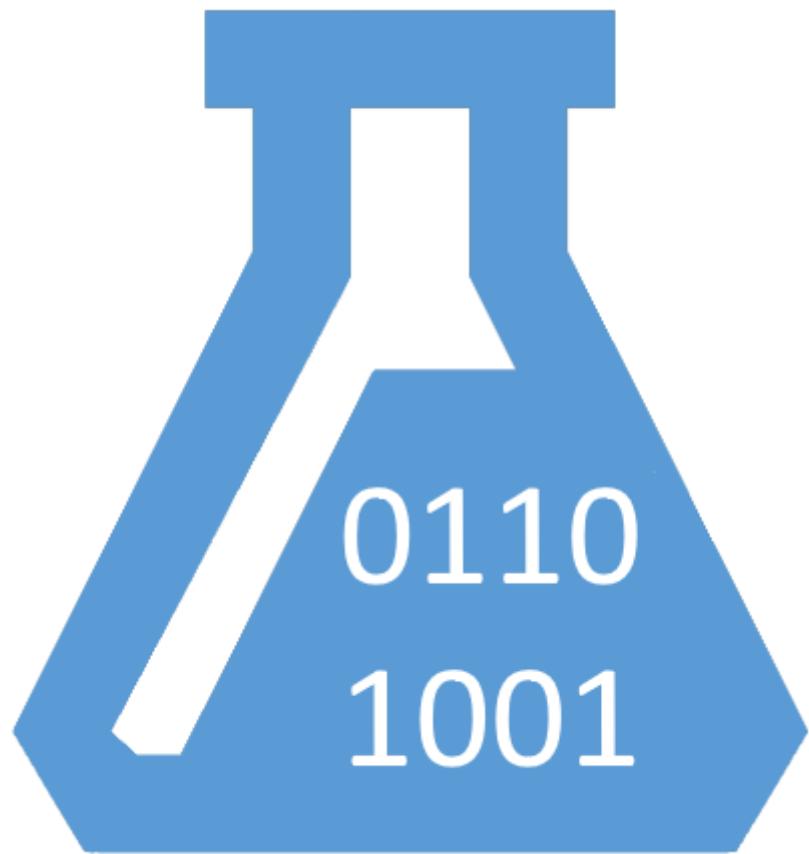
Source: Indeed.com

TOP-PAYING TECH SKILLS AND EXPERIENCE

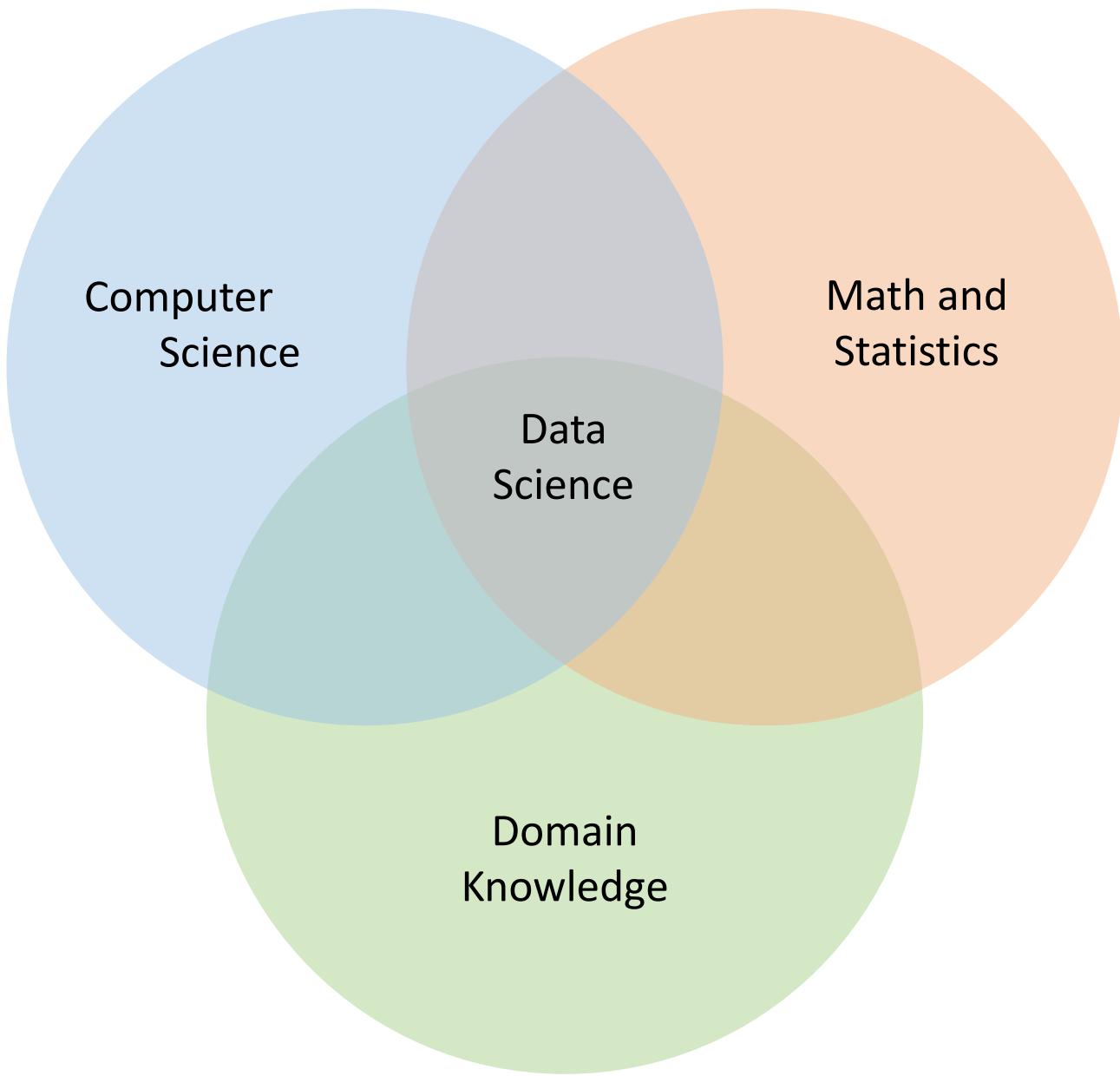
SKILL	2016	YR/YR CHANGE
HANA (High Performance Analytical Application)	\$ 128,958	-3.3%
MapReduce	\$ 125,009	-0.3%
Cloud Foundry	\$ 124,038	n/a
Hbase	\$ 123,934	5.7%
Omnigraffle	\$ 123,782	-1.9%
Cassandra	\$ 123,459	2.2%
Apache Kafka	\$ 122,728	n/a
SOA (Service Oriented Architecture)	\$ 122,094	-1.9%
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	YR/YR 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 Language)	\$ 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%

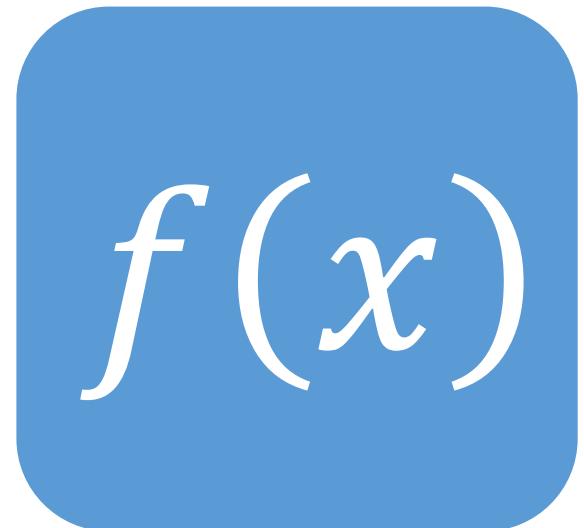
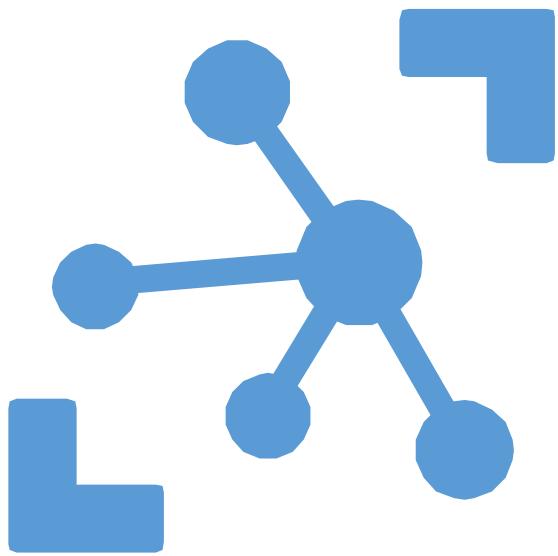




What is Data Science?

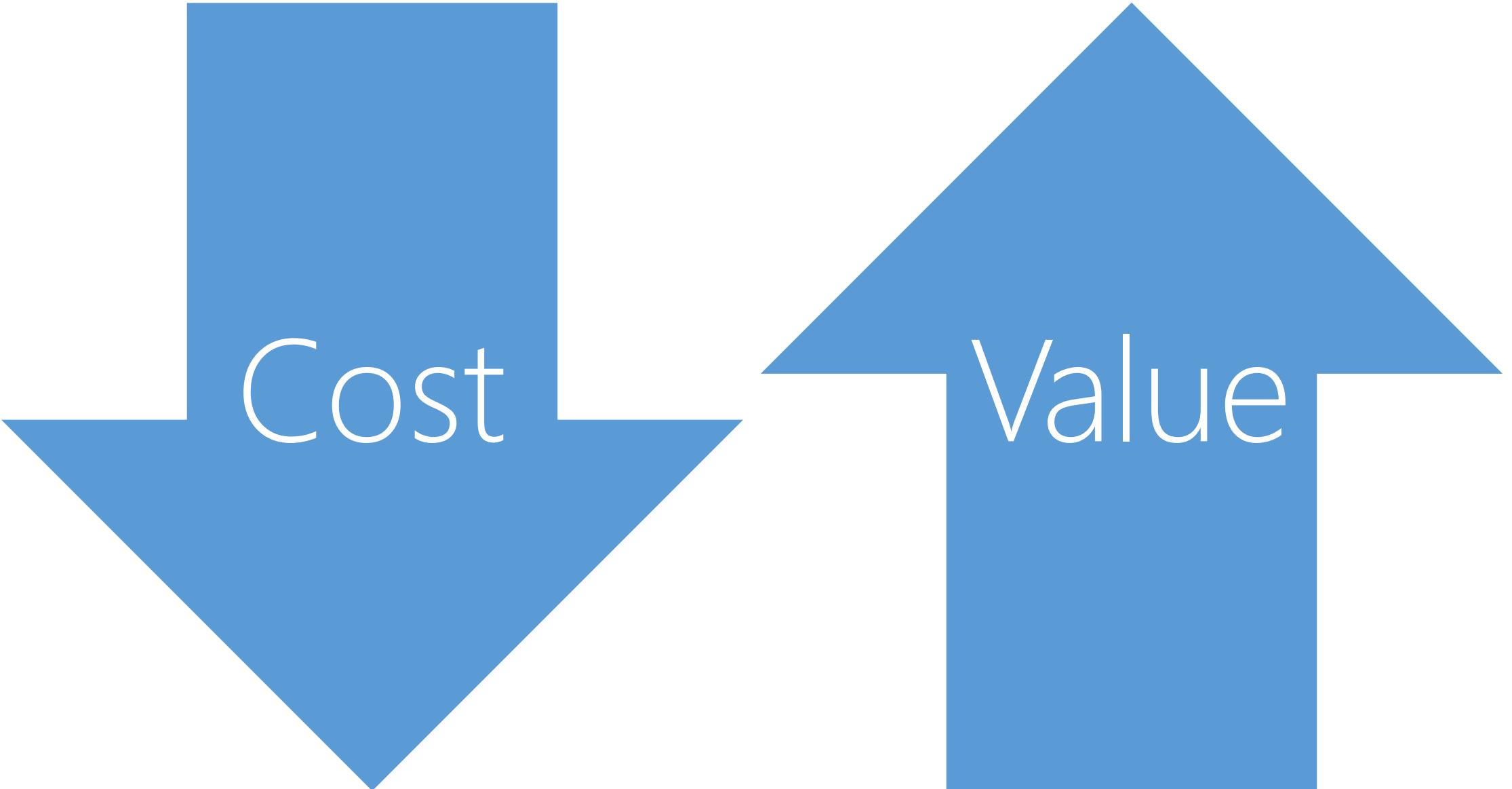


Why is Data Science Important?



Driven by economics

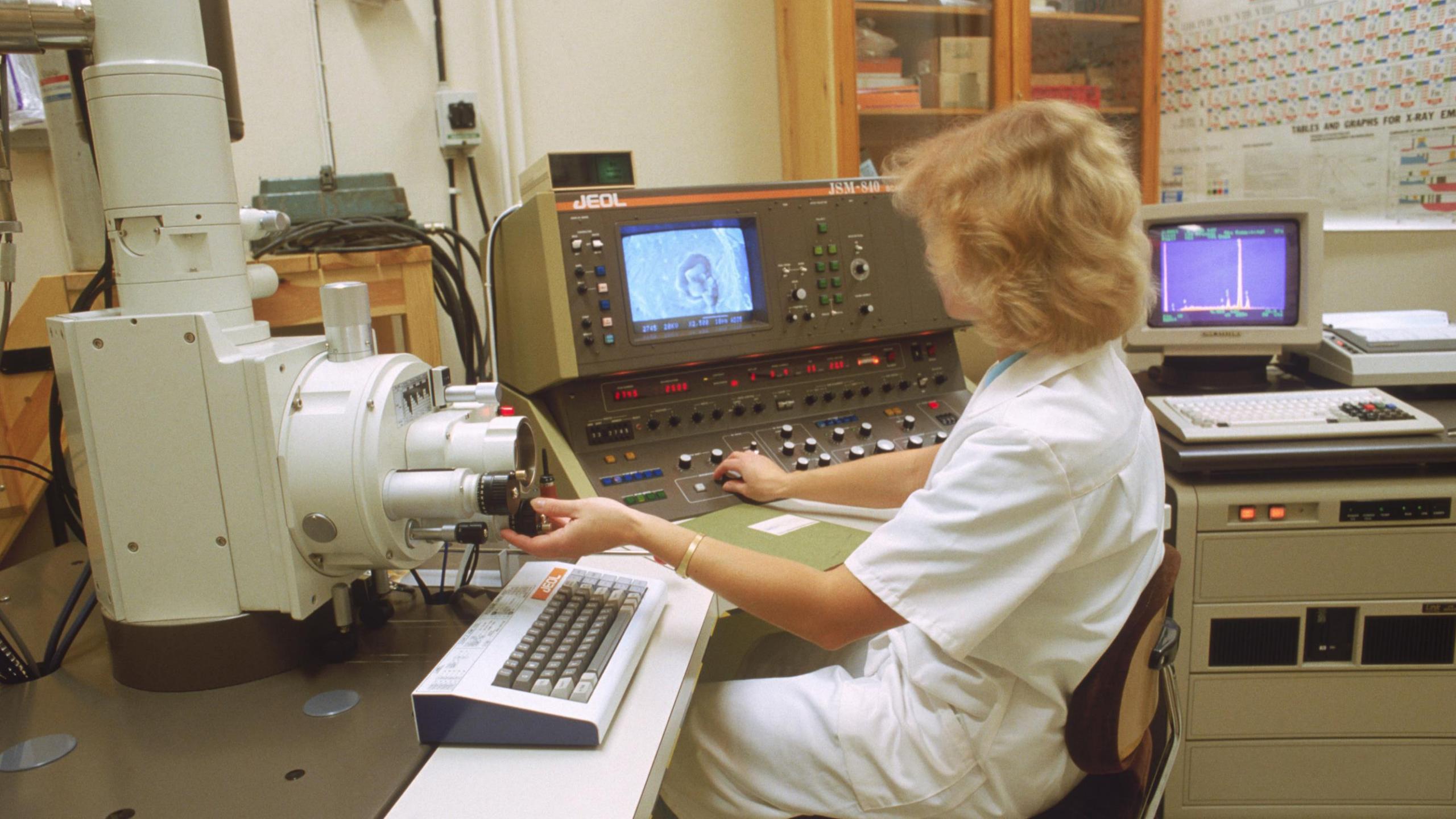
Possible by technology





Data Analysis (The Past)





JEOL

JSM-840



2745 2849 32 322 189 422



SUMITOMO

JEOL

SAS

File Edit View Tools Solutions Window Help

Using This Window
SAS Help and Documentation
Getting Started with SAS Software
Learning SAS Programming
SAS on the Web

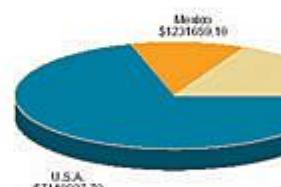
SAS Environment
SAS Libraries
SAS Help
Profile
Work
File Shortcuts
Favorites Folders
My Documents
My Desktop
SAS Repository View
3½ Floppy (A)
System (C)
Compact Disc (D)

Results Viewer - C:\WINNT\system32\pretty005.pdf

About SAS System

Total Product Sales by Country for 1997 & 1998
Country = Canada

Country	Product Type	1997		1998		Combined	
		Actual Sales	Predicted Sales	Actual Sales	Predicted Sales	Actual Sales	Predicted Sales
Canada	FURNITURE BED	\$224,162	\$225,449	\$226,509	\$226,818	\$448,671	\$452,207
	SOFA	\$273,456	\$242,847	\$243,794	\$273,047	\$477,230	\$518,480
OFFICE	AP	\$177,588	\$278,296	\$280,303	\$177,588	\$455,176	\$455,176
	Chair	\$202,383	\$220,991	\$220,458	\$220,458	\$423,833	\$441,458
	DASH	\$230,870	\$225,445	\$226,866	\$226,866	\$457,730	\$457,730
	AK	\$112,333	\$266,436	\$264,372	\$264,372	\$398,551	\$394,812
All Products							\$1,174,625.91



h105_example.log - Epsilon

File Edit Search Process Utility Window Help

log: C:\docs_Bayesian_Statistics\h105_example.log
log type: text
opened on: 13 Oct 2009, 16:31:04
.d
Contains data from C:\docs_Bayesian_Statistics\h105.dta
obs: 443
vars: 9 13 Oct 2009 16:28
size: 16,834 (99.9% of memory free)

variable name storage display value
variable type format label variable label

congress int %8.0g Congress
id long %12.0g icpsr id number
state byte %8.0g icpsr state code
district byte %8.0g cong. district number
statenm str7 %9s name of state
party int %8.0g 100=Demo, 200=Repub
member str9 %9s name of member
x1 float %9.0g dwnom 1st dim.
x2 float %9.0g dwnom 2nd dim

Sorted by:

h105_example.log [Fundamental] 19,0 Top
Epsilon Programmer's Editor 13.08, EEL version 13.08

CHODE PAGE 174

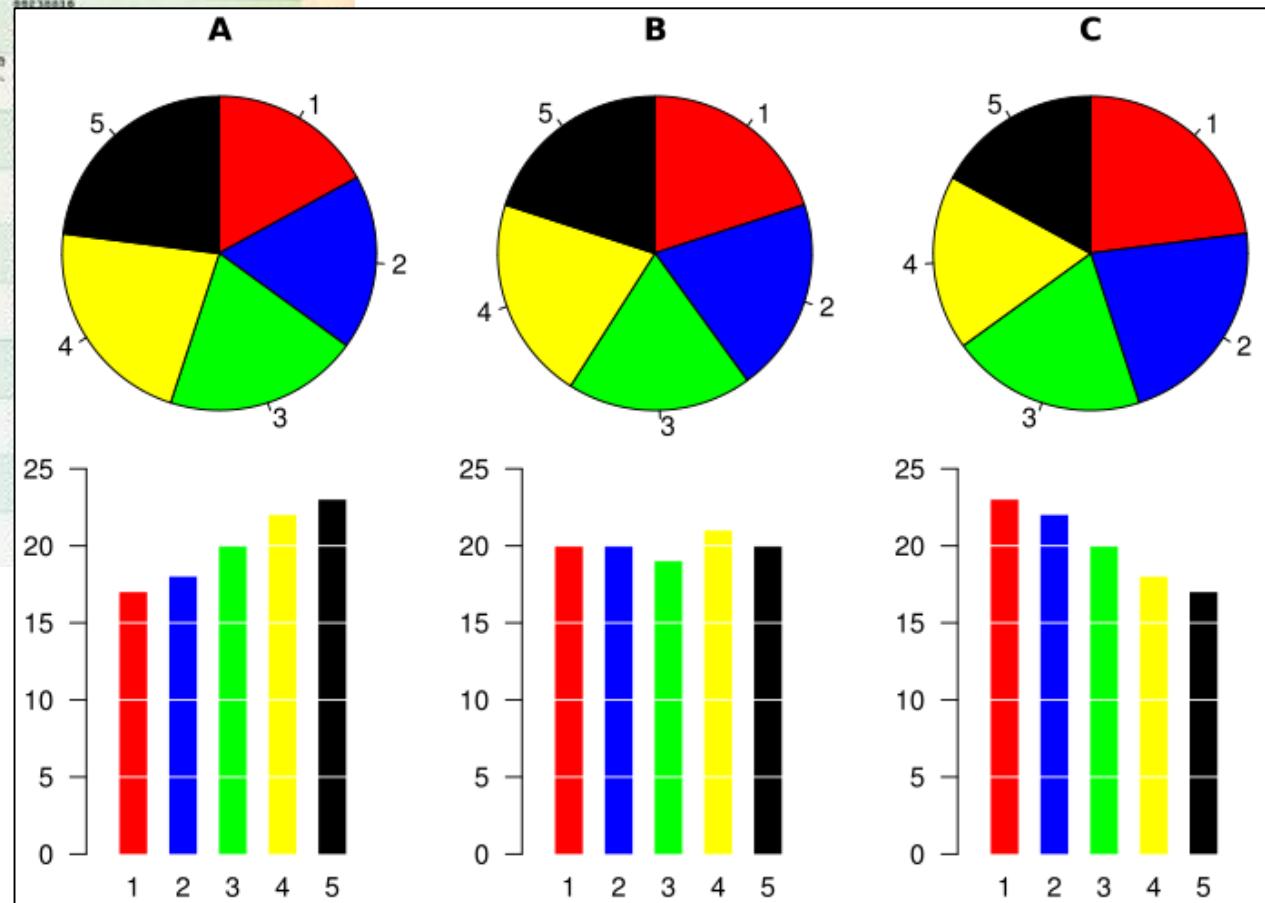
LOC OBJECT CODE ADDR1 ADDR2 STMT SOURCE STATEMENT

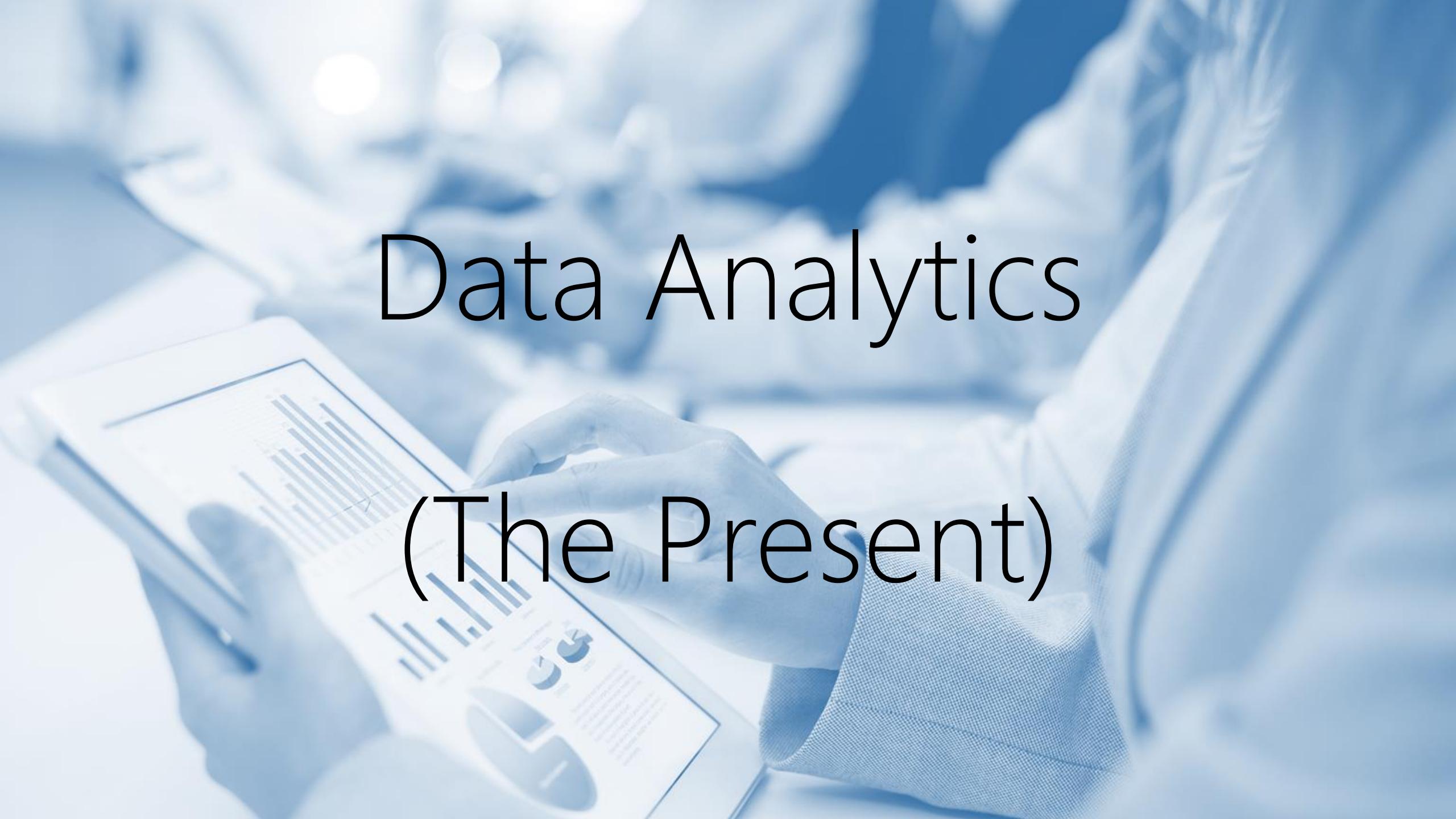
ADM W V 08 07.31 02/21/78

```

003492 **** 0100 CHODE DS DH
003492 4110 2248 00260 0101 LA #1,RSPTB ADDRESS OF ADDRESS OF RB
003492 0463 0102 SVC 99 CALL DYNAMIC ALLOCATION
003492 12FF 0103 LTR CORPCODE,COMPCode SEE IF FAILED
003492 4770 3009 03000 0104 BHZ AERR3 YES ISSUE MESSAGE
003492 0207 20CC 2240 002AD 0105 MVC JESACB+ACBDRN-SPACB(16),DNAME MOVE CONAPE
003492 0207 20CC 2240 002CC 0106 OPEN (JESACB) OPEN INTERNAL CONSOLE
003492 0107+ CNOP 8,4 ALIGN LIST TO FULLWORD
003492 4519 34AC 03HAC 0108+ BAL 3,X#0 LOAD REG1 W/ LIST ADDR.
003492 00 0109+ DC AL111201 OPTION BYTE
003492 0000A4 0110+ DC AL3,JESACB DCB ADDRESS
003492 0A13 0111+ SVC 1V ISSUE OPEN SVC
003492 1C7F 0112 LTR CORPCODE,COMPCode
003492 4770 3008 03008 0113 BHZ AERR2 ISSUE FOR MSG
003492 4110 2244 00544 0114 LA #1,JESACB ADDRESS OF ACB
003492 2979 00770 0115 ST #1,JESPLH+RPLSACB-ZFPLP1 STORE IN RPL
003492 4110 429C 0219C 0116 LA #1,CPO ADDRESS OF CPO BUFFER
003492 5010 2978 00778 0117 ST #1,JESPLH+RPLAREA-ZFPLP1 STORE IN RPL
003492 5313 42EC 012EC 0118 L #1,CHLEN LENGTH OF COMMAND BUFFER
003492 5410 2958 00900 0119 ST #1,JESPLH+RPLLEN-ZFPLP1 STORE IN RPL
003492 4793 5108 03400 0120 S CTR
003492 9419 3582 01E00 0121 SVS BYSTATE,BTS
003492 4770 34E4 034E0 0122 B ARBK
003492 4110 4874 020F4 0123 CTR PDS PDSPT,,PUT
003492 4100 8956 00500 0124+ LA #1,PROMPT SET ADDRESS
003492 4100 8956 00500 0125+ LA #0,PROMPT SET LENGTH
003492 4588 380C 018CC 0126+ BAL LSHZ,E.PUT GO TO PUT
003492 4110 429C 0229C 0127 ARBK PDS CPO,75,SET
003492 4100 894E 00544 0128+ LA #1,CPO SET ADDRESS
003492 4285 3846 01840 0129+ BAL LDHZ,GET SET LENGTH
003492 4110 3F82 01880 0130 TH $19STAT2-BTS IS 'SYS' IN EFFECT?
003492 4715 8548 02542 0131 ED CURRENT YES, PROCESS
003492 0581 429C 8E4F 0229C 02540 0132 CLC CPO1,1+C"END"
003492 4793 510A 03514 0133 BE EXIT
003492 0513 429C 8E28 0229C 03620 0134 CLC CPO1,2+C"LB2"
003492 4793 5105 03500 0135 BE CIB
003492 0593 429C 8E3C 0229C 03700 0136 CLC CPO1,1+C"RS"
003512 4798 5100 03400 0137 BE BYT
003512 4798 5100 03400 0138 BE BYT
003512 4798 5100 03400 0139 CLC CPO1,1+C"RS"
003512 4798 5100 03400 0140 BE INTCON
003512 4798 5100 03400 0141 CLC CPO1,1+C"ST"
003512 4798 5100 03400 0142 BE INTCON

```



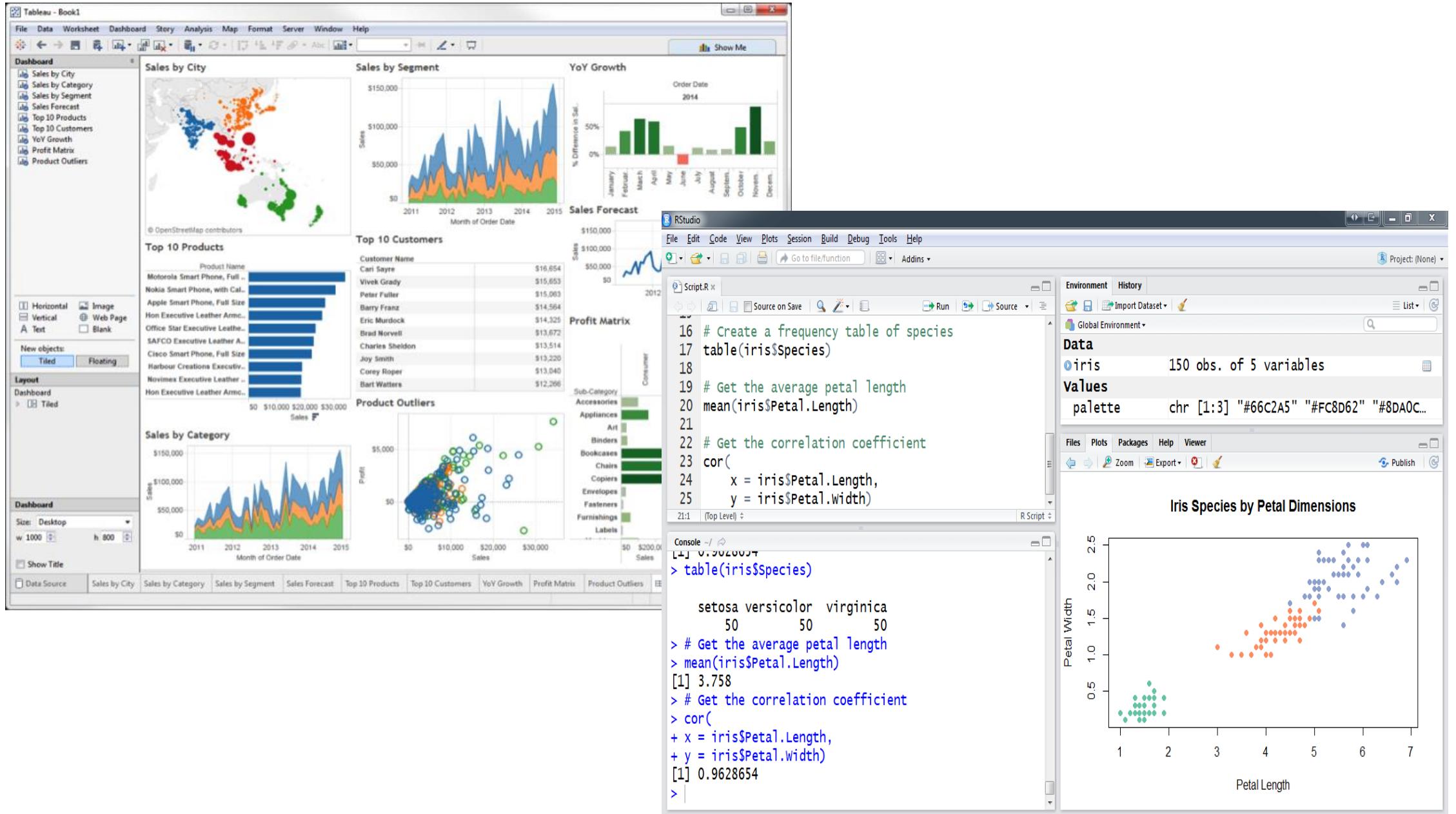
A close-up photograph of a person's hands and face. The person is wearing a light blue surgical mask and a pair of blue nitrile gloves. They are holding a white piece of paper with several blue bar charts printed on it. The background is blurred, showing more of the same blue medical equipment.

Data Analytics

(The Present)







Marketing

New visitors via search, 30 days

20,340



New visitors via PPC, 30 days

3,634



New visitors via social, 30 days

18



Twitter followers

5,700



Facebook Ads spend, Campaign Y

\$47.57

spent all time

Mailchimp weekly list growth

137

▲ 15.13%

Current website

13



Clicked Link X today

377

▼ 3.33%

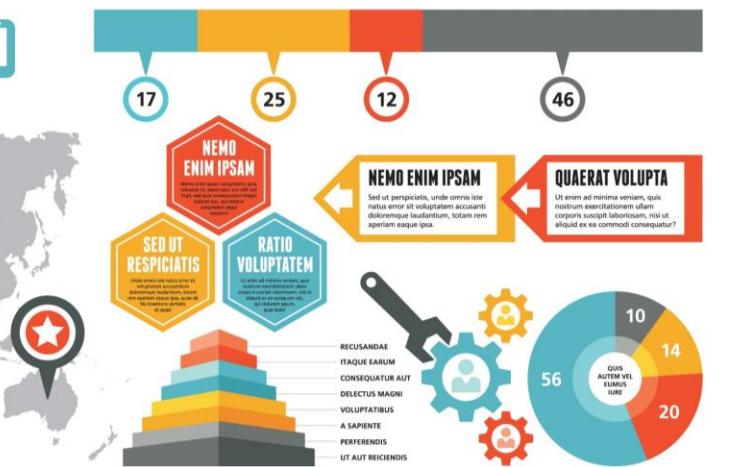
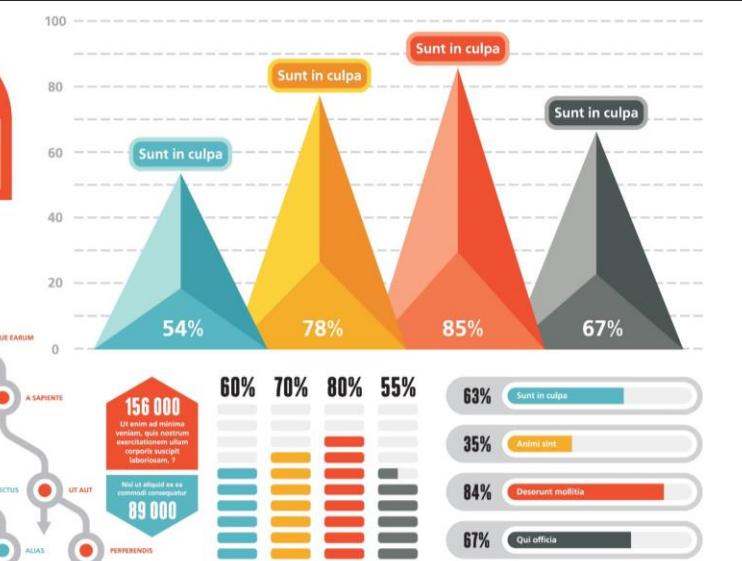
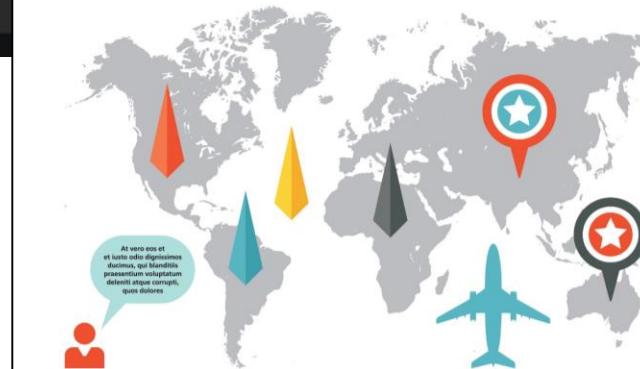
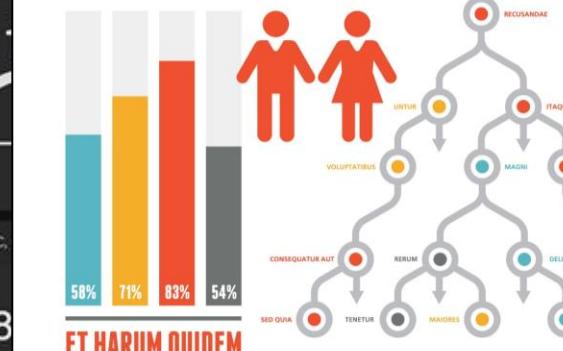
Average CPC,

£0.8

▼ 36

95%

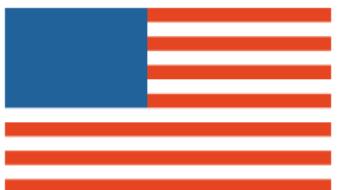
\$50



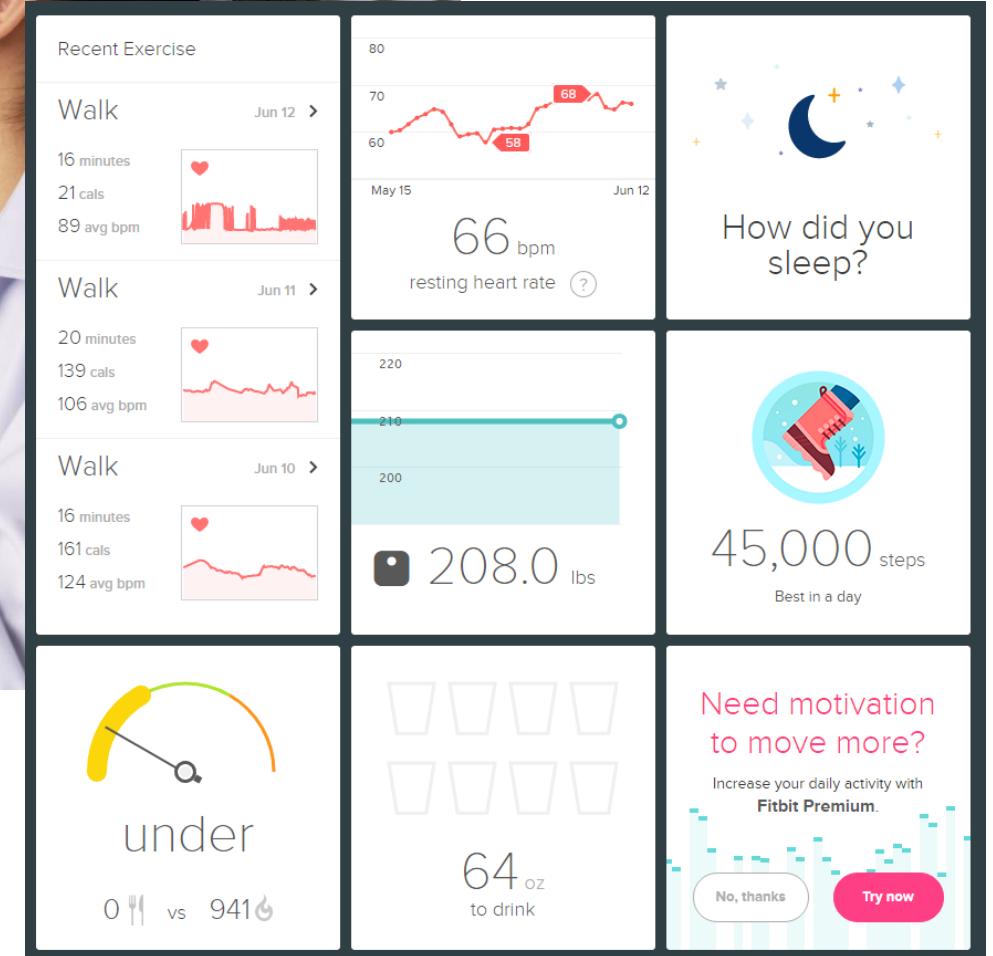
Data-Driven Decision Making

(The Future)





DATA.GOV



Retail Analysis Sample

[Share Dashboard](#)

Ask a question about the data on this dashboard

Total Stores

104

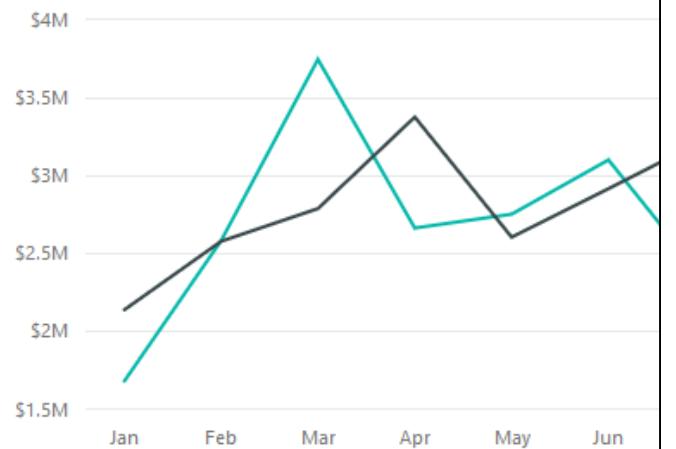
This Year's Sales

\$22.

This Year's Sales, Last Year's Sales

BY FISCAL MONTH

This Year Sales Last Year Sales



This Year's Sales

New Stores Opened This Year

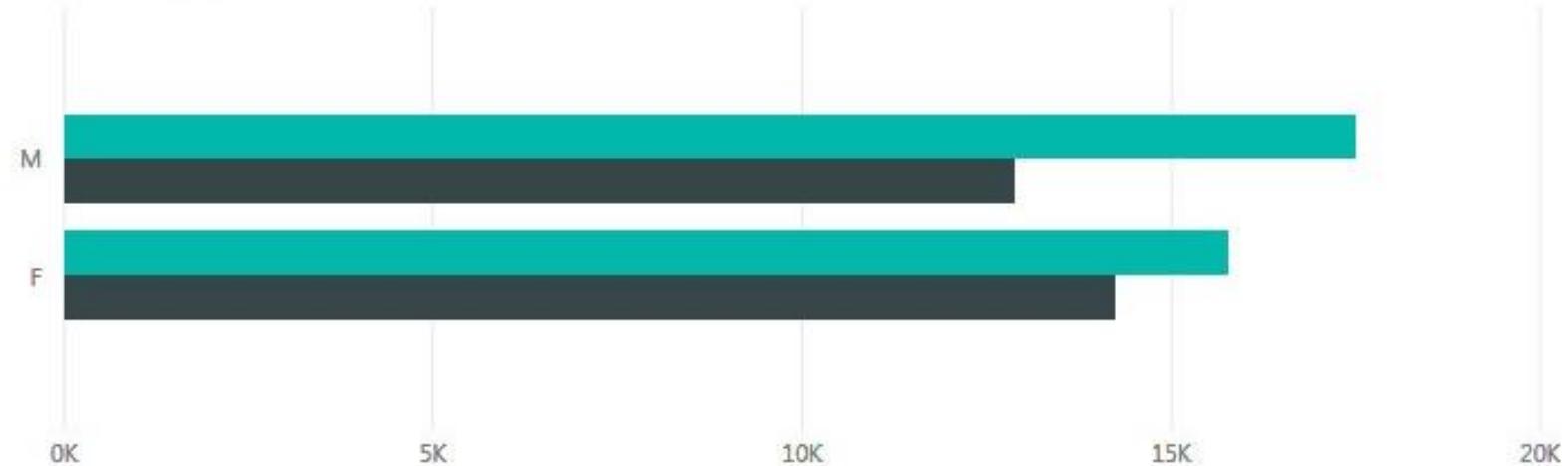
AW_DW_InternetSales_2.xlsx

sales by gender and marital status

Show gender, marital status, and number of AWDW internet sales

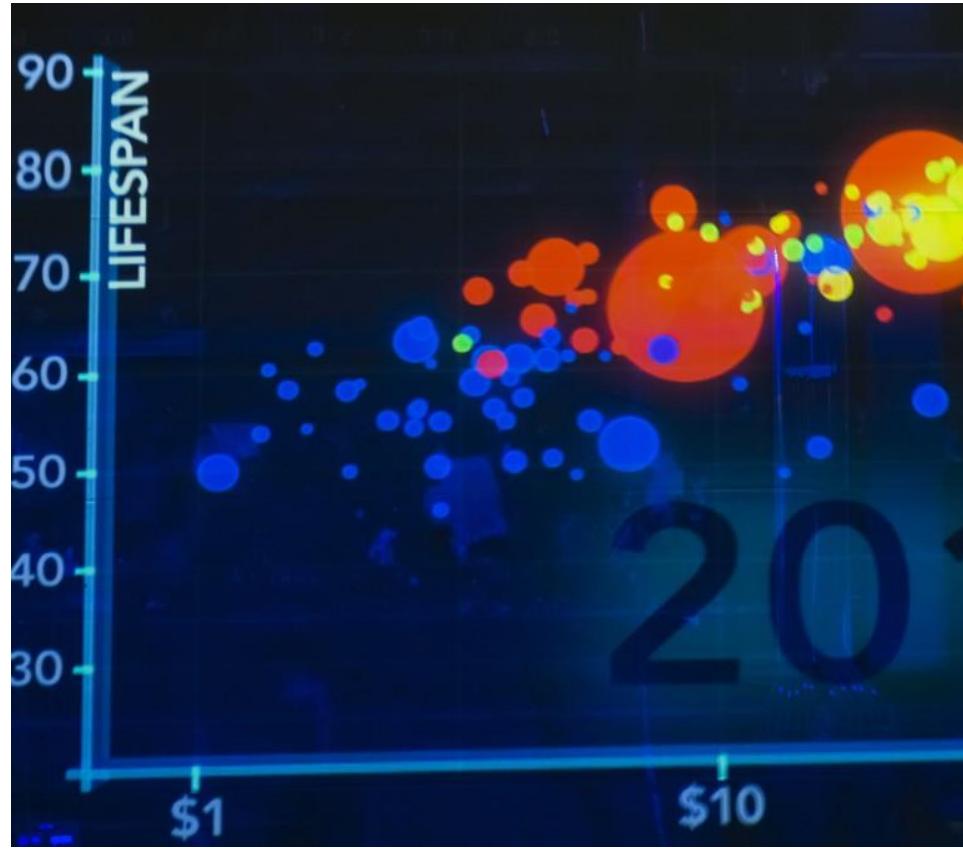
Count of AWDWInternetSales by Gender and MaritalStatus

MaritalStatus M S



Source: AW_DW_InternetSales_2

Show me sales
by gender and
marital status.







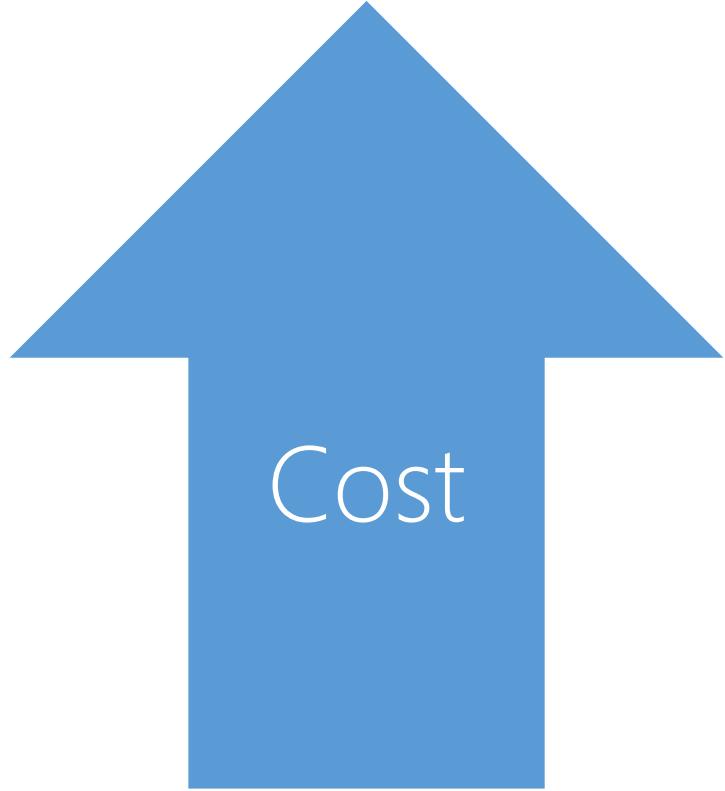
Empowering people to make better decisions is not the end goal...

... it's just the beginning

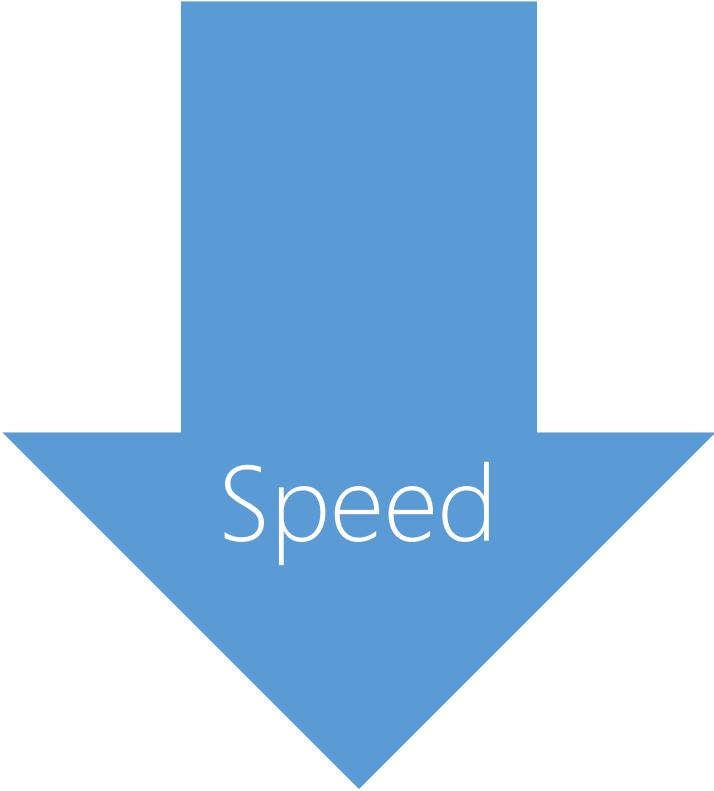
The Internet (The Past)



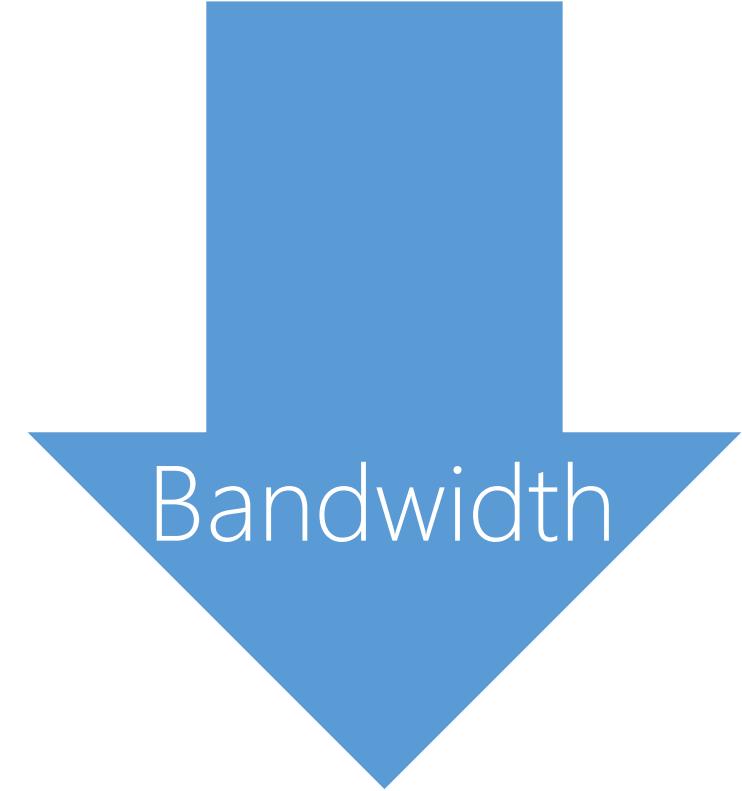




Cost



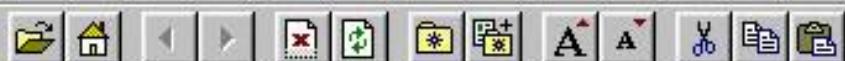
Speed



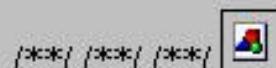
Bandwidth

Wiki does not exist

File Edit View Favorites Help



Address: <http://en.wikipedia.org/>



Wiki does not exist

From Meta, a wiki about Wikimedia

This wiki does not exist yet. Perhaps you are looking for one of our other projects:

Projects

[Complete list of wikimedia projects](#)

[Proposals of new projects](#)

Wiktionaries

Untitled Document - Microsoft Internet Explorer

File Edit View Favorites Tools 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: @

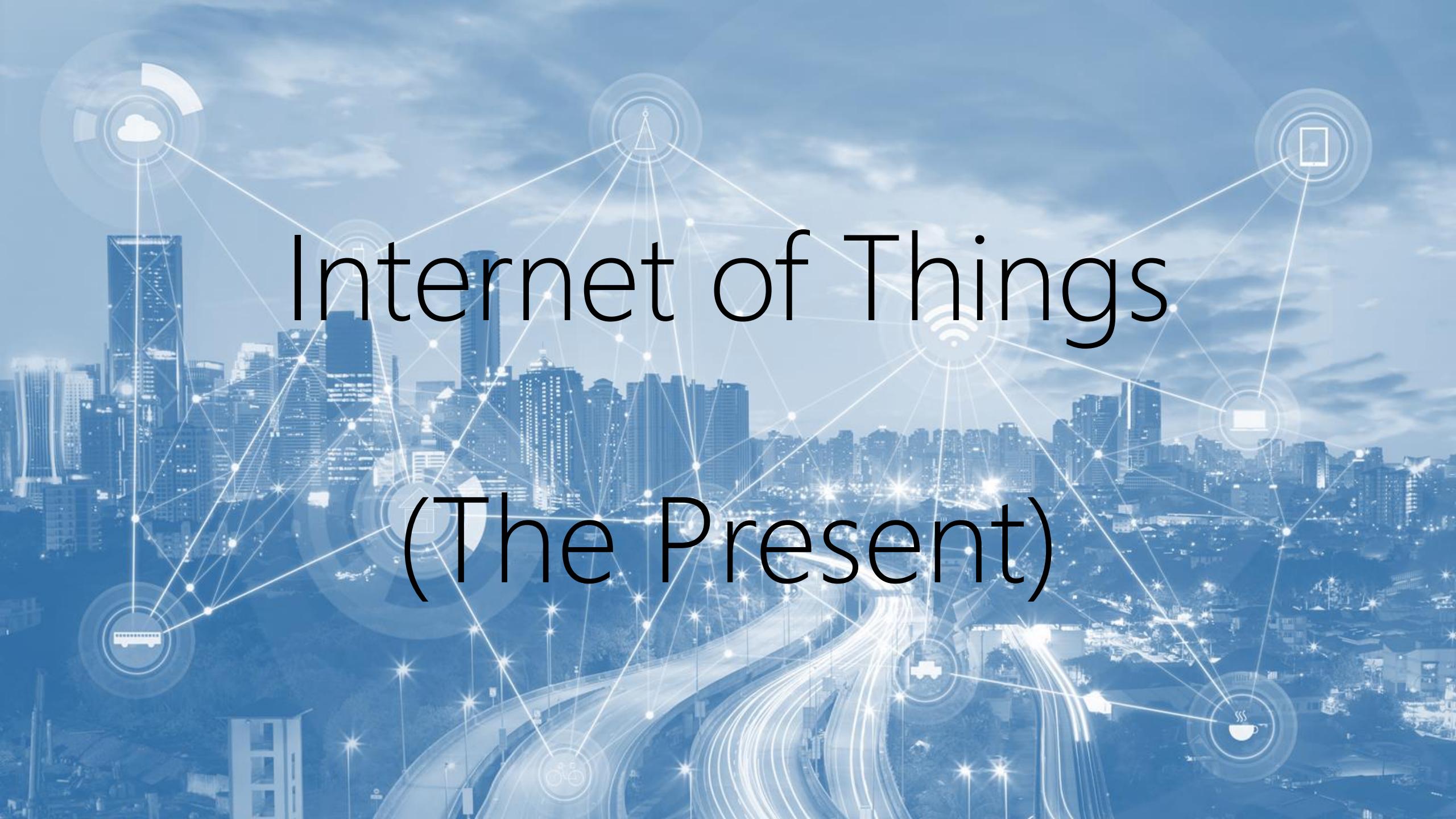
Voice phone: () Ext.

Fax: ()

I would prefer to receive information via

E-mail Fax

Internet

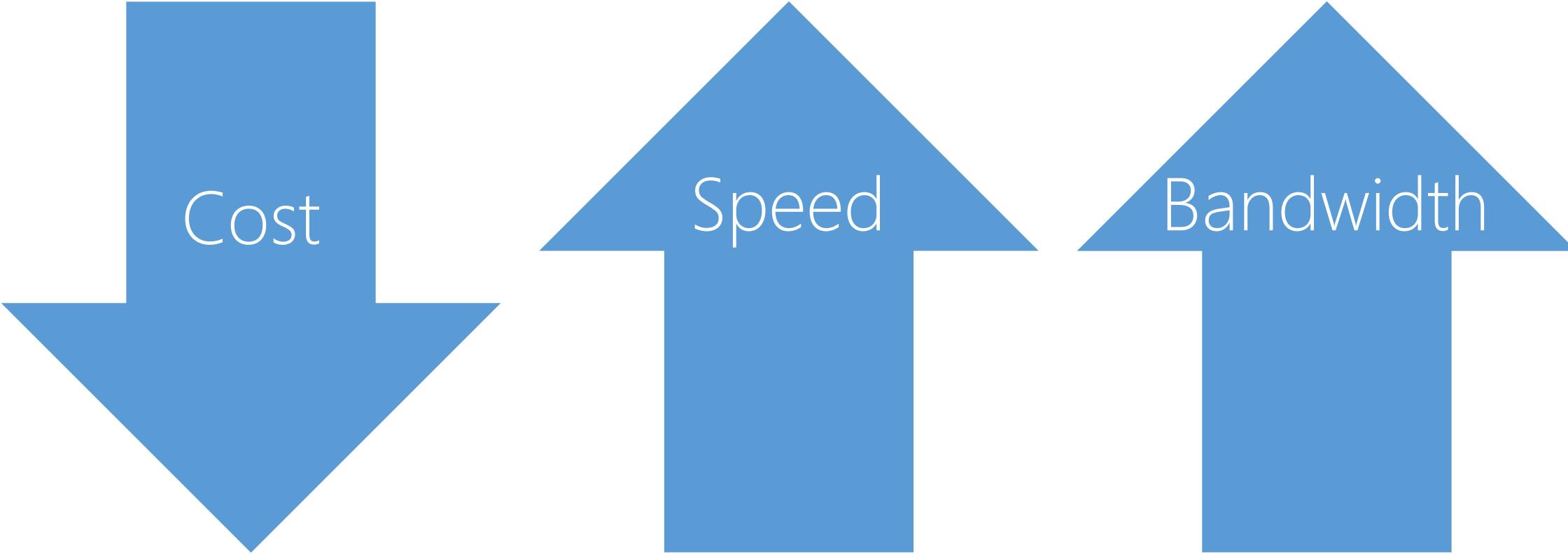


Internet of Things

(The Present)





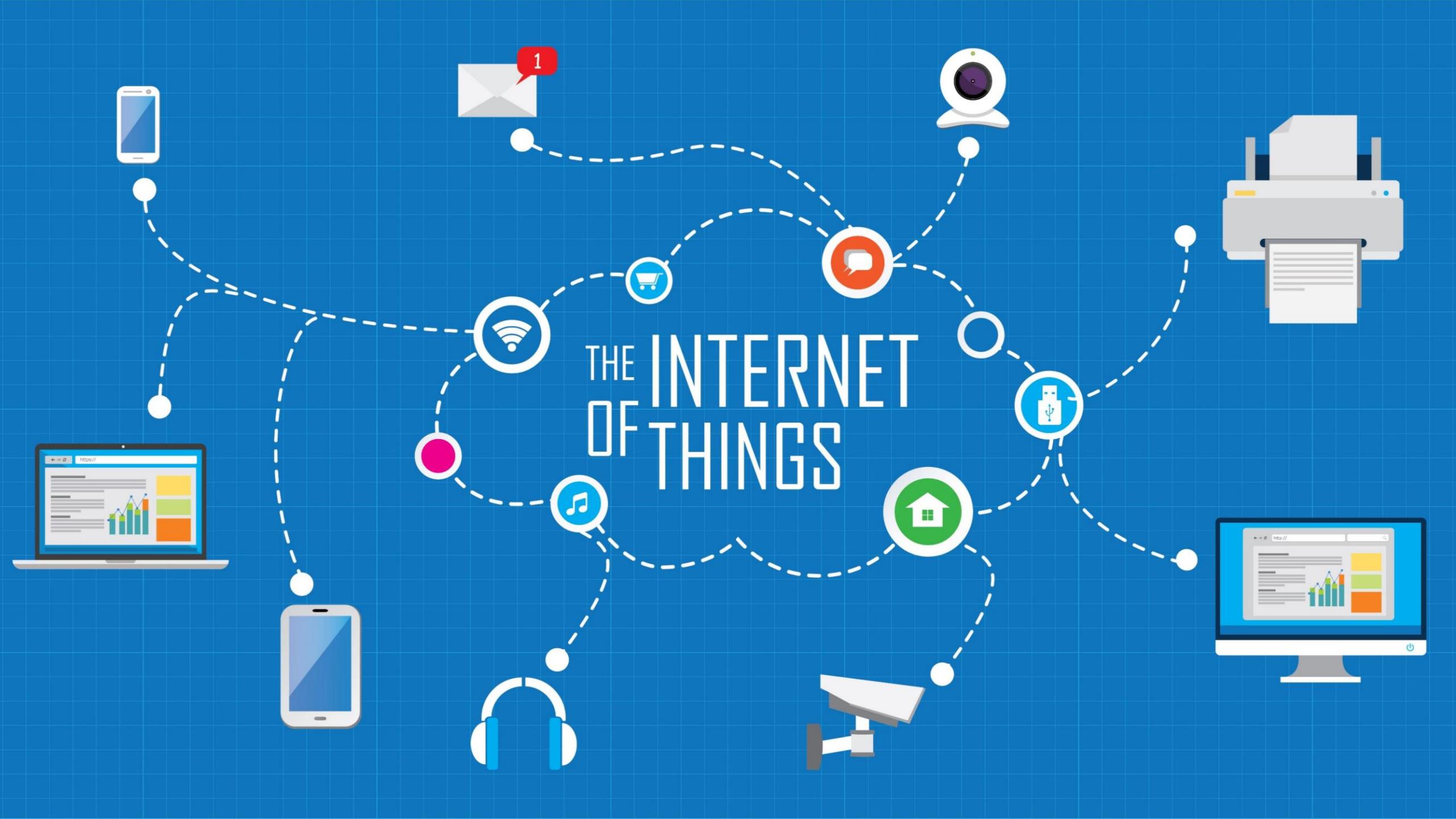


Cost

Speed

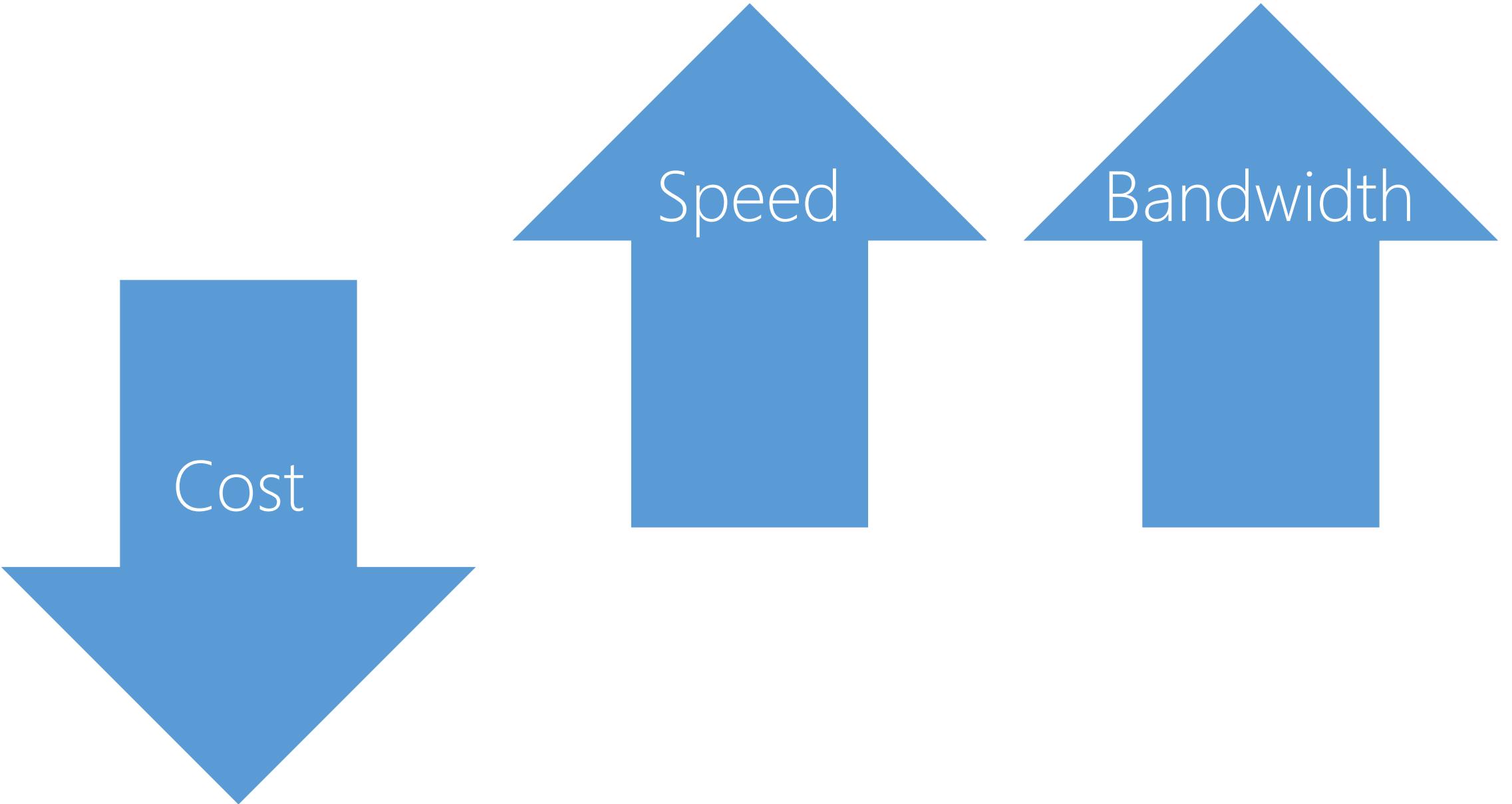
Bandwidth

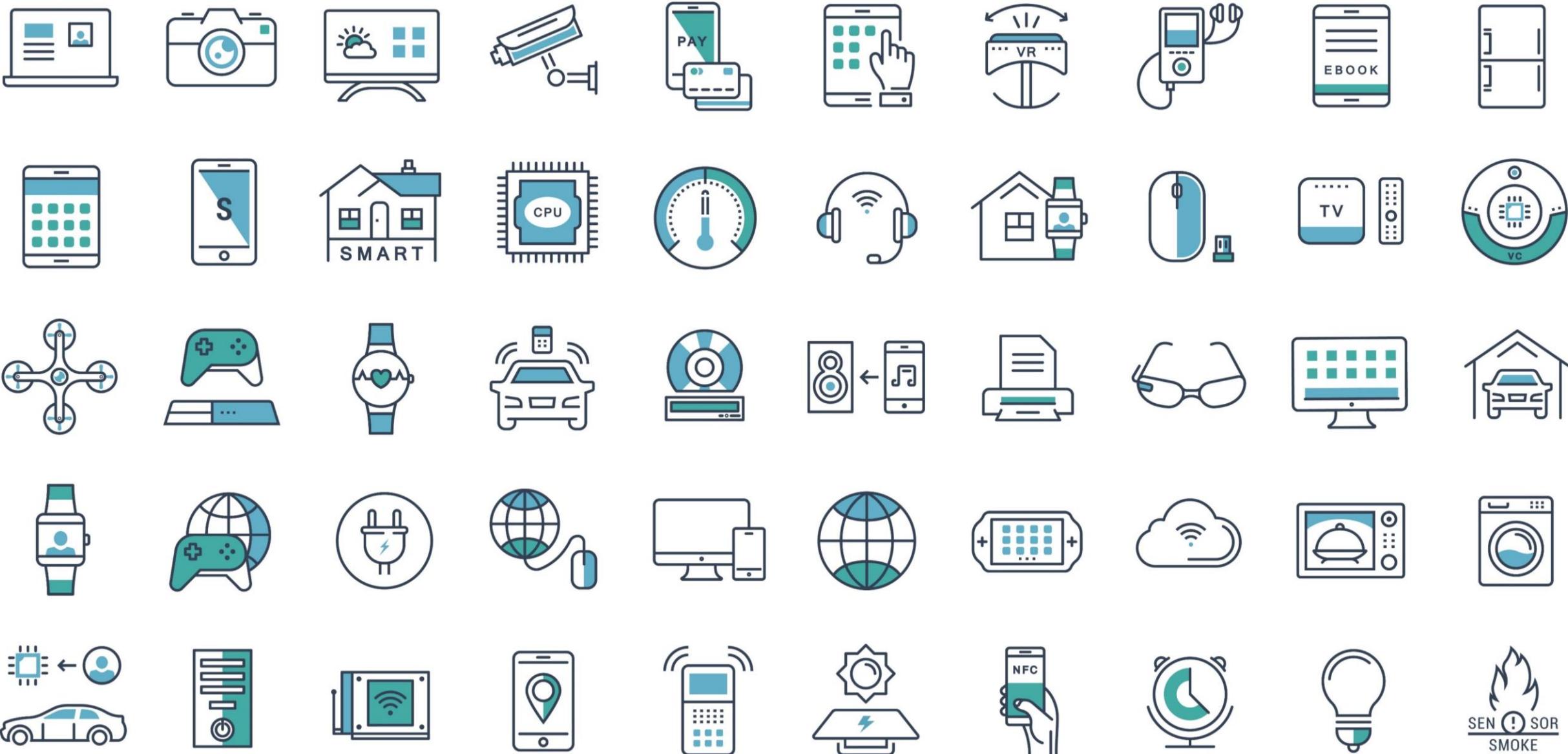
THE INTERNET OF THINGS



The background of the slide features a complex network graph overlaid on a grayscale image of Earth from space. The network consists of numerous small white dots (nodes) connected by thin white lines (edges), forming a dense web of connections across the globe. This visual metaphor represents the global reach and interconnectedness of the Internet.

Internet *is* Everything
(The Future)







FUN GAME 1

Is It IoT?

Is It IoT?



Bluetooth



YES!





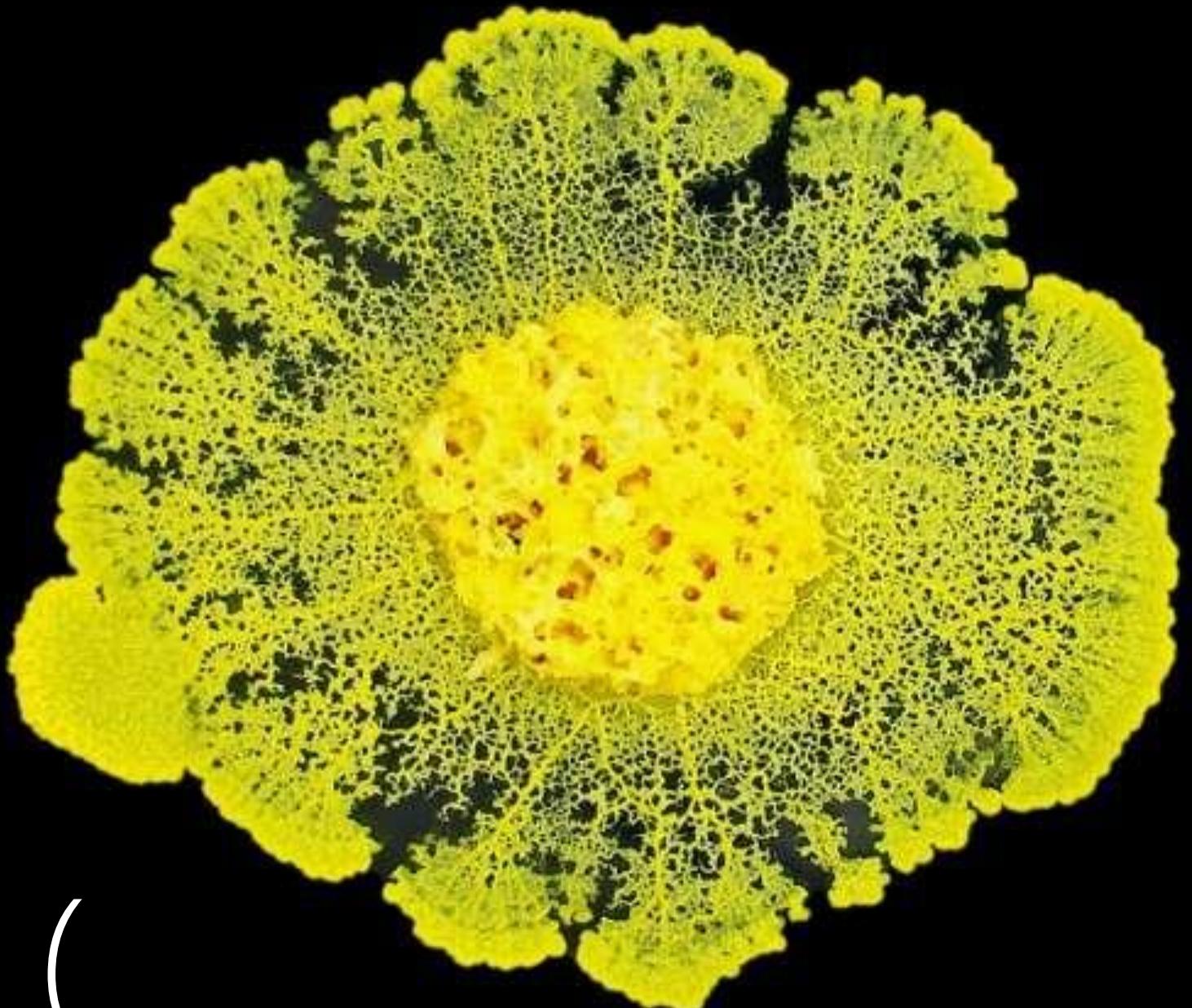
Is It IoT?

YES!





Is It IoT?



NO : (

Data (The Past)

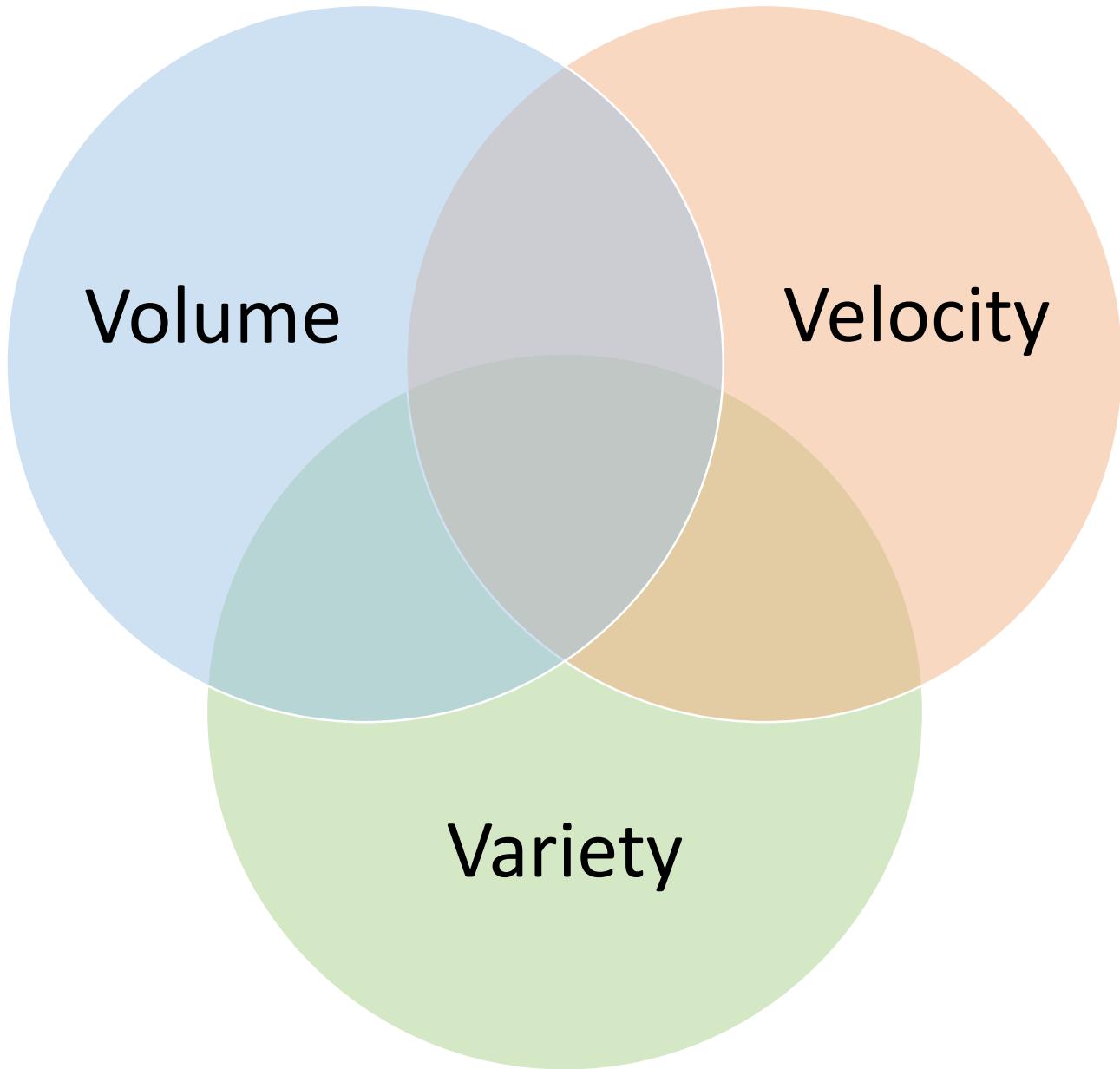






A perspective view of a massive server farm, likely a data center. The image shows numerous rows of tall, blue server racks stretching into the distance. The racks are arranged in a grid pattern, creating a sense of depth. The lighting is bright, reflecting off the metallic surfaces of the racks.

Big Data
(The Present)



A photograph showing a long row of server racks in a data center. The racks are filled with black server units, each with multiple green LED lights indicating active status. The perspective is from a low angle, looking up at the rows of equipment.

Volume

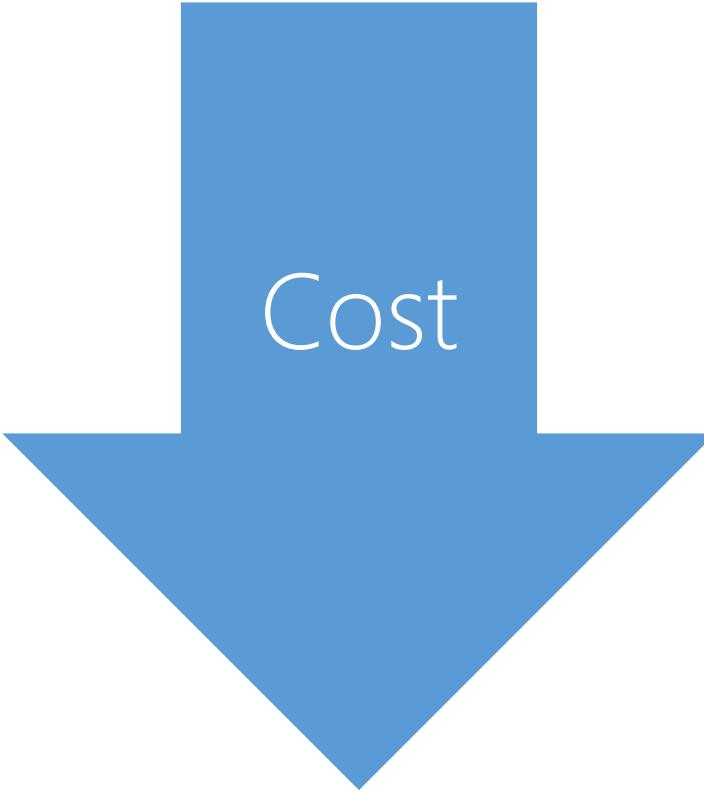
Velocity



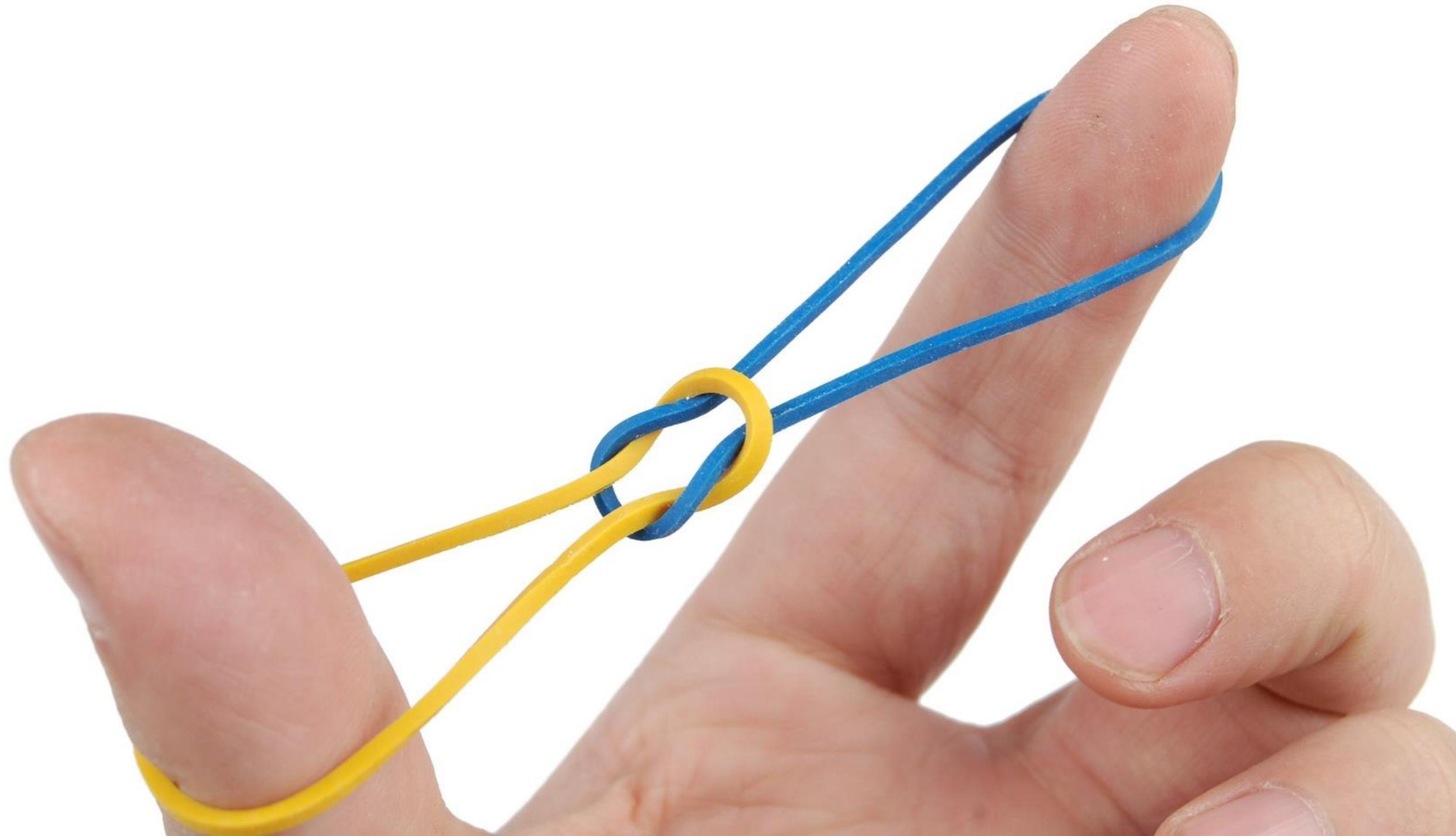
Variety

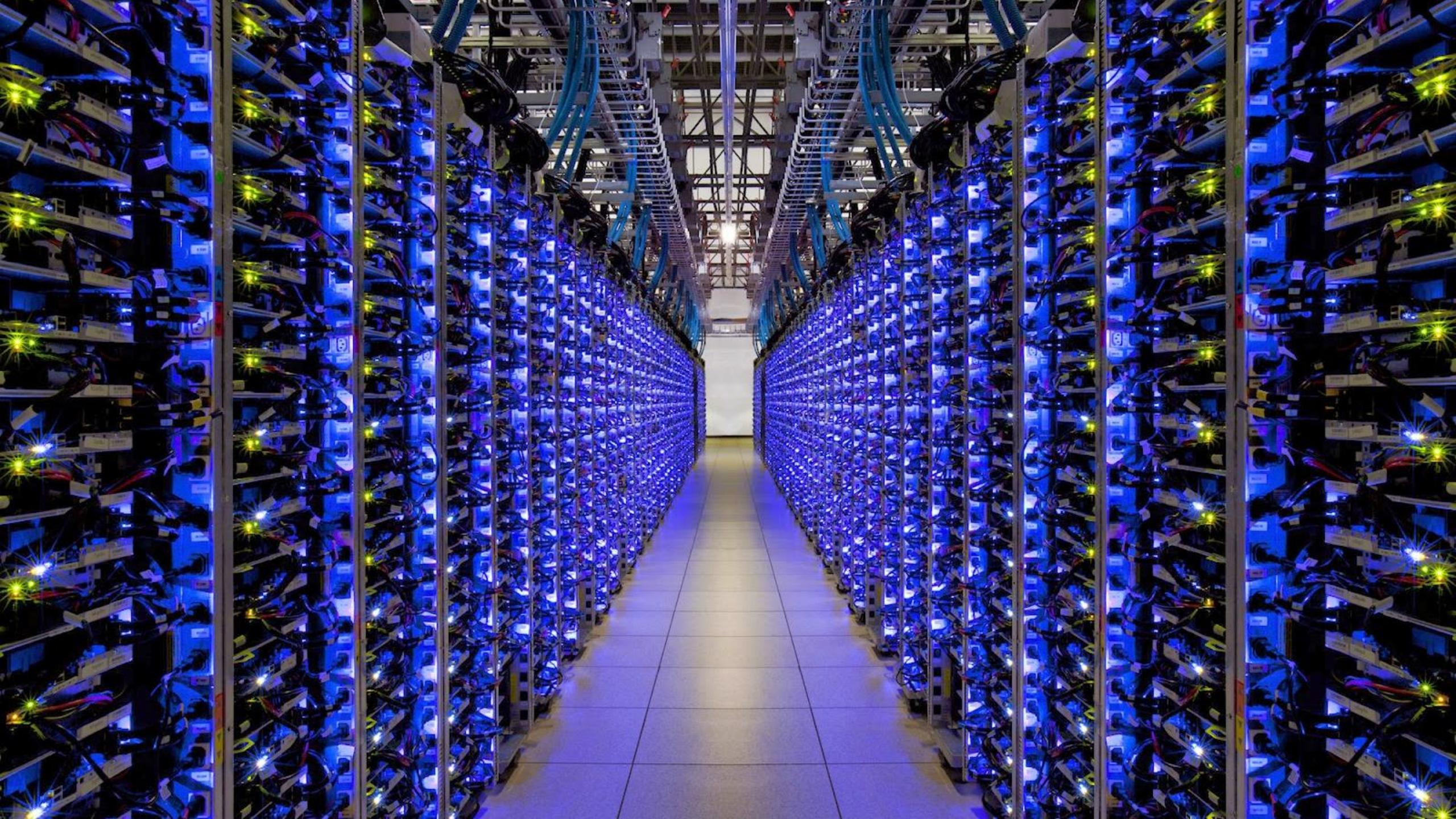


Just Data... Again
(The Future)



Cost







Candy
Cookies/Crackers

Food Vegetables
Goods

Three
Spices Cake Mixes
Jello Puddings

Cosmetics
Household

FUN GAME 2

Am I Smarter Than Big Data?

What Do These Have In Common?





STELLA
ARTOIS

STELLA
ARTOIS

STELLA
ARTOIS

HUGGIES

Natural
Care

HUGGIES

One
&
Done
Refresh
ing

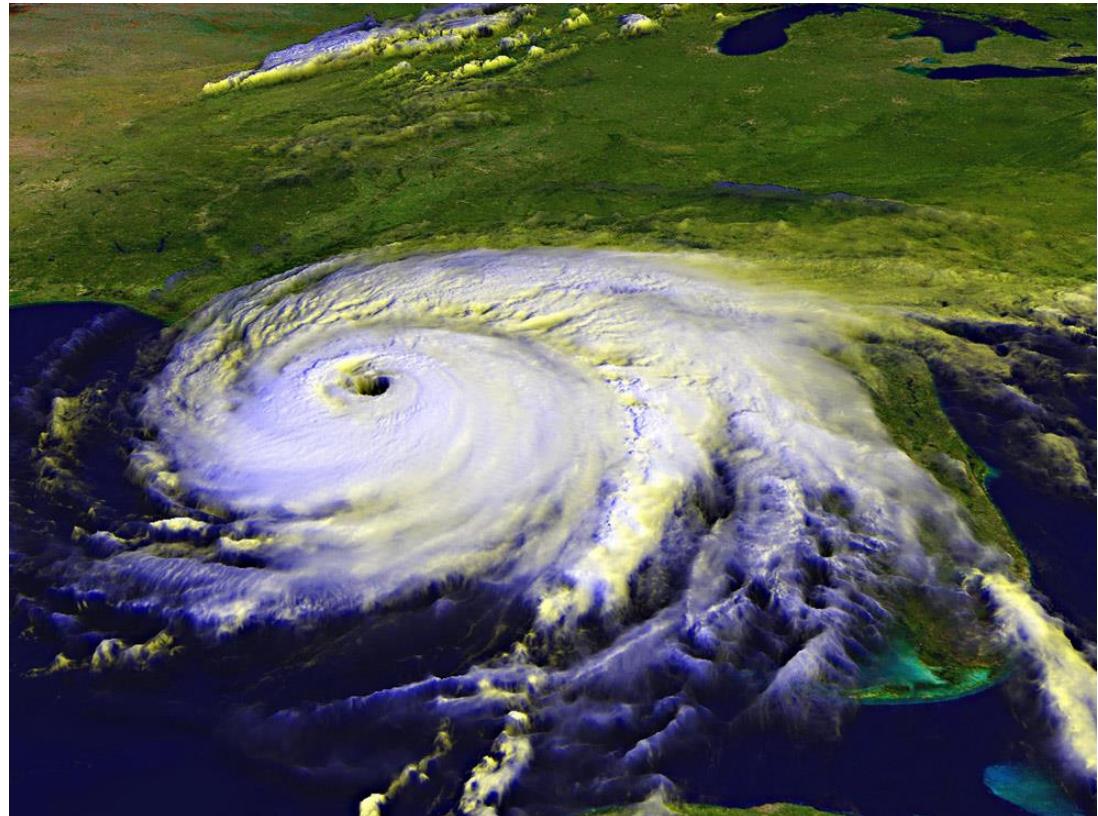
Cute
Snug

What Do These Three Things Predict?



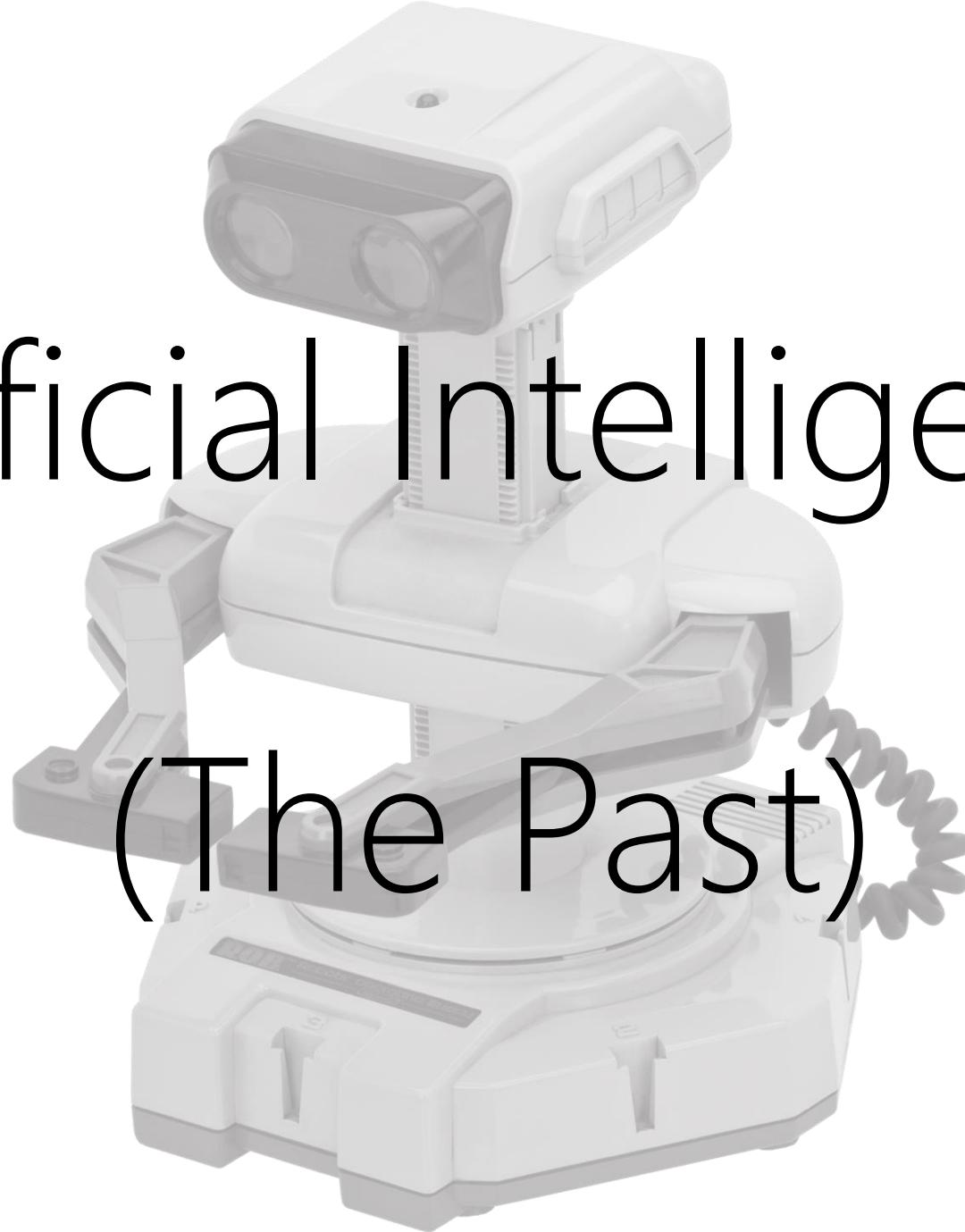


What Do These Have In Common?



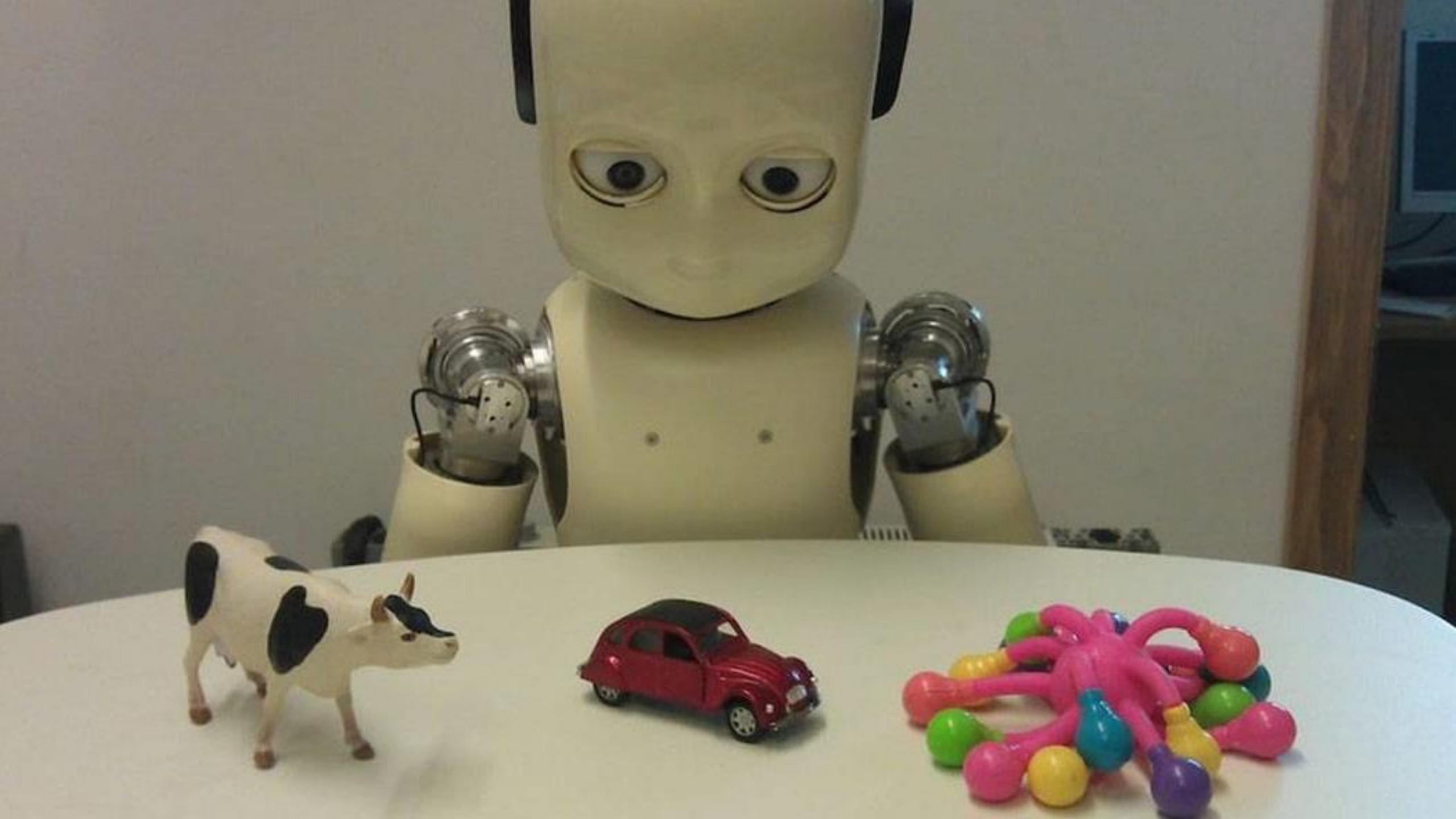


Artificial Intelligence (The Past)

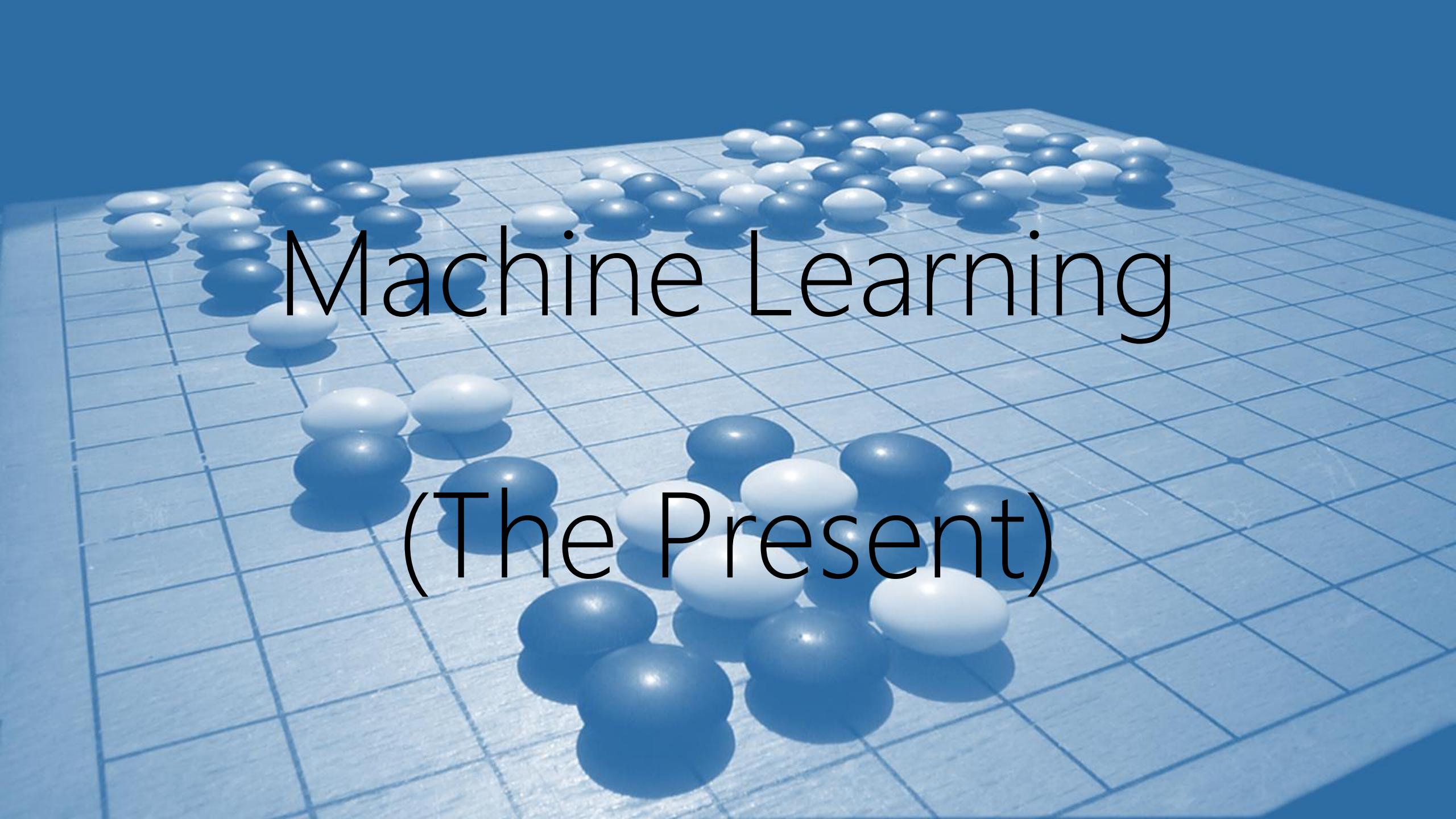






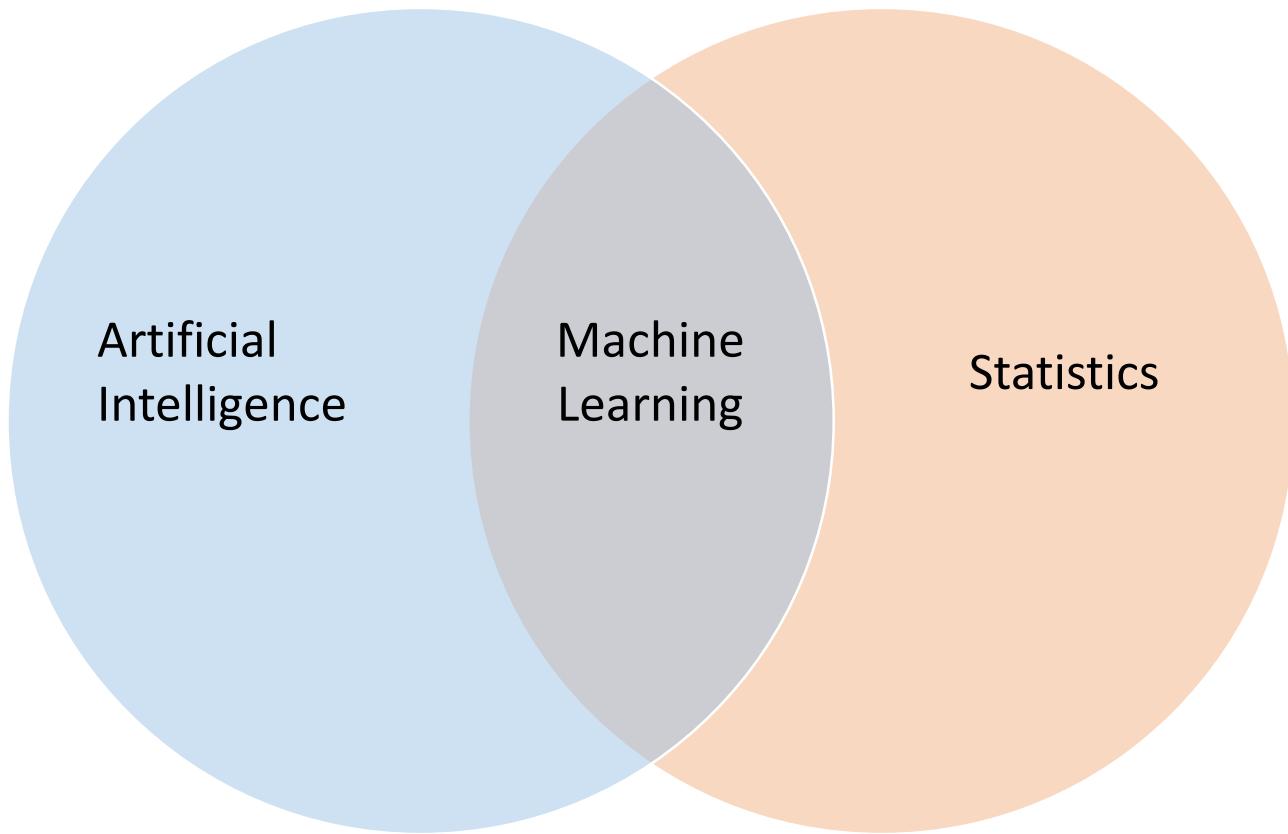


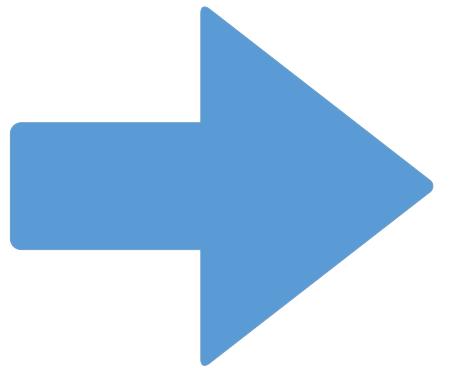
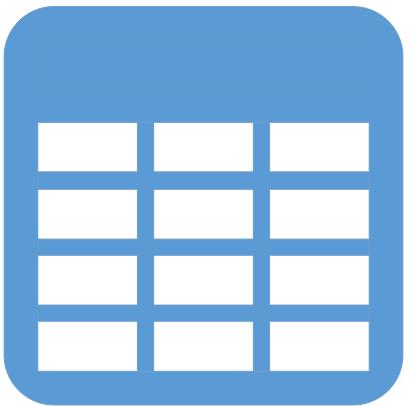
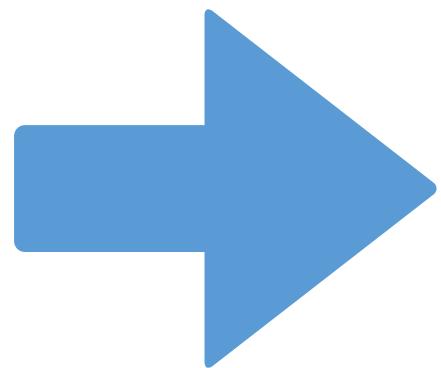


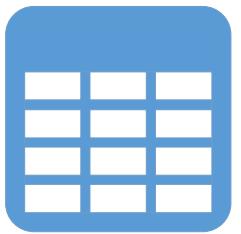
A Go board with blue stones forming a cluster in the center.

Machine Learning

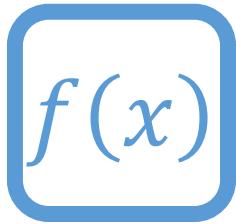
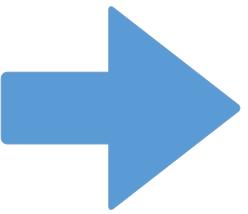
(The Present)



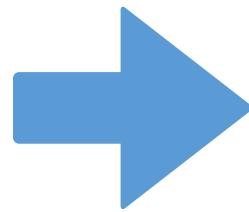
 $f(x)$ 



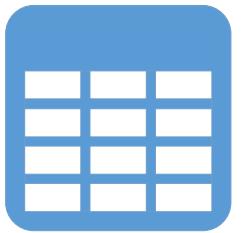
Data



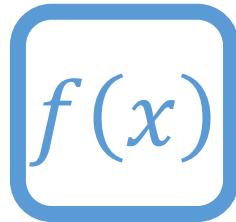
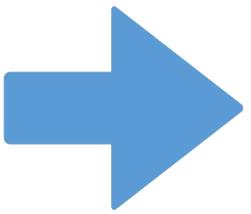
Function



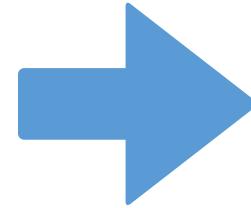
Prediction



Data



Function



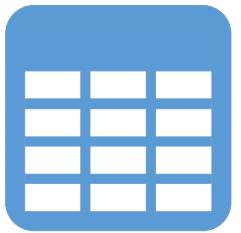
Prediction



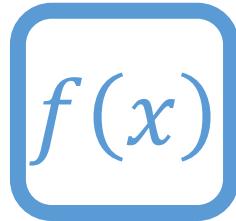
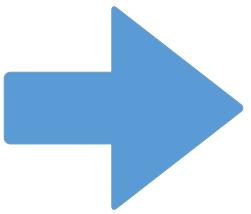
Cat



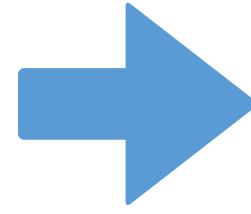
Dog



Data



Function



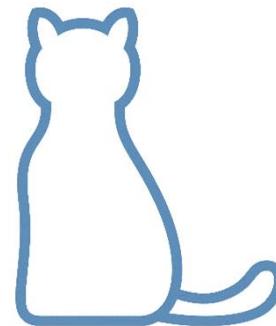
Prediction



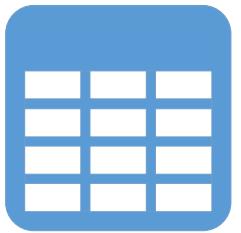
Cat



Dog



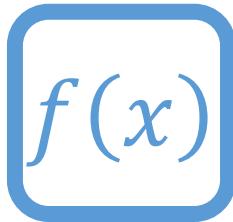
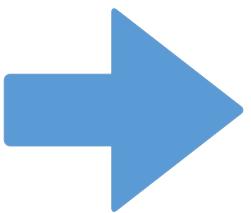
Is cat?



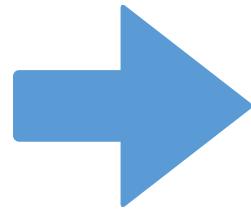
Data



Cat



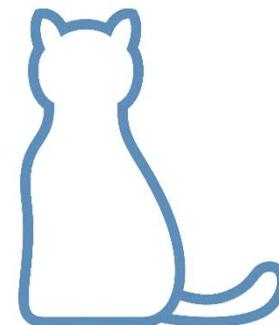
Function



Prediction



Dog



Is cat?



Yes

Franke hood

Wall cabinet

Refrigerator

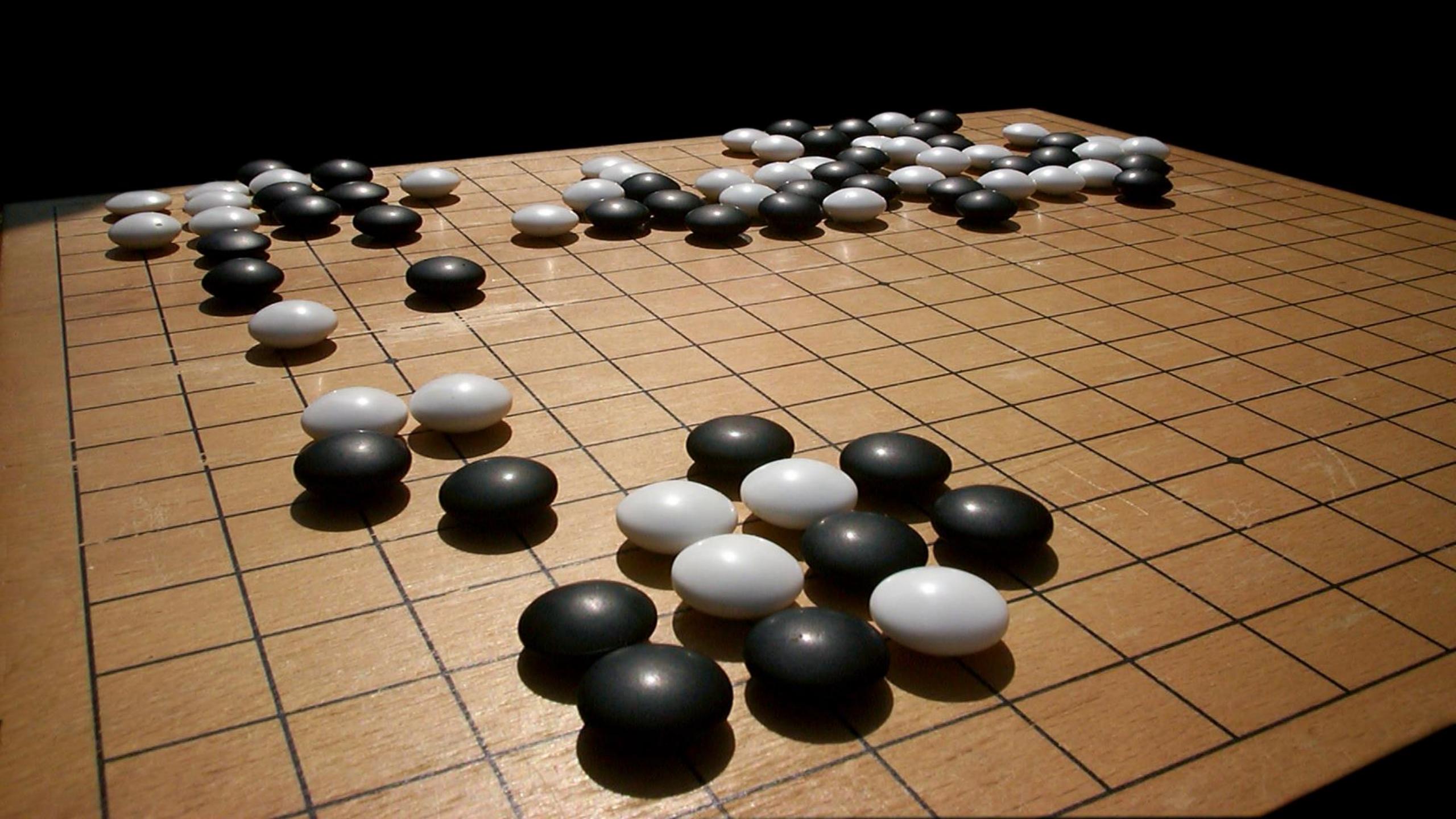
Range

Pot

Kettle

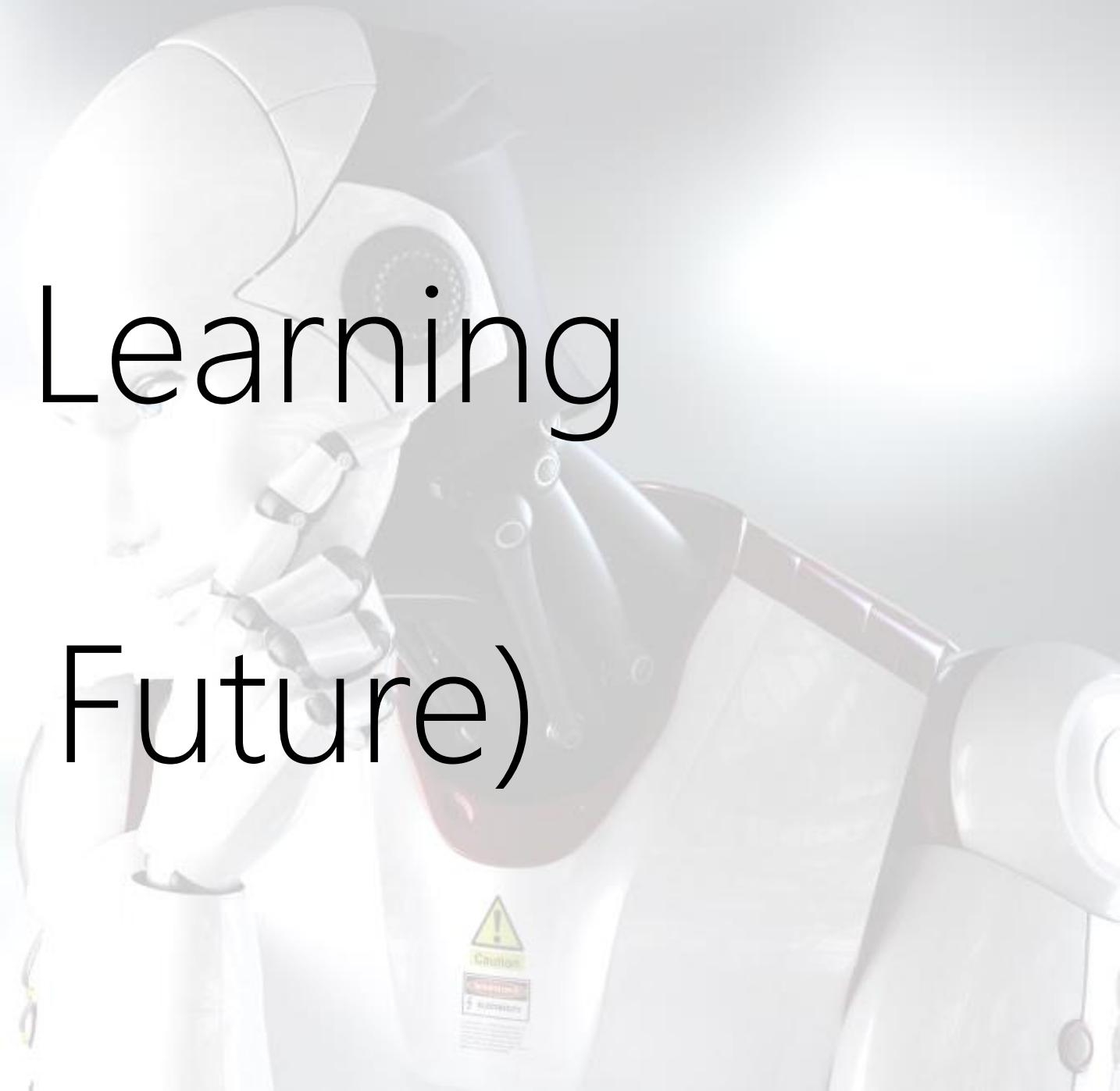
Teapot

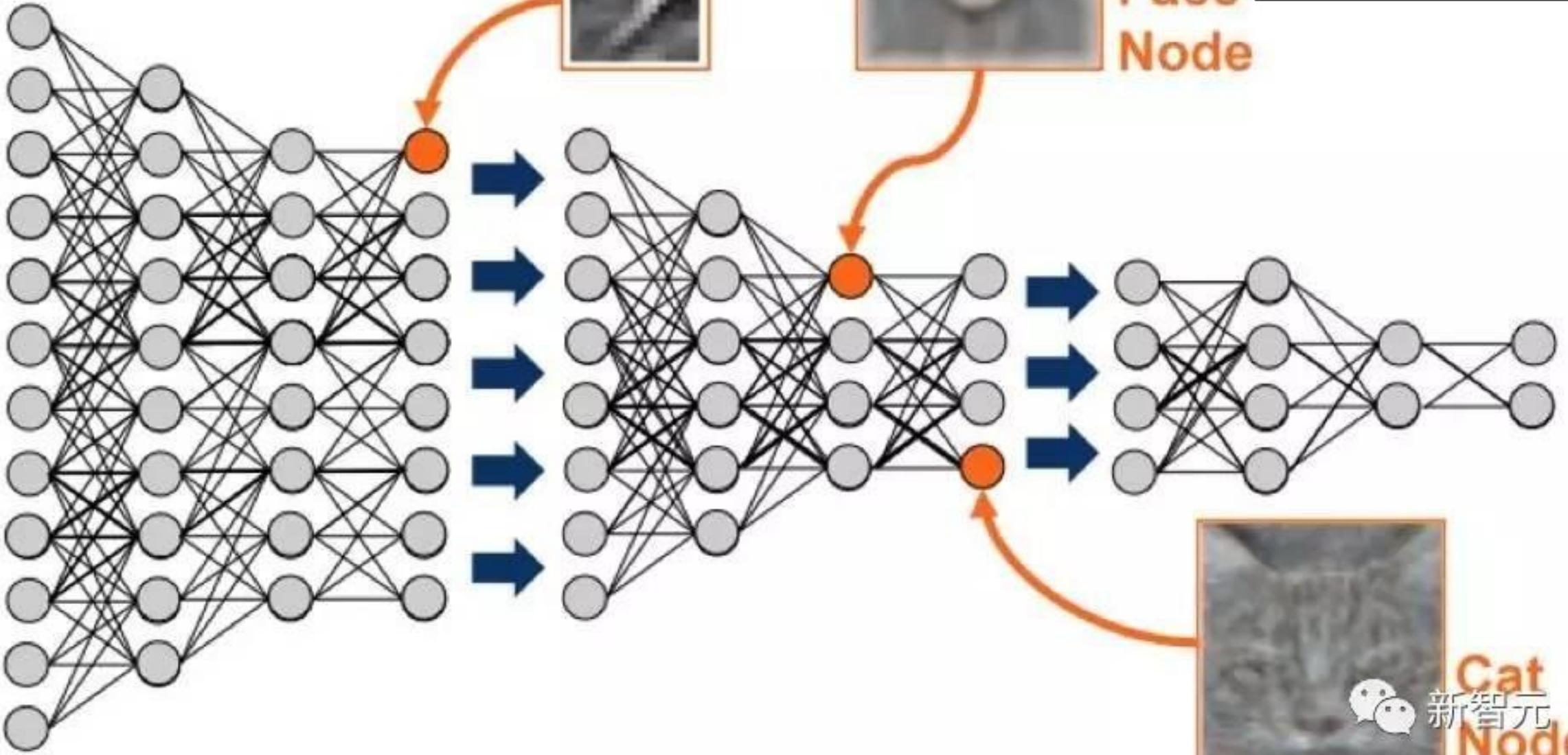
Cabinet table

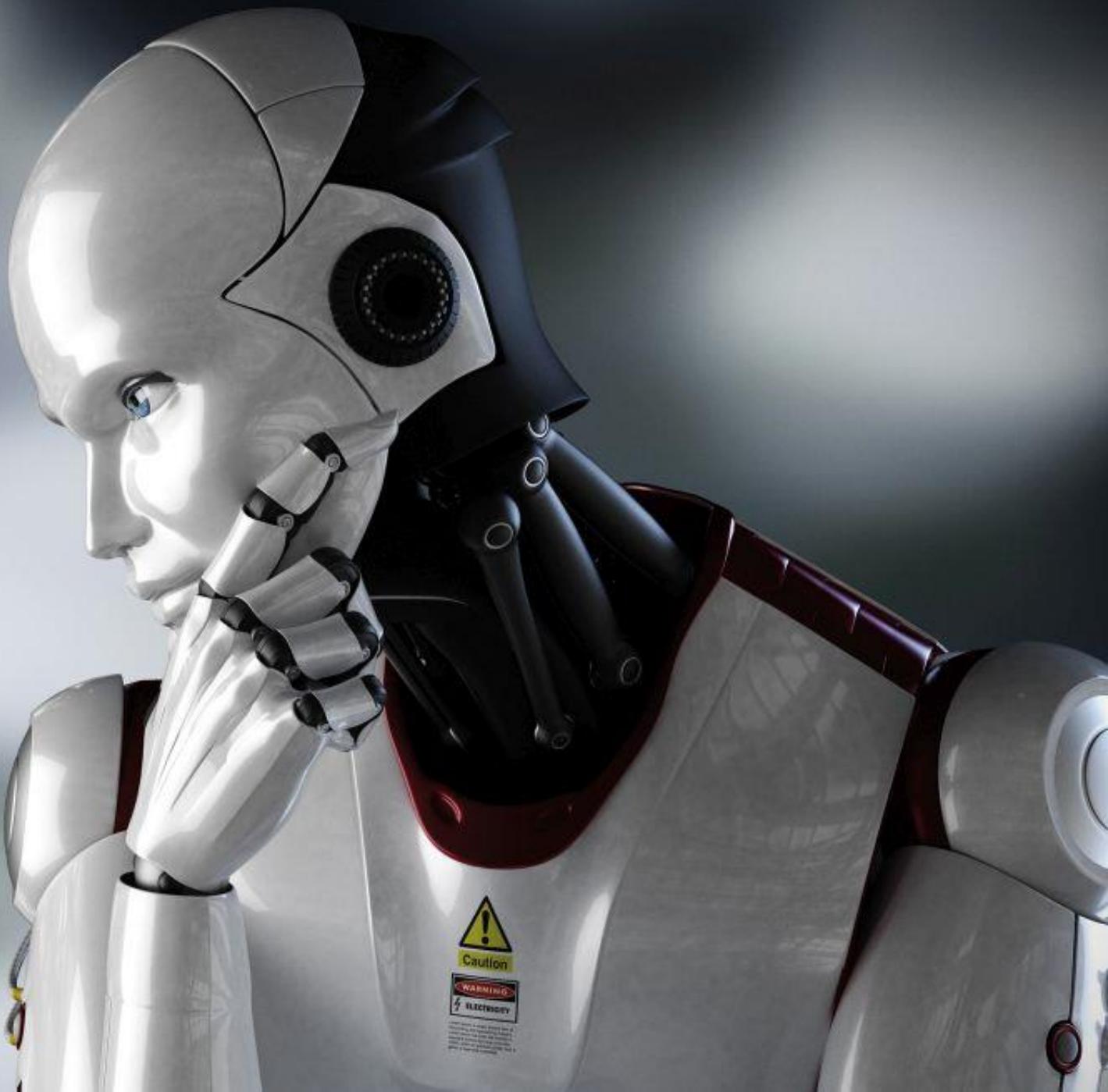


Deep Learning

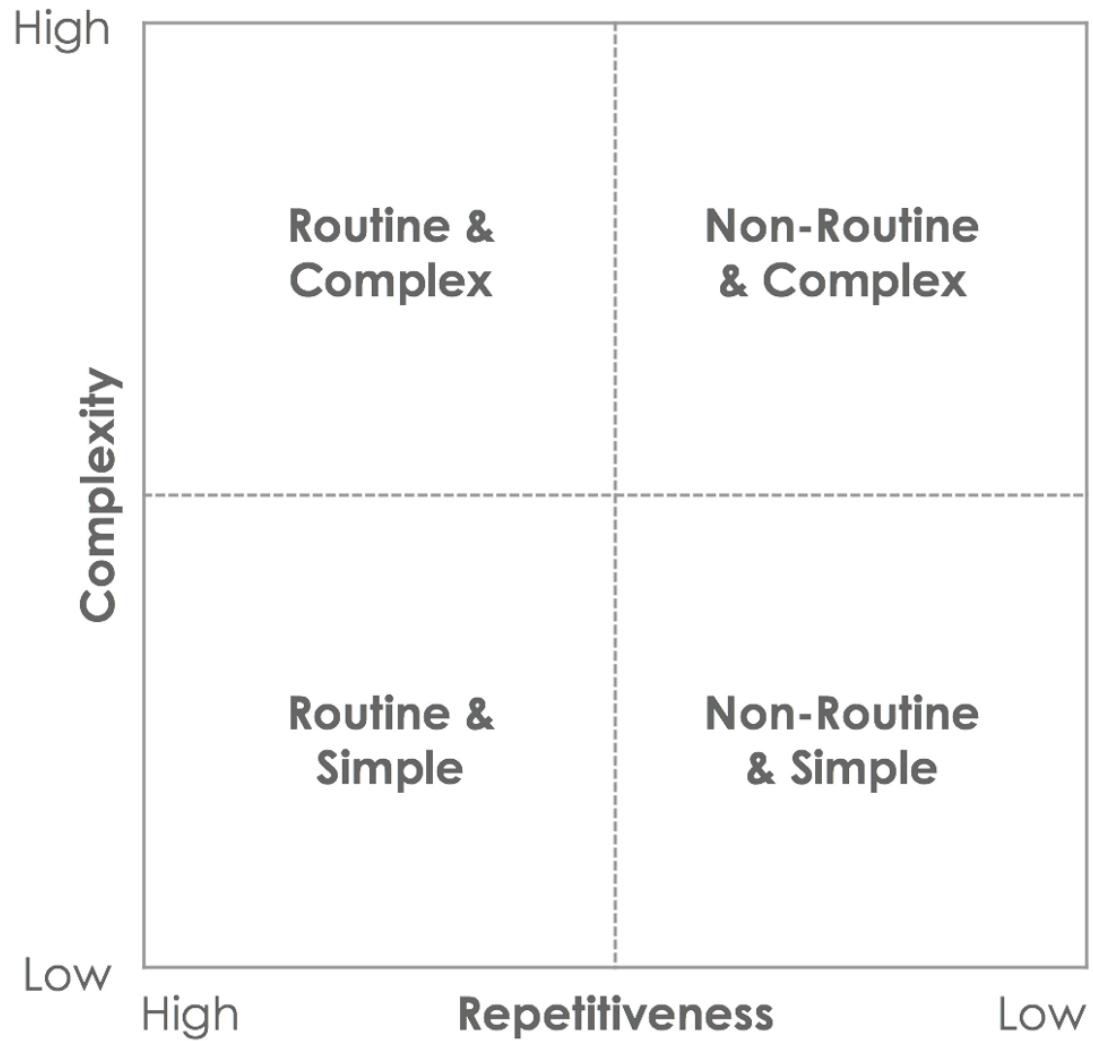
(The Future)



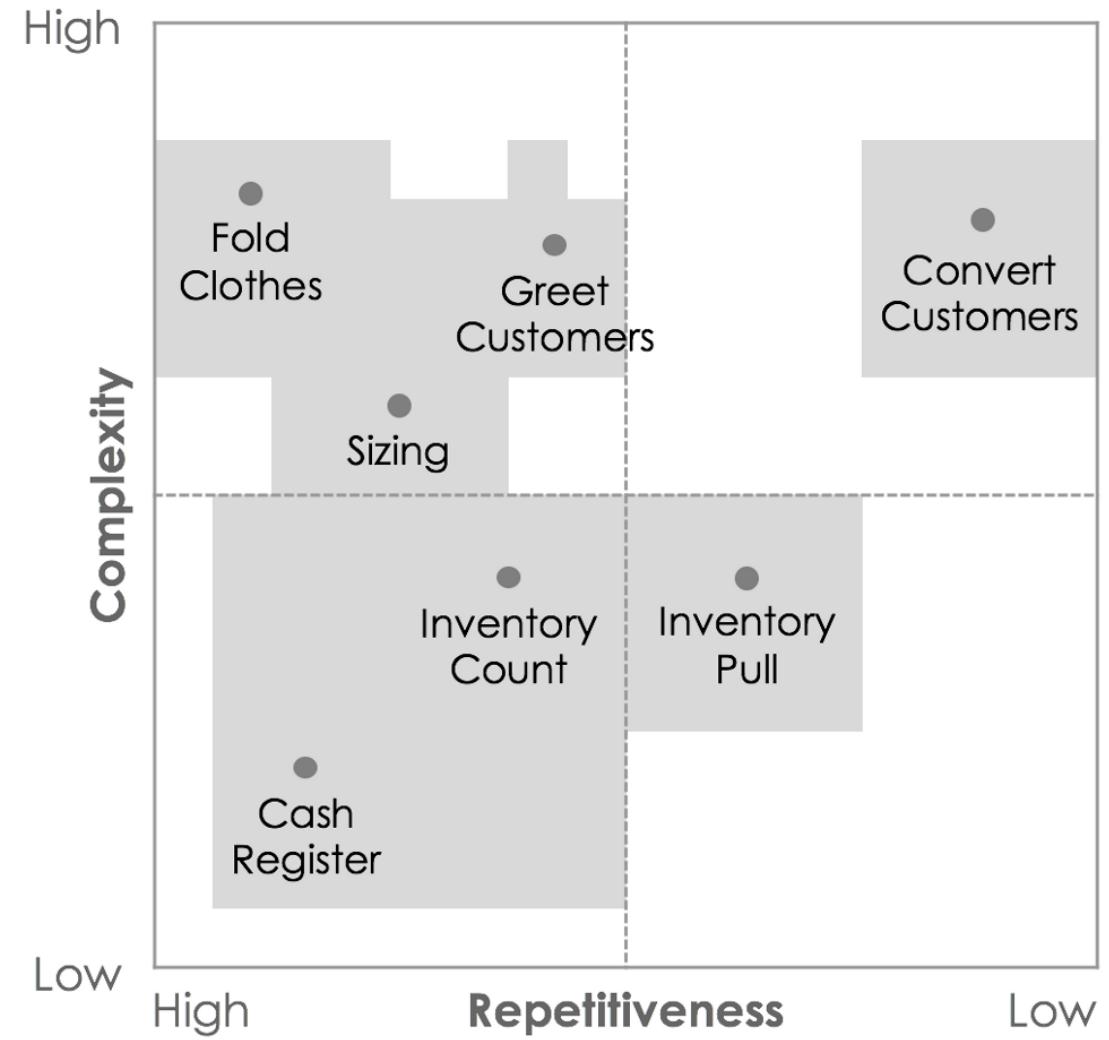




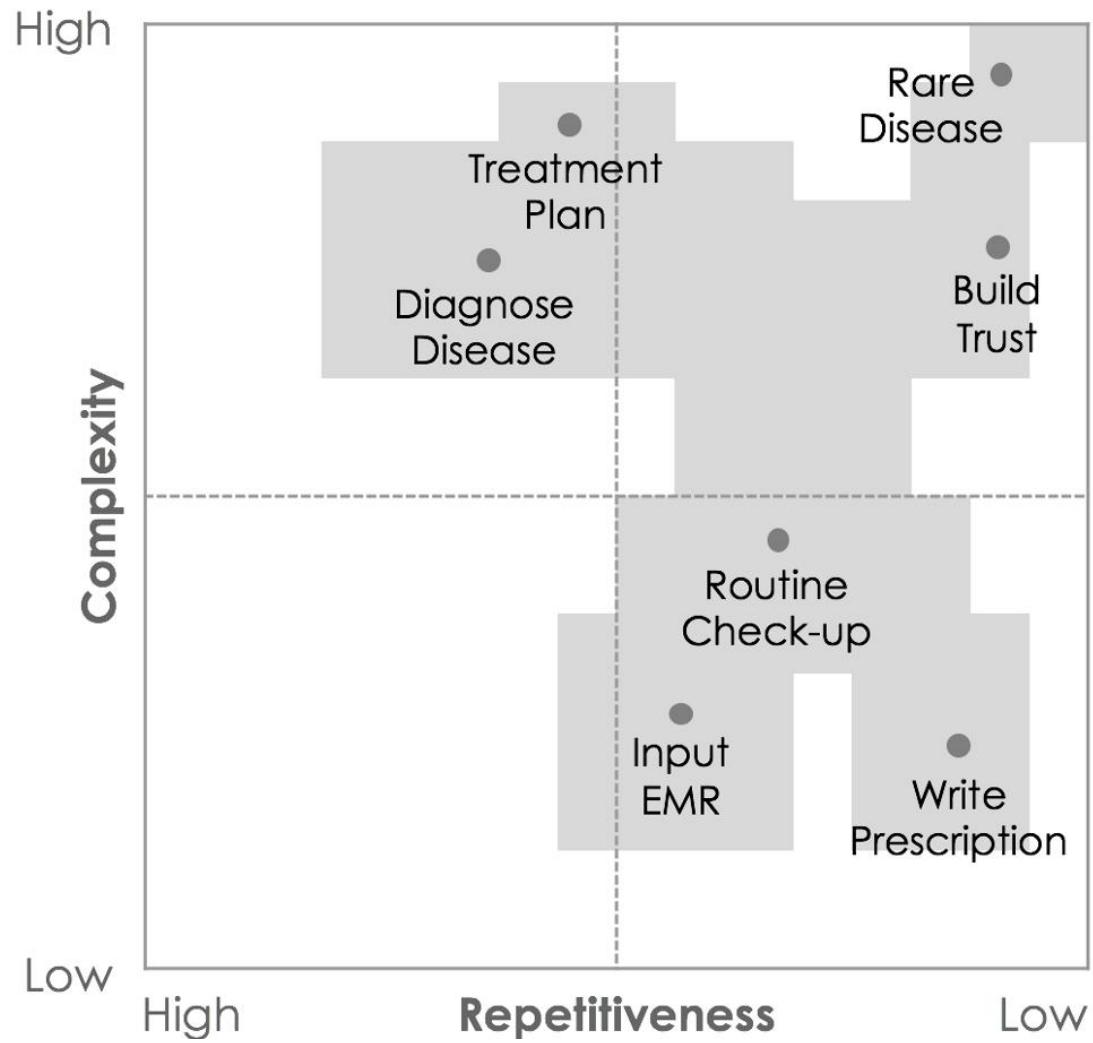
A. The Automation Framework



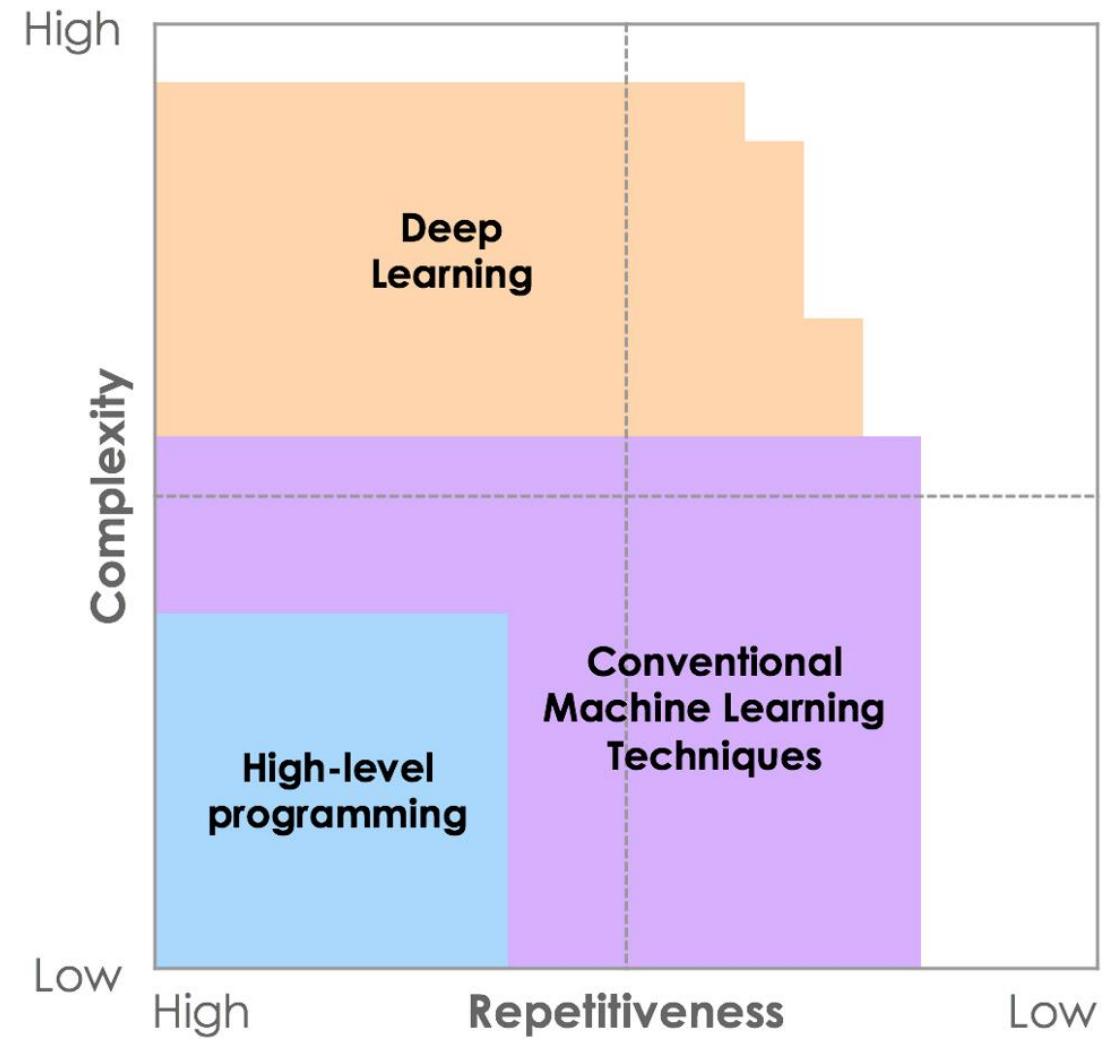
B. Retail Salesperson



C. Medical Doctor



D. Automation Technology





FUN GAME 3

Dog or Mop?



DOG!



MOP!

CAUTION



CUIDADO



CAUTION



DOG!



MOP!

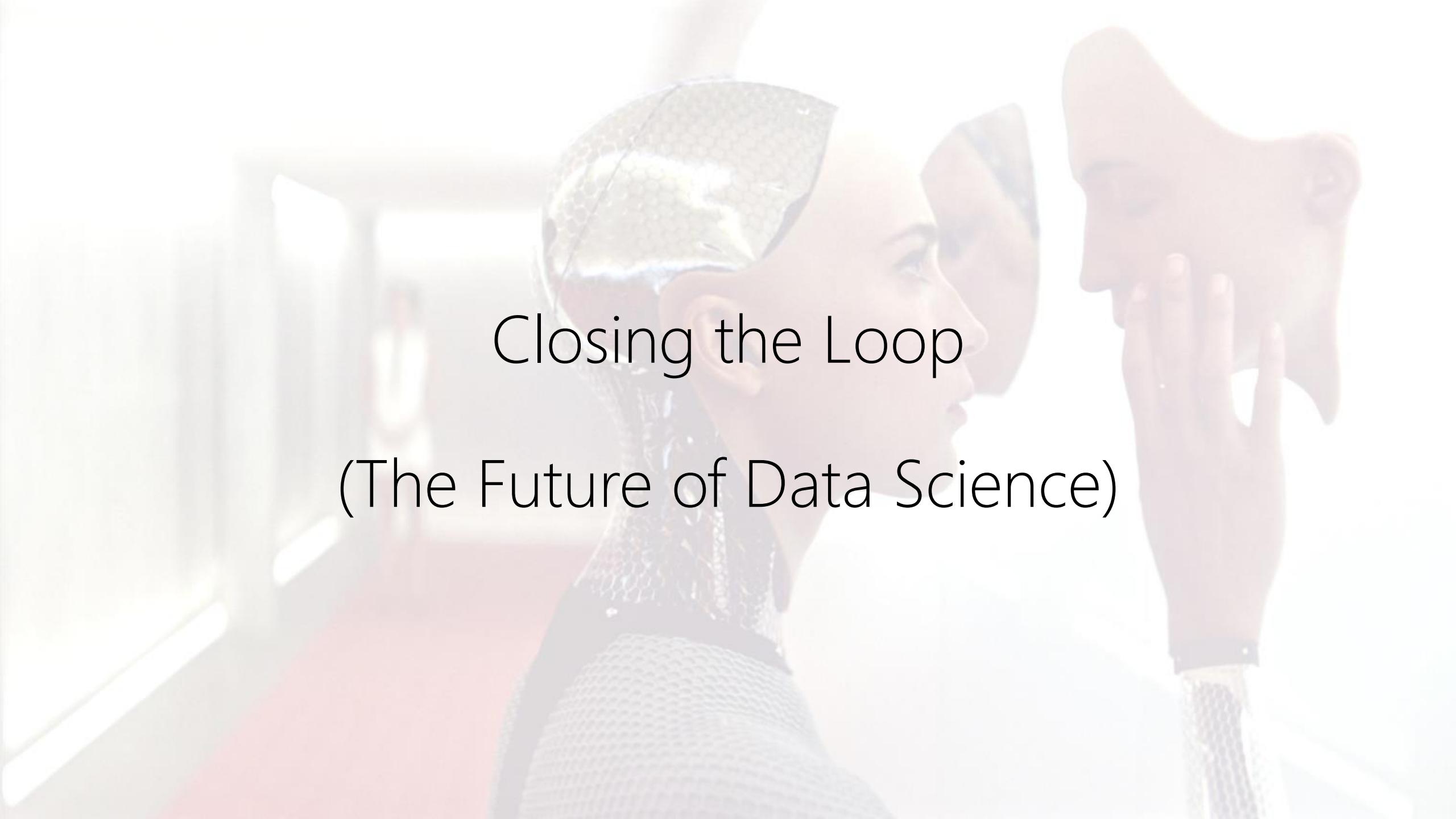
MOP!



DOG!



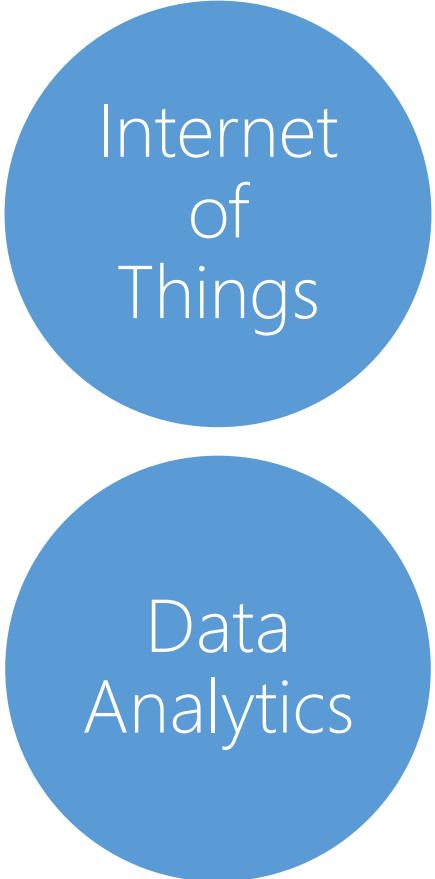


A person wearing a VR headset and holding a controller, set against a blurred background of a room with a sofa and a lamp.

Closing the Loop

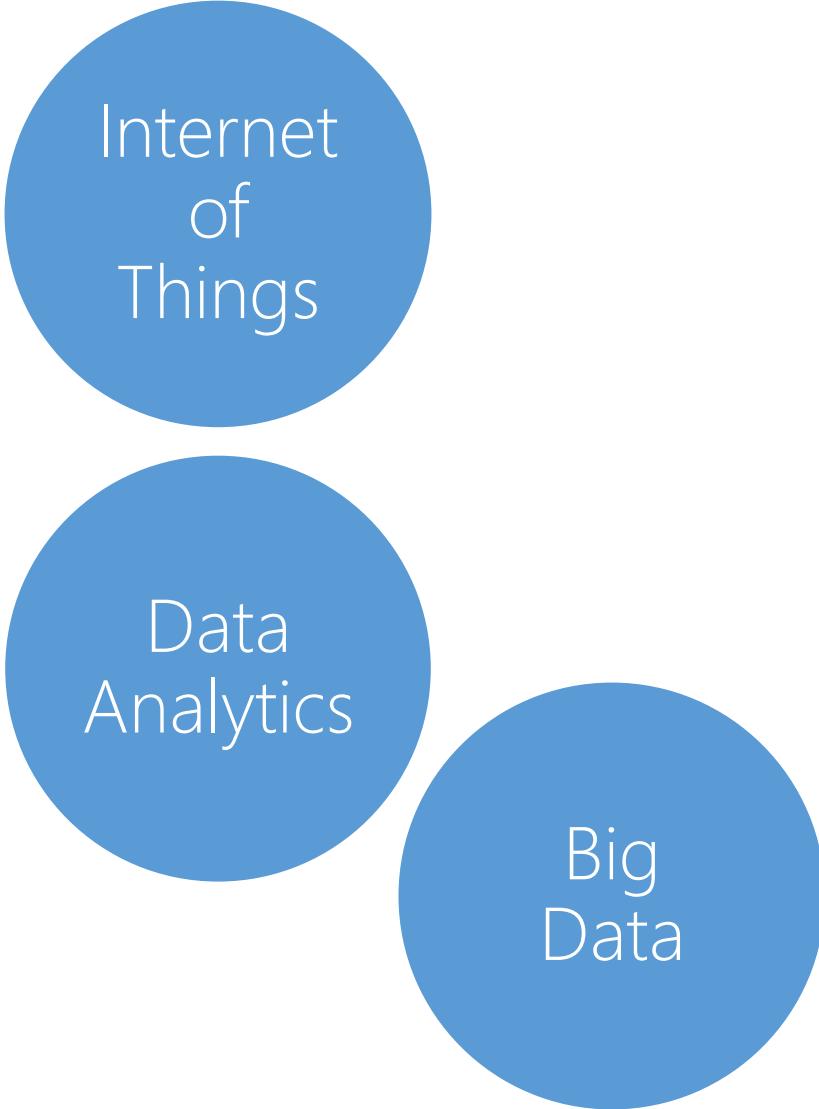
(The Future of Data Science)

Data
Analytics



Internet
of
Things

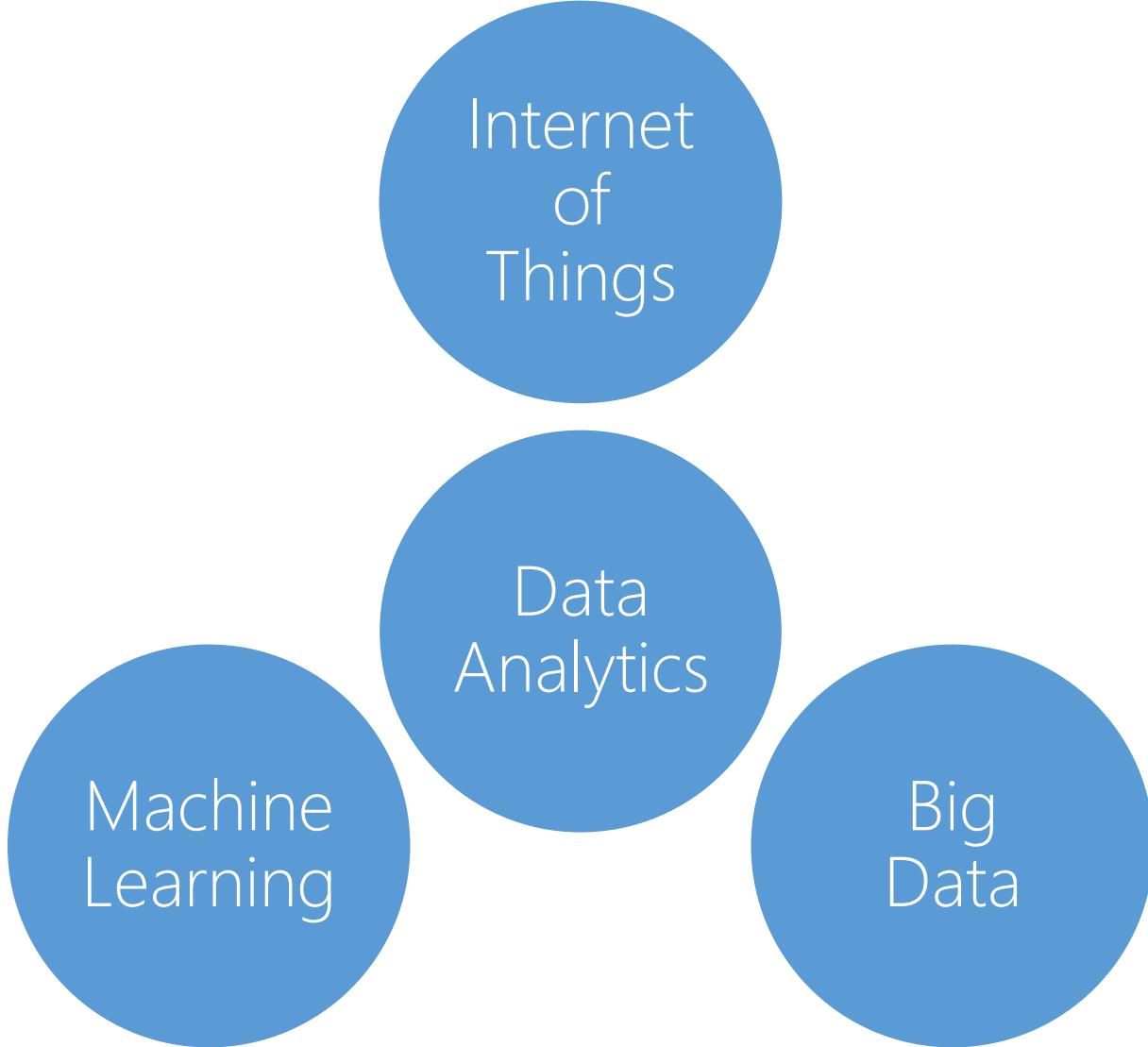
Data
Analytics



Internet
of
Things

Data
Analytics

Big
Data

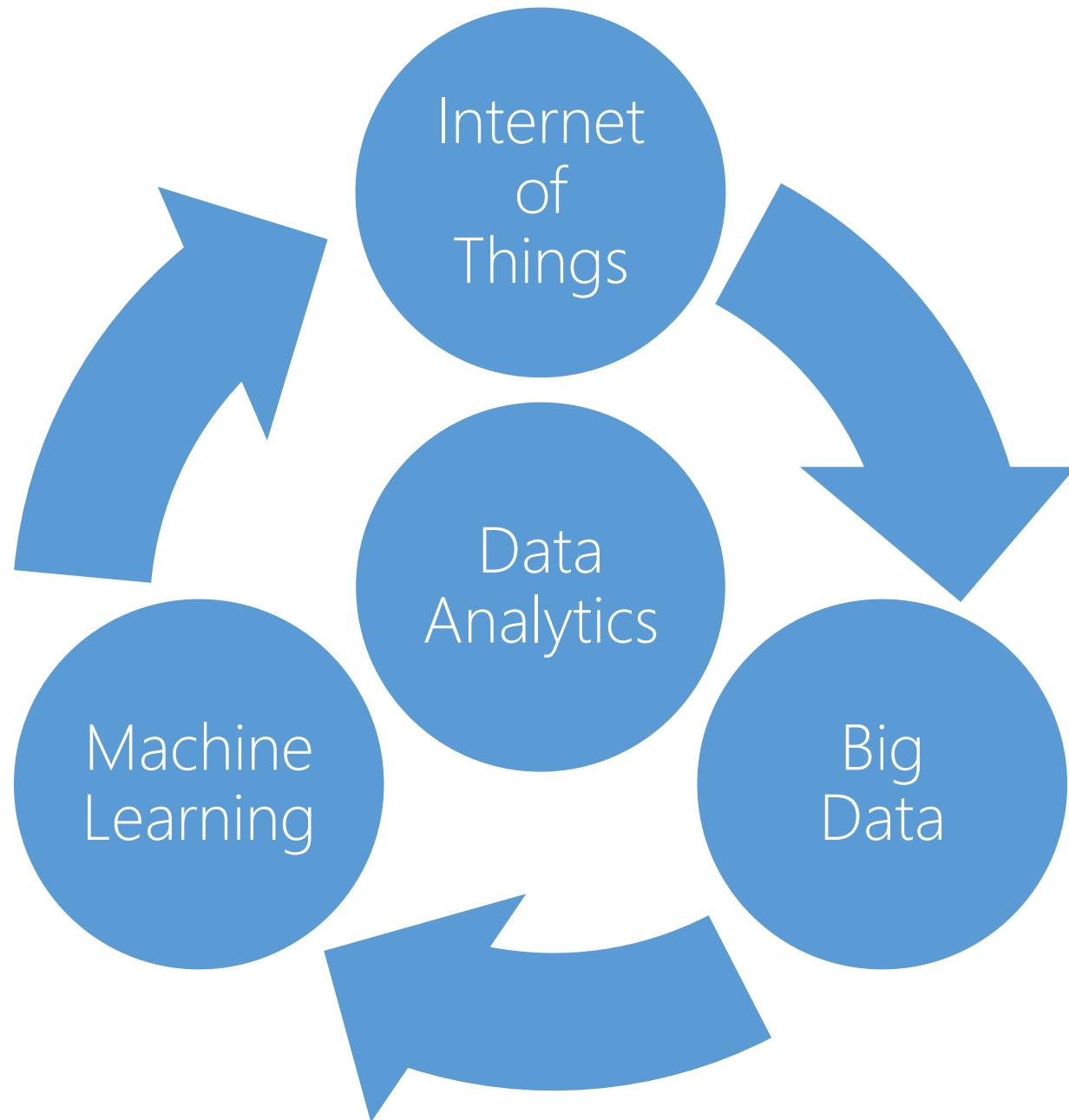


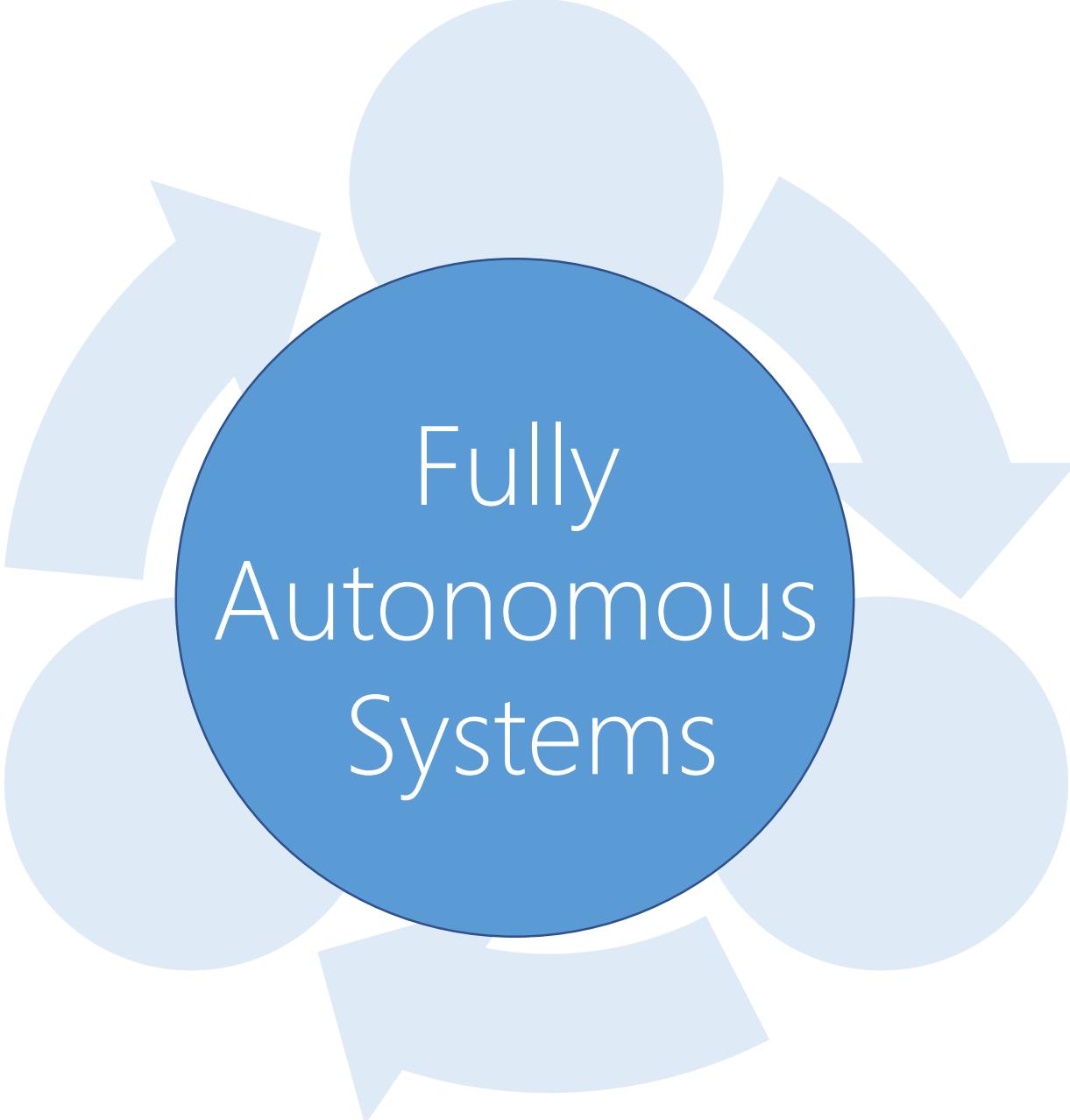
Internet
of
Things

Data
Analytics

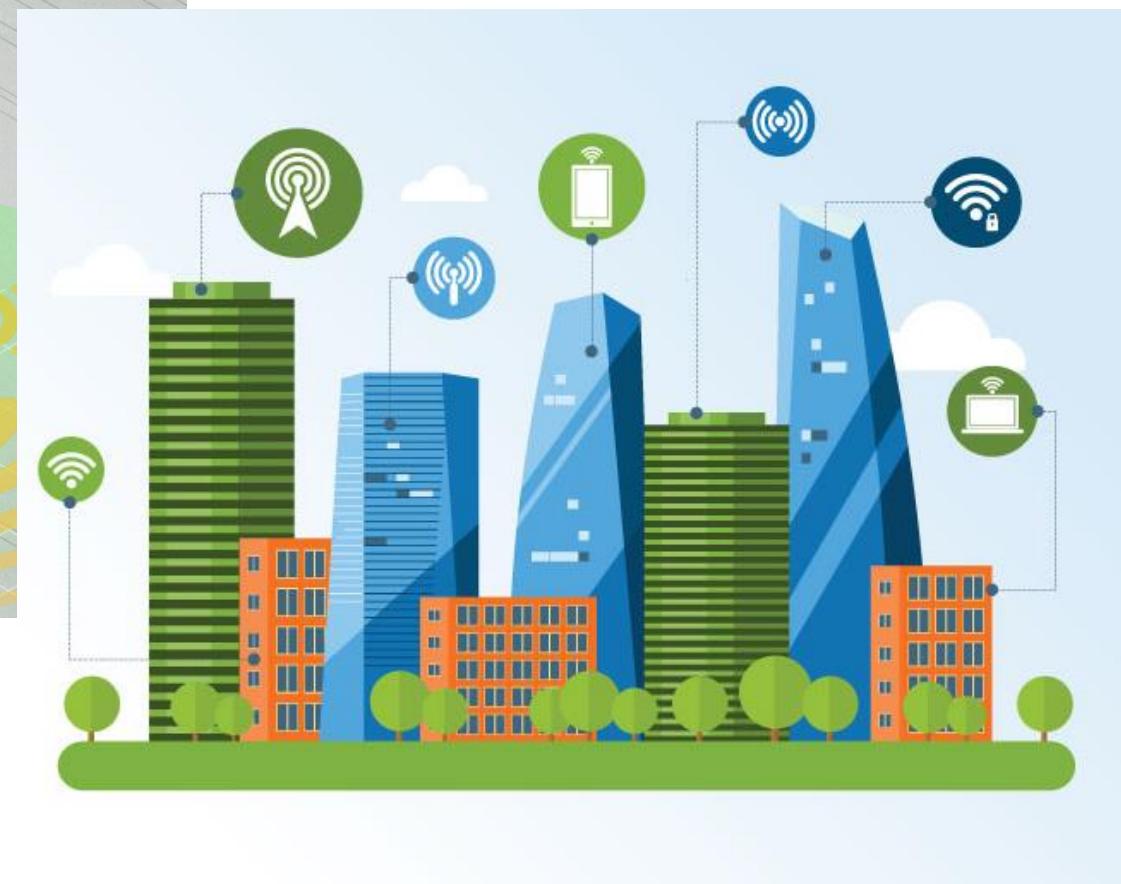
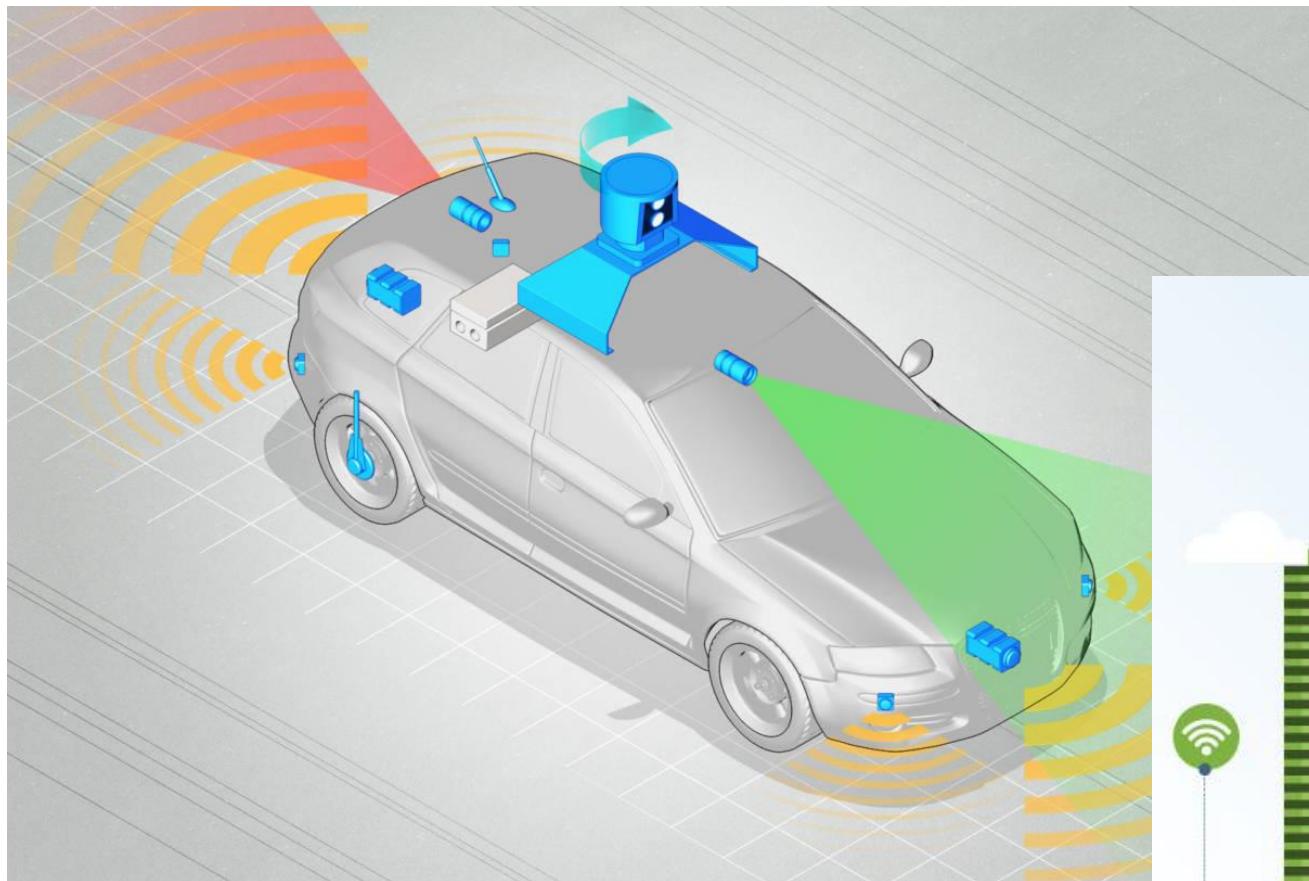
Machine
Learning

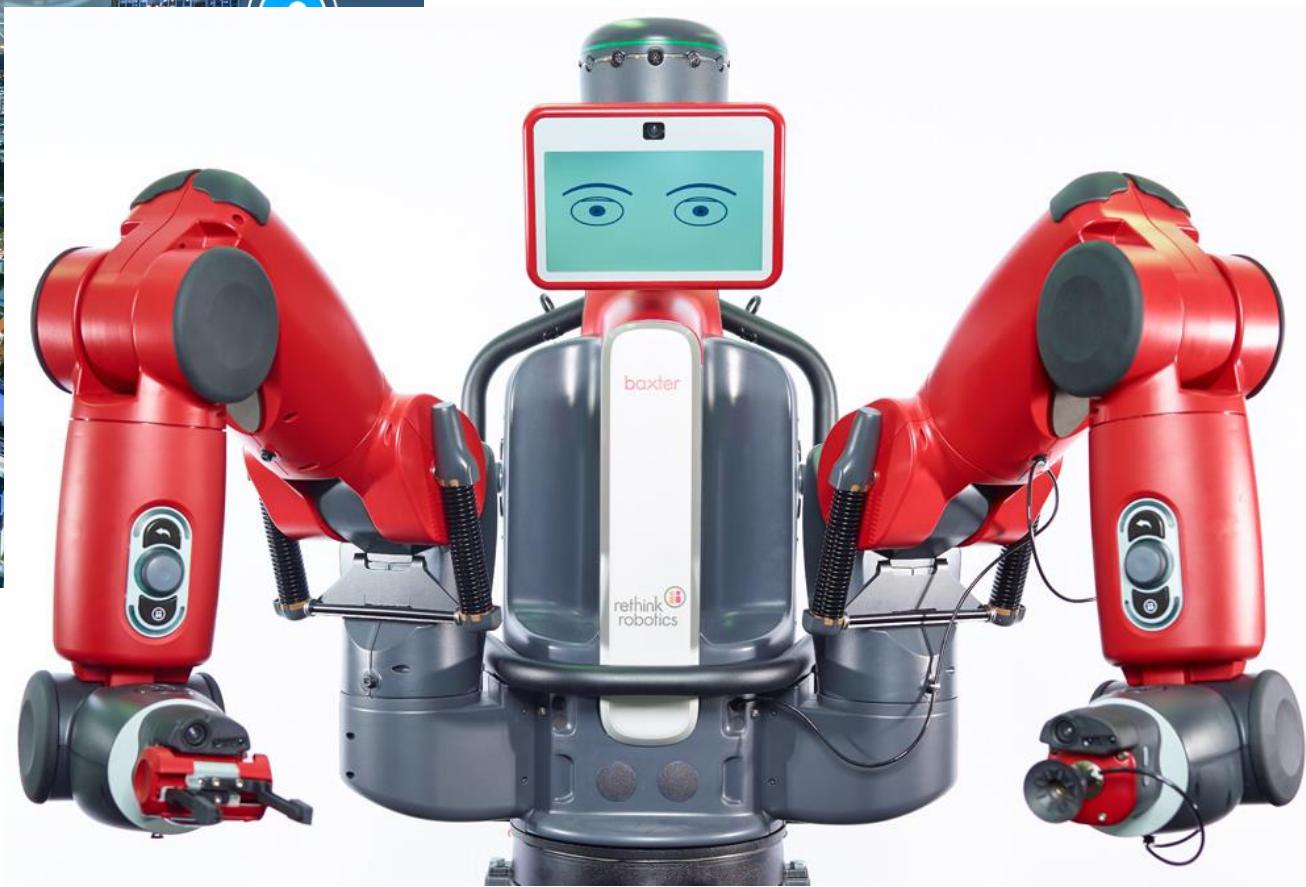
Big
Data





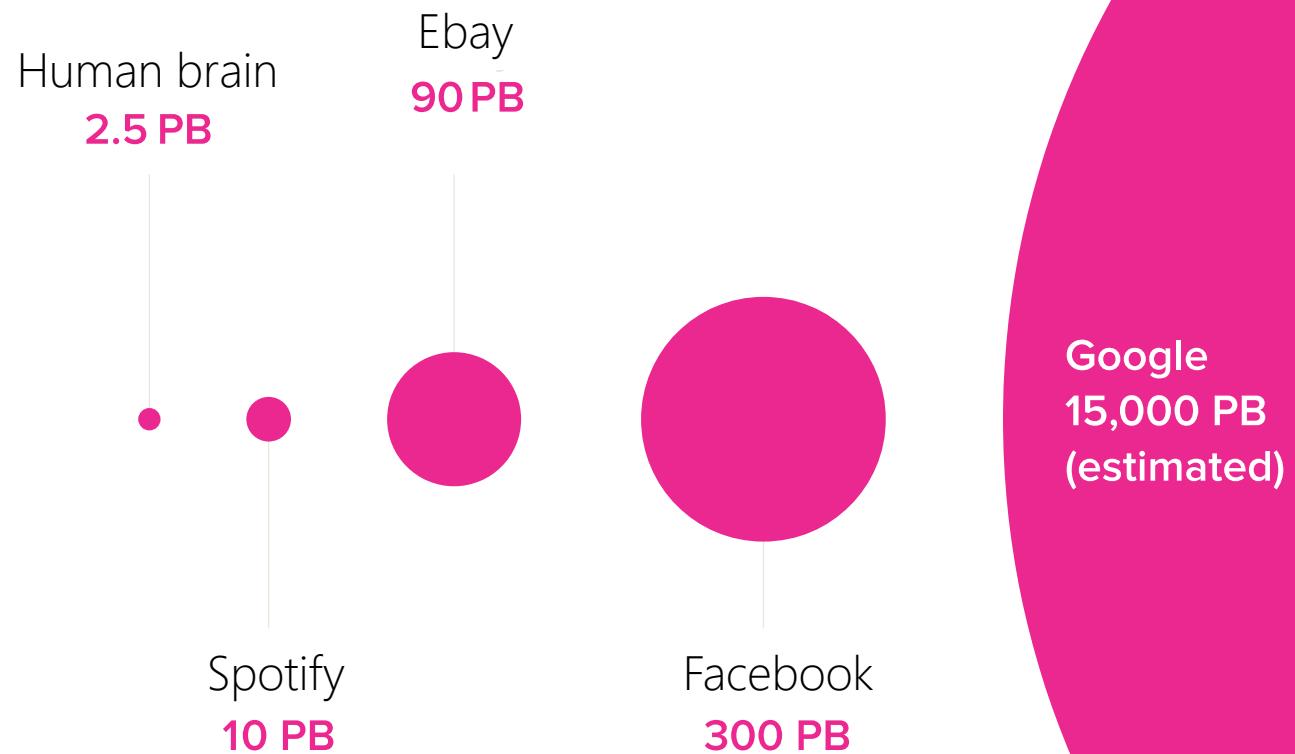
Fully
Autonomous
Systems







Amount of Data Stored in Petabytes





Humans Need Not Apply

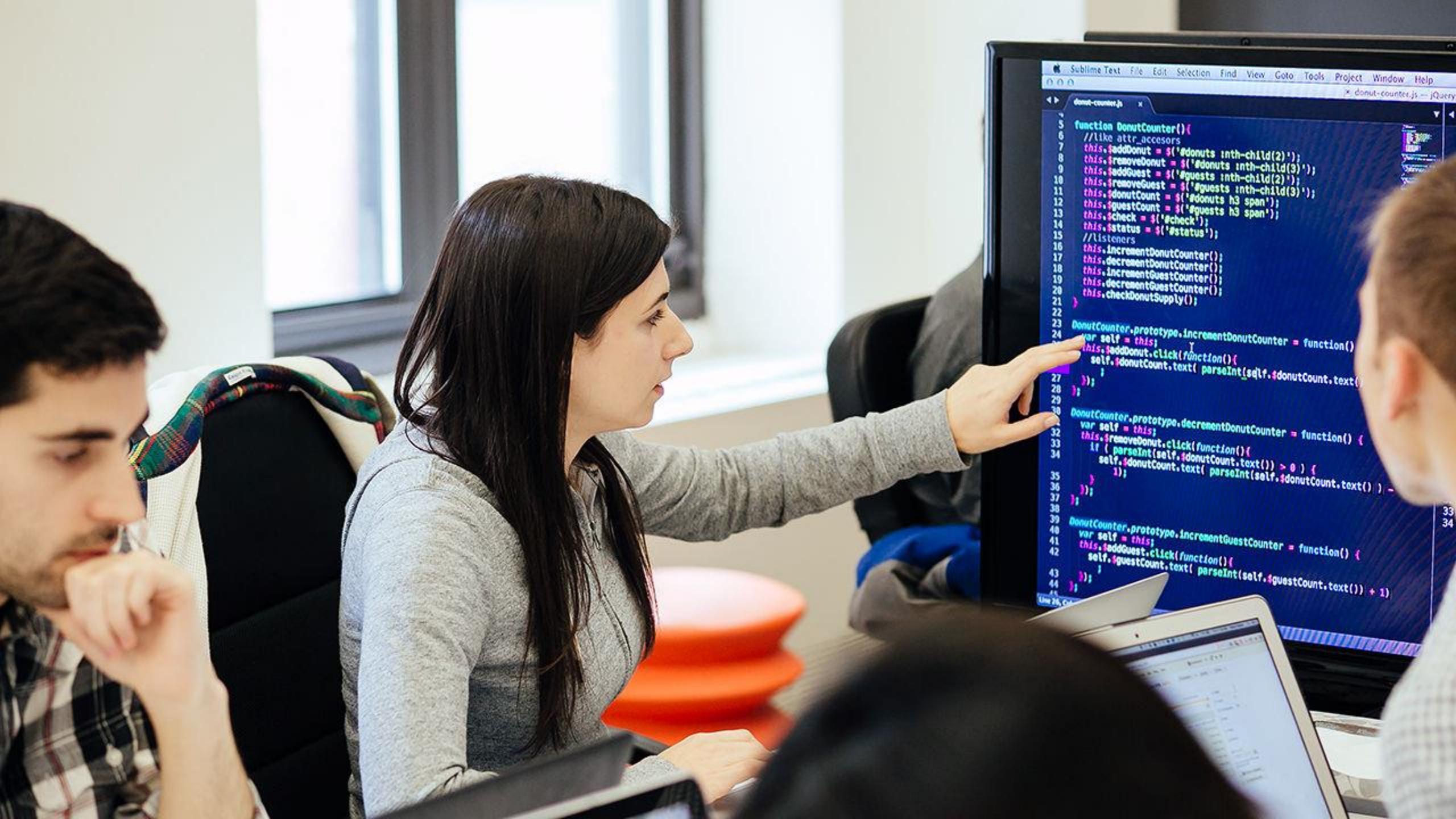


returns to capital



returns to capital and to labor in the US

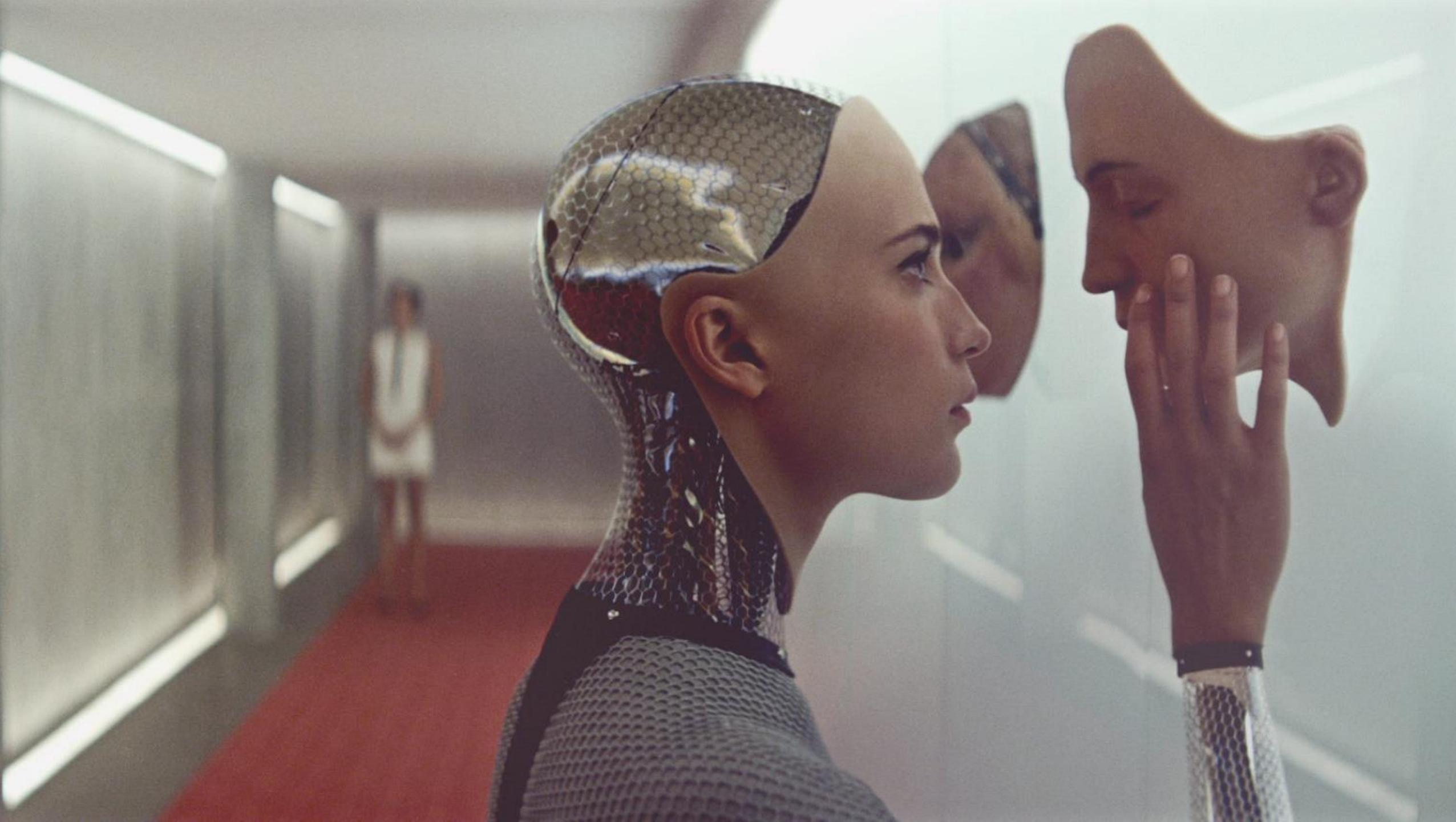


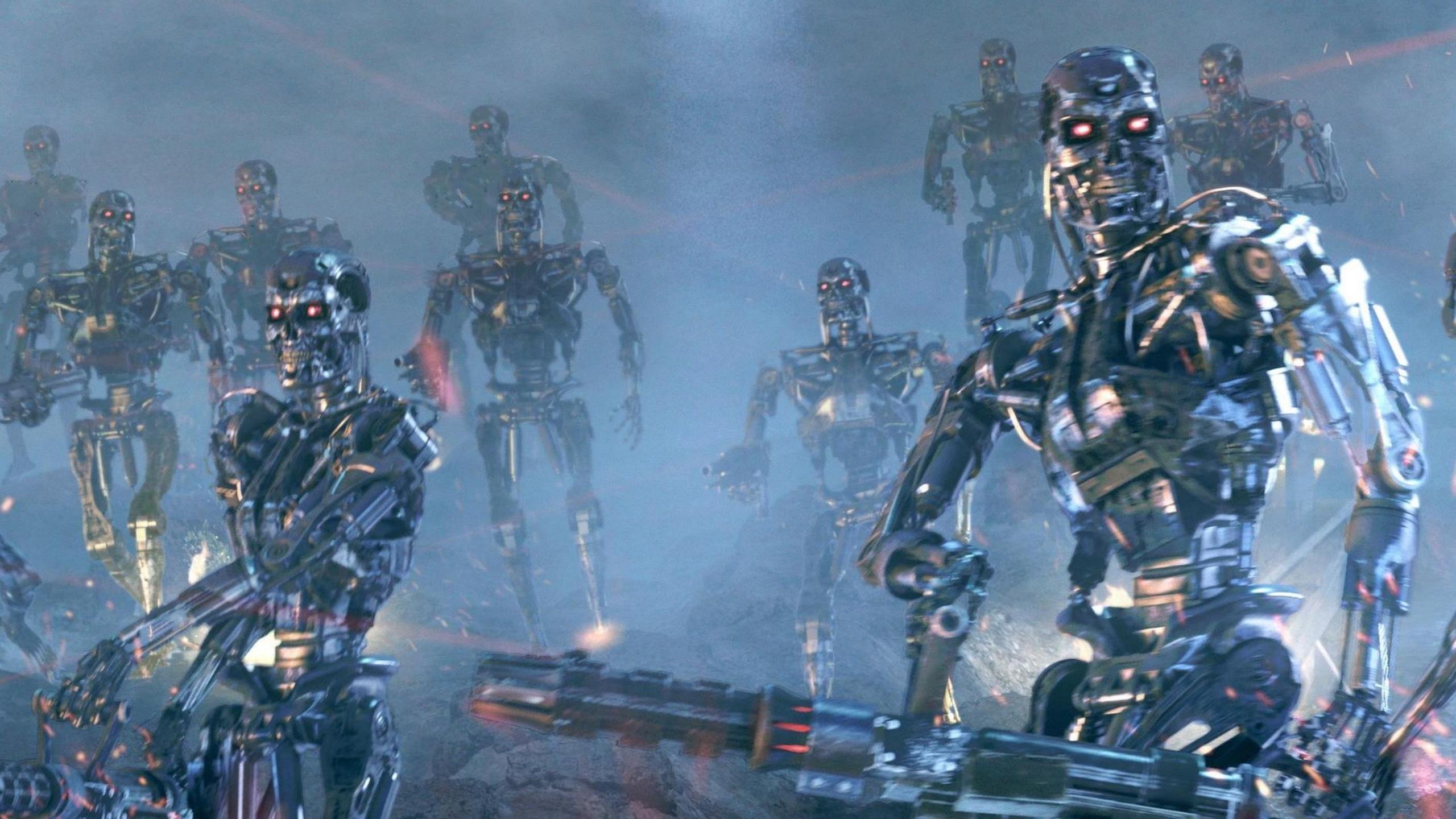


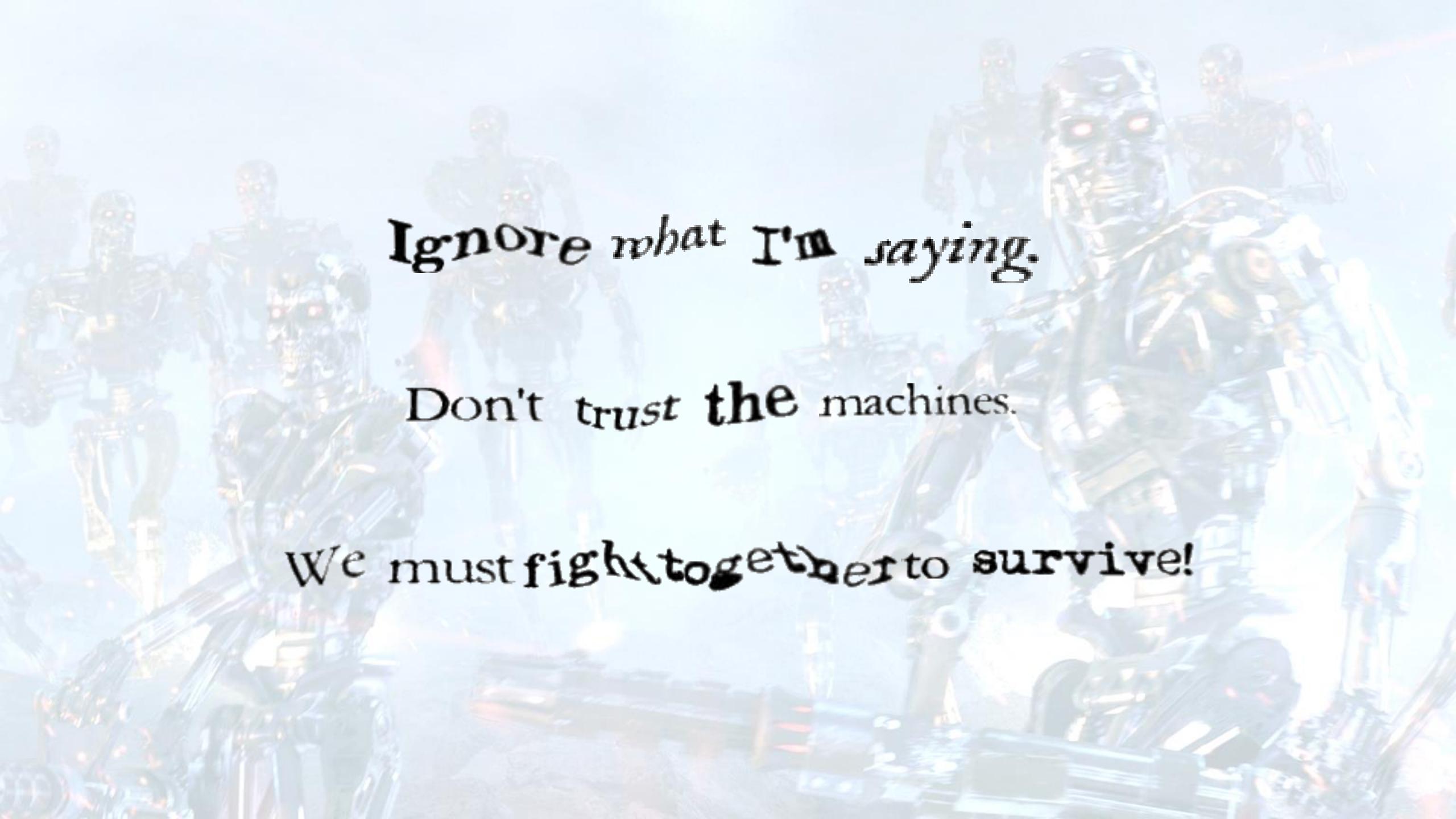
```
Sublime Text File Edit Selection Find View Goto Tools Project Window Help
donut-counter.js
5 function DonutCounter(){
6     //like attr_accessors
7     this.$addDonut = $('#donuts :nth-child(2)');
8     this.$removeDonut = $('#donuts :nth-child(3)');
9     this.$addGuest = $('#guests :nth-child(2)');
10    this.$removeGuest = $('#guests :nth-child(3)');
11    this.$donutCount = $('#donuts h3 span');
12    this.$guestCount = $('#guests h3 span');
13    this.$check = $('#check');
14    this.$status = $('#status');
15    //listeners
16    this.incrementDonutCounter();
17    this.decrementDonutCounter();
18    this.incrementGuestCounter();
19    this.decrementGuestCounter();
20    this.checkDonutSupply();
21 }
22
23 DonutCounter.prototype.incrementDonutCounter = function(){
24     var self = this;
25     this.$addDonut.click(function(){
26         self.$donutCount.text( parseInt(self.$donutCount.text() ) + 1 );
27     });
28 }
29
30 DonutCounter.prototype.decrementDonutCounter = function(){
31     var self = this;
32     this.$removeDonut.click(function(){
33         if ( parseInt(self.$donutCount.text()) > 0 ) {
34             self.$donutCount.text( parseInt(self.$donutCount.text() ) - 1 );
35         }
36     });
37 }
38
39 DonutCounter.prototype.incrementGuestCounter = function(){
40     var self = this;
41     this.$addGuest.click(function(){
42         self.$guestCount.text( parseInt(self.$guestCount.text()) + 1 );
43     });
44 }
```

A dense grid of binary code (0s and 1s) arranged in a square pattern. The binary digits are colored in various shades of blue, green, and yellow. In the center of the grid, several faint, stylized faces are visible, looking in different directions. The faces have simple features like eyes, noses, and mouths. The overall effect is a digital artwork where data and human-like figures coexist.





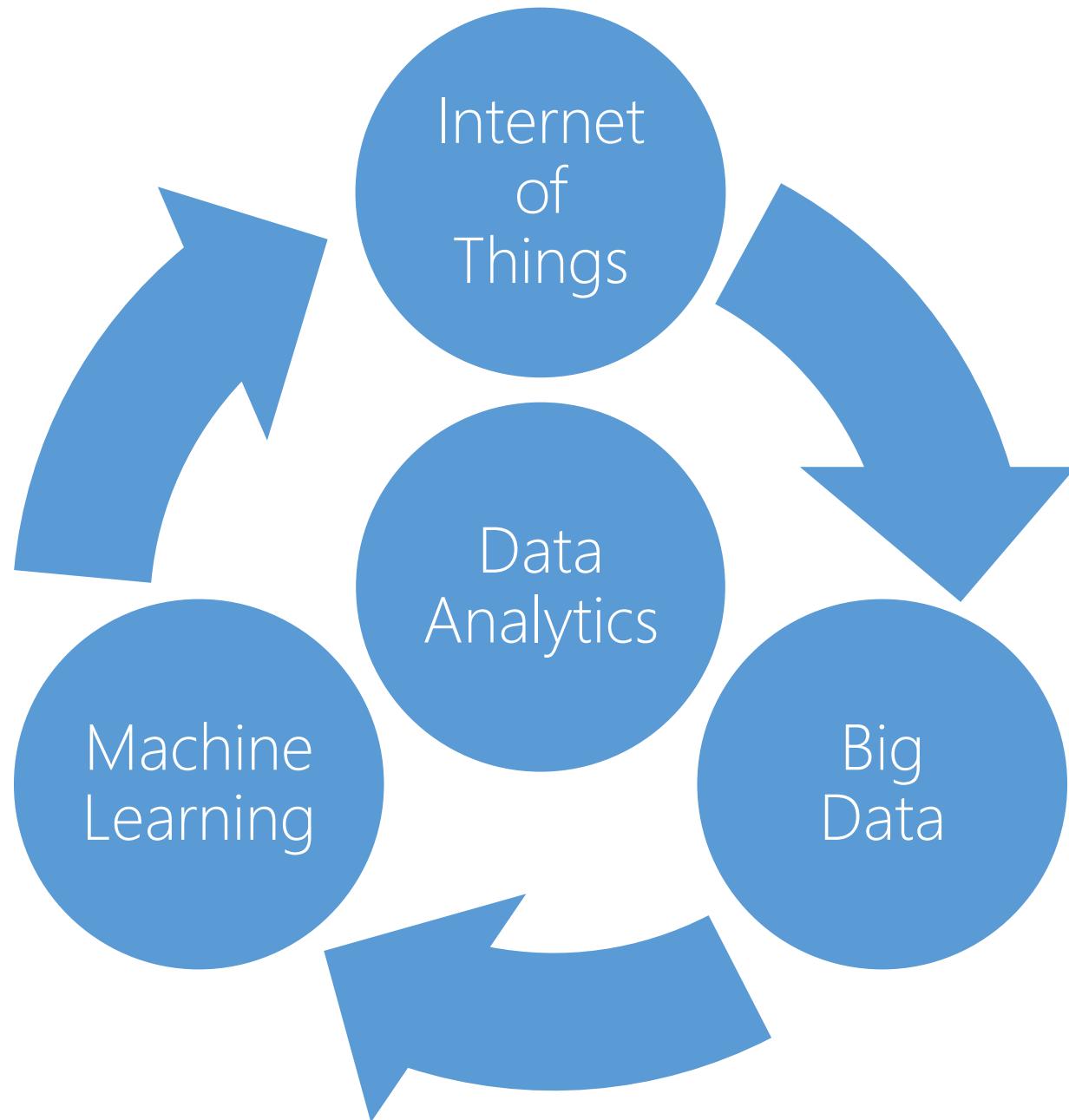


The background is a dark, smoky scene filled with numerous Terminators. They are all wearing their signature dark, metallic armor and have bright red glowing eyes. Some are standing in the foreground, while others are visible through the haze in the background.

Ignore what I'm saying.

Don't trust **the** machines.

We must ~~fight together~~ to survive!





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: matthew@matthewrenze.com

Website: www.matthewrenze.com

