

# Machine Learning for the IoT

Welcome to the IoT

---

# Objectives

---

- An overview of the IoT ecosystem and its taxonomy
- Understand the complexity IoT systems and the design challenges
- Give additional details on the main contents of the course

# Contents

---

- Definitions
- Practical use-cases and proof-of-concepts
- Key features
- Enabling technologies
- Resource constraints
- Architecture
- Characteristics of IoT data

# IoT Concept

*"The Internet of Things connects everyday consumer objects and industrial equipment onto the network, enabling information gathering and management of these devices via software in order to increase efficiency, enable new services, or achieve other health, safety, or environmental benefits"*

Kevin Ashton, a British technologist in 1999 was Executive Director at MIT's Auto-ID Center, an RFID research consortium

[Wired talk on youtube](#)

1999!

“THE INTERNET OF THINGS IS  
ABOUT EMPOWERING COMPUTERS  
...SO THEY CAN SEE, HEAR  
AND SMELL THE WORLD FOR  
THEMSELVES”

KEVIN ASHTON  
INVENTOR OF THE TERM  
“INTERNET OF THINGS”



Interview at: [https://www.youtube.com/watch?v=PXncS2\\_63o4](https://www.youtube.com/watch?v=PXncS2_63o4)

# Definition (a practical one)

- The Internet of things (IoT) is a system of interrelated computing devices, mechanical and digital machines provided with unique identifiers (UIDs) and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction [source: wiki]
- Terminologies similar to IoT
  - USN (Ubiquitous Sensor Networks)
  - M2M (Machine-to-Machine)
  - IoE (Internet of Everything) – Cisco's favorite term
  - Cloud of Things
  - Web of Things



# IoT from different players

---

- Dr. John Barrett at TEDxCIT -- The Internet of Things
  - <http://www.youtube.com/watch?v=QaTIt1C5R-M>
- Intel IoT -- What Does The Internet of Things Mean?
  - <http://www.youtube.com/watch?v=Q3ur8wzzhBU>
- Cisco IoT -- Networking Overview
  - <https://www.youtube.com/watch?v=XYD-rHkCfyM>

# Sensing & Communicating: the core of IoT

- Sensing & Communicating (S&C) capability = smartness
  - Quite old concept in the industrial segment (\$\$\$), e.g. oil&gas
  - First working device: vending machine at University of Michigan (1982)
  - The concept at large scale: mobilephone (Apple?)

I am a physical coke machine with some special hardware attached to my empty lights. Since my empty lights blink on a dispense, it is possible to count the number of sodas still inside me. The custom hardware feeds the data from these sensors into a small dedicated machine that runs custom software to send this information over the local net to a local Unix machine. That machine is also running some custom software which translates the status into a plan file, so that you can finger the machine from all over the world and get the status.

The CMU CS Department Coke Machine



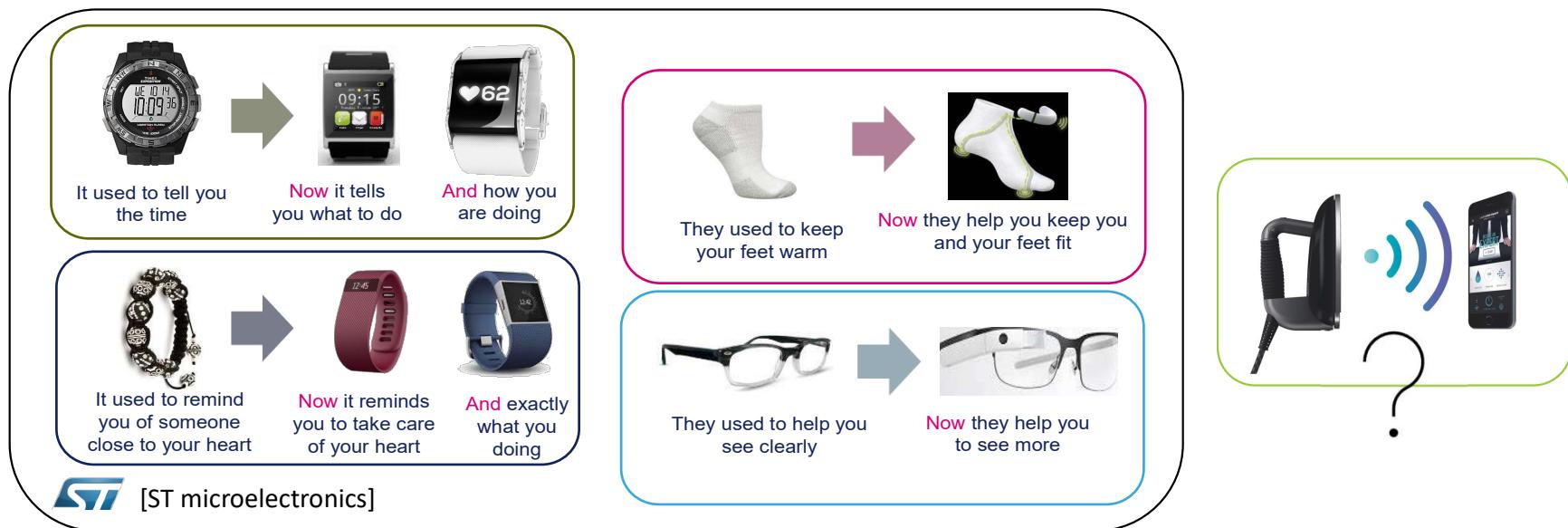
Position  
Acceleration  
Pressure  
Temperature  
Humidity  
Light  
Vision  
Touch



3/4/5G  
WiFi  
Bluetooth  
NFC  
InfraRed

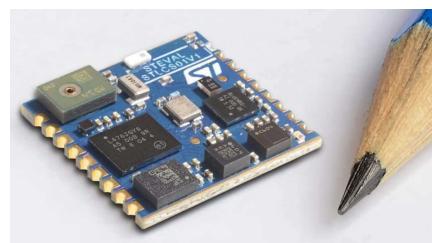
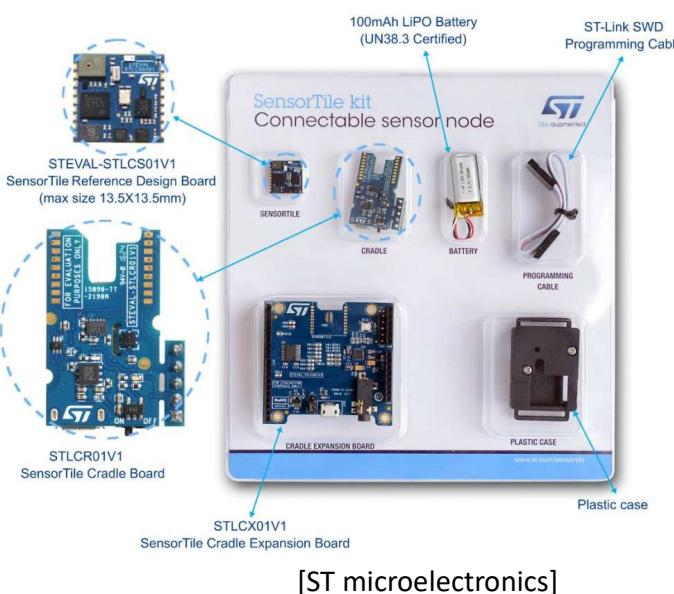
# Beyond smartphones: the rise of SmartThings

- Everything (or almost everything) had become (or is becoming) smart
  - Massive integrations of embedded sensors and wifi connectivity
  - .... even where it is not needed, at all!



# Enabling technology: sensor miniaturization

- Miniaturization and mass production of low-cost, composable and embedded wifi Multi-Sensorial platforms

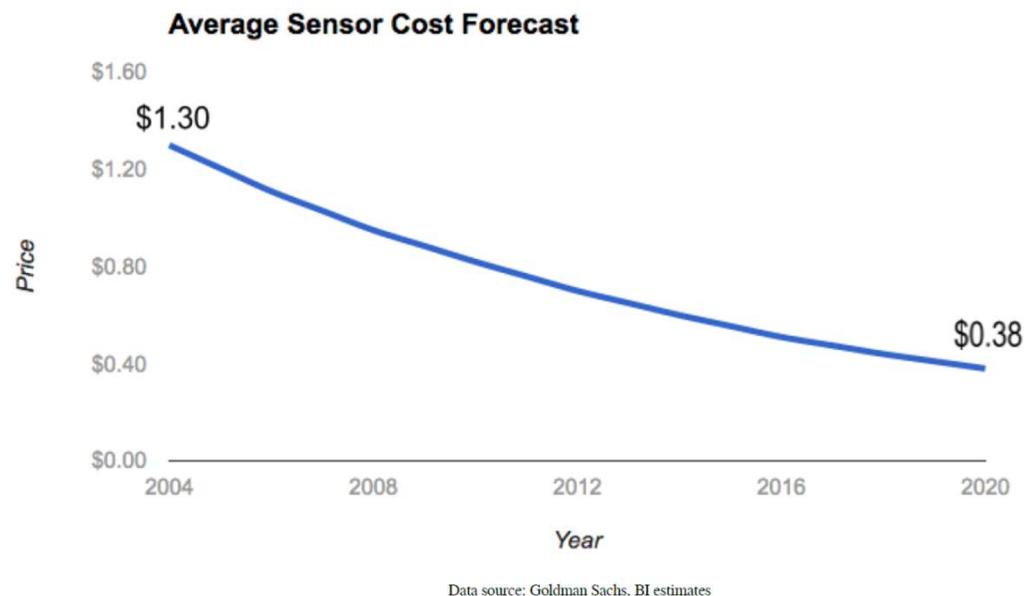


STMicroelectronics' **13.5mm x 13.5mm SensorTile** is one of the smallest turnkey sensor board of its type, containing a **MEMS accelerometer, gyroscope, magnetometer, pressure sensor, and a MEMS microphone**. With the on-board low-power STM32L4 microcontroller, complete **Bluetooth® Low-Energy (BLE)** transceiver, a broad set of system interfaces that support use as a sensor-fusion hub or as a platform for firmware development, and can be **plugged into any STM32 Nucleo board and Arduino**.

Cost in the range of **70 EUR** (kit).

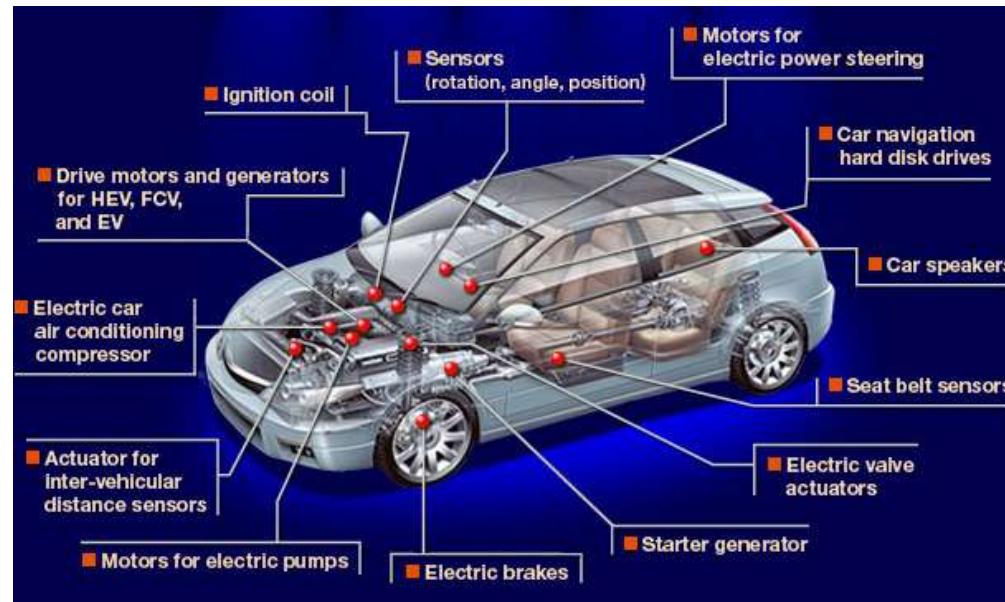
# Technology scaling

- Goldman Sachs predicts an average sensor cost in 2020 of under \$0.40 USD, 30% of what it was in 2004



# SmartThings: Not just sensors

- A smart object, or smart thing, or Cyber-Physical System (CPS), is a rather complex electro-mechanical system made of many parts that operate in synergy
  - Sensor(s)
    - $\times 10^3$
  - Actuator(s)
    - $\times 10^2$
  - Controller(s)
    - $\times 10^2$
  - Connectivity

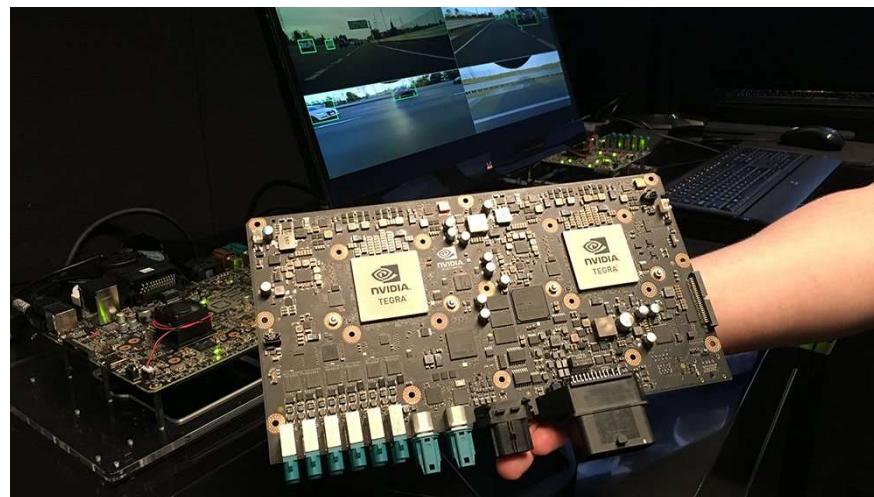


# SmartThings: Not just sensors

- A smart object, or smart thing, or Cyber-Physical System (CPS), is a rather complex electro-mechanical system made of many parts that operate in synergy
  - Sensor(s)
    - $\times 10^3$
  - Actuator(s)
    - $\times 10^2$
  - Controller(s)
    - $\times 10^2$
  - Connectivity



[Tesla]



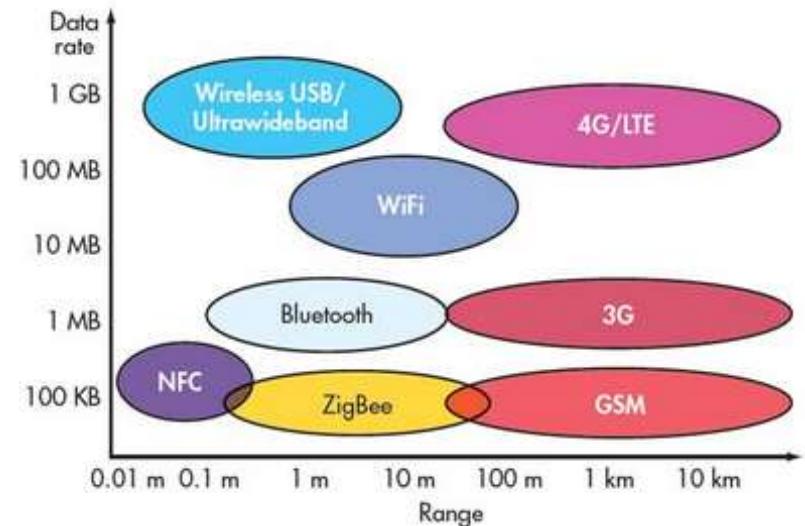
Part-I: Welcome to the IoT

Full Self Driving Chip in the Tesla Model S

When powered on and engaged, **sensory input** is fed to the board from a variety of sources. Those include current car readings such as **inertial measurement unit (IMU)**, **radar**, **GPS**, **ultrasonic sensors**, **wheel ticks**, **steering angle**, and **maps data**. There are **8 vision cameras** and **12 ultrasonic sensors**. Data is fed to both FSD chips **simultaneously** for processing. The two chips independently form a future plan for the car - a detailed plan of what the car should do next. [wikichip]

# SmartThings: Not just sensors

- A smart object, or smart thing, or Cyber-Physical System (CPS), is a rather complex electro-mechanical system made of many parts that operate in synergy
  - Sensor(s)
    - $\times 10^3$
  - Actuator(s)
    - $\times 10^2$
  - Controller(s)
    - $\times 10^2$
  - Connectivity
    - Multiple channels/cards
      - From 1 to 4/5



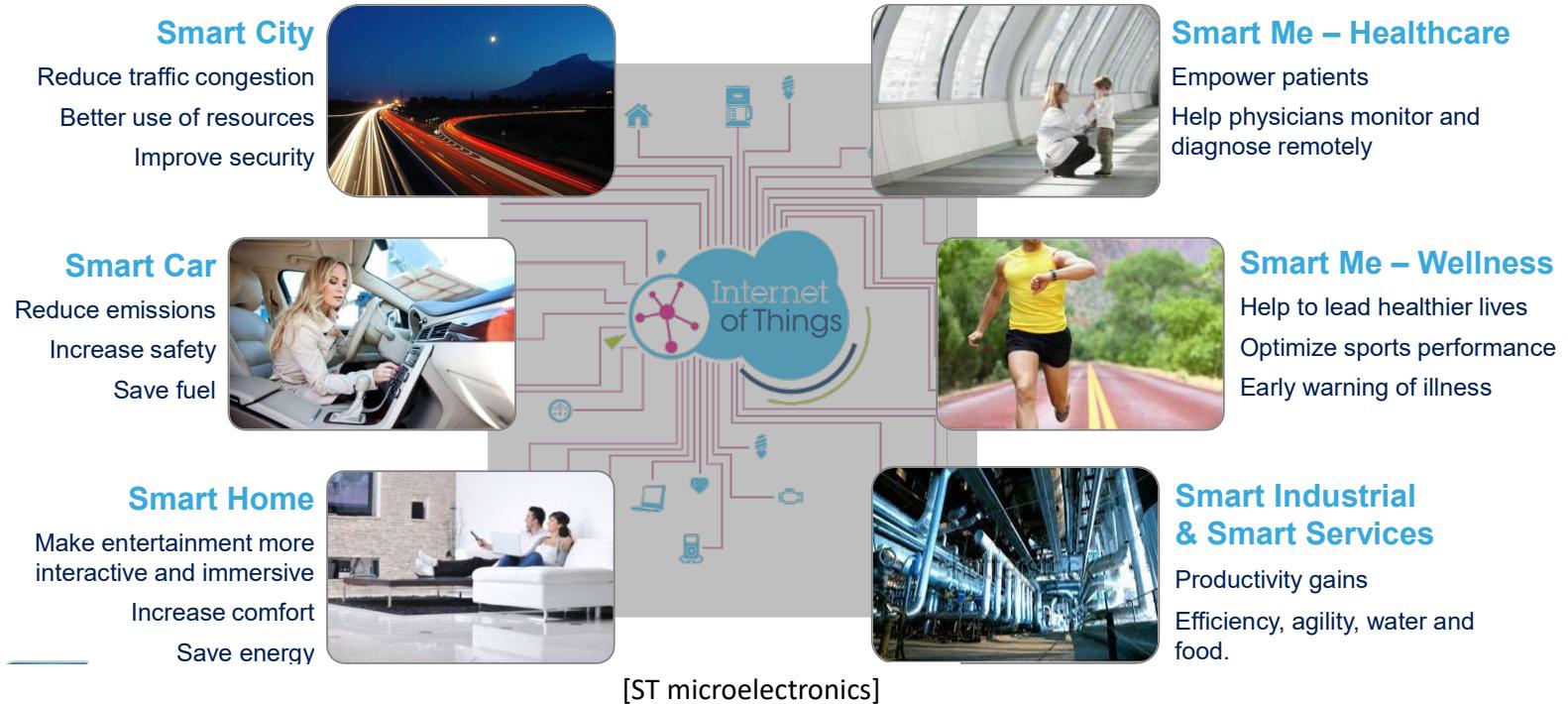
# SmartThings: Not just sensors (II)

- Even the simplest (and more common) smart-objects are the integration of the four parts, yet with limited capability
  - Sensor
  - Actuator
  - Controller
  - Connectivity



# From Smart Things to Smart Systems

- Smart devices alone have no value, together (system) can enable new services



# Smart Industry

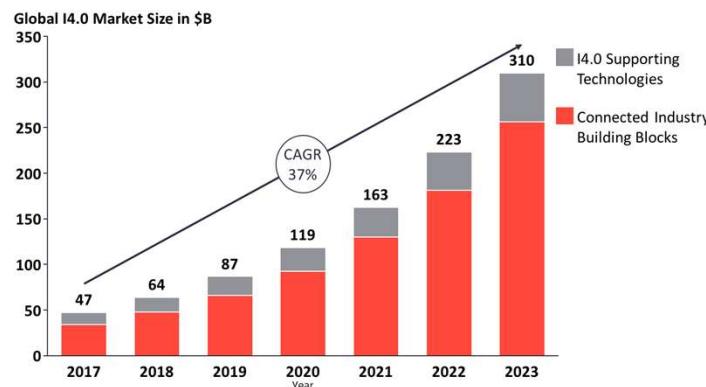


<https://www.youtube.com/watch?v=iyj-NKA91yg>

# Industrial 4.0, main driving force in EU

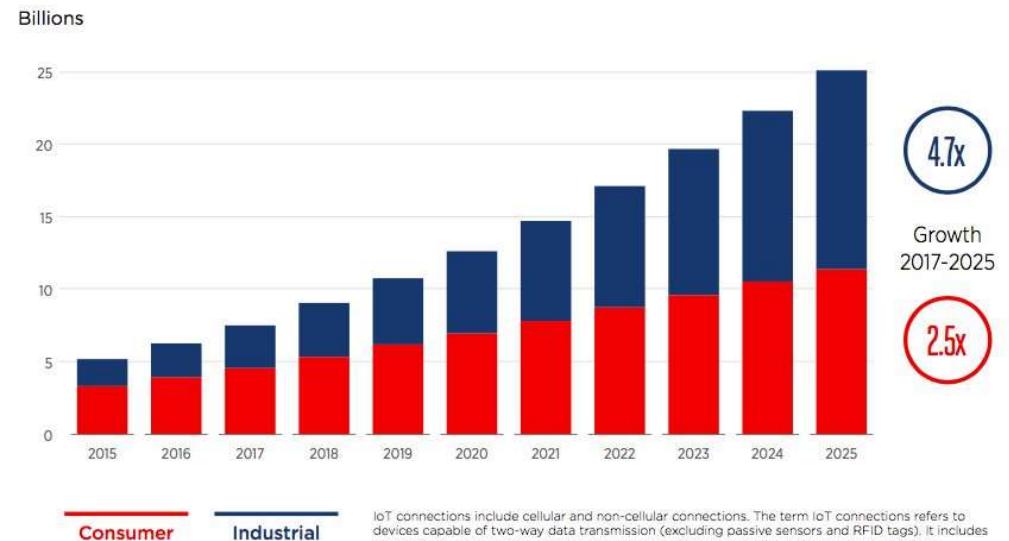
IoT ANALYTICS

## Global Industry 4.0 Market Size 2017-2023



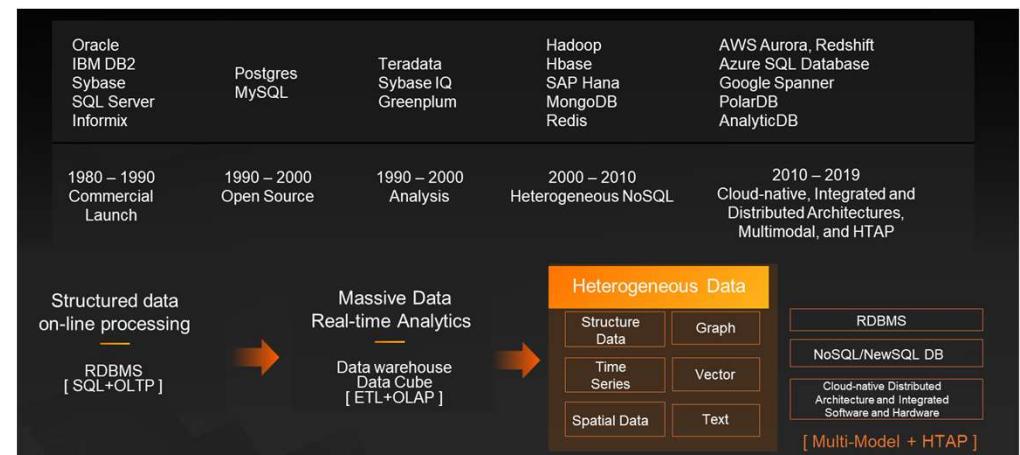
Note: The overall market for I4.0 refers to global spending on the six connected industry building blocks and six I4.0 supporting technologies  
Source: IoT Analytics – November 2018

## IoT connections worldwide



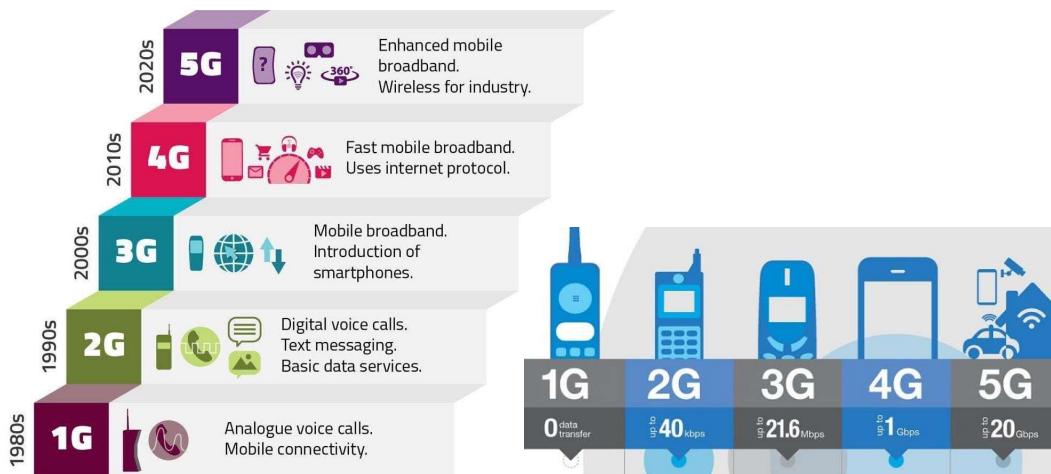
# Enabling technology: DBM

- Data-base technologies
  - Relational (yesterday)
    - Table based
    - Vertically scalable
  - Non-relational or Distributed Database (today)
    - Document-based, key-value pairs, graph databases
    - Horizontally scalable



# Enabling technology: ubiquitous wifi coverage

- Mobile broadband
  - GSM/3G/4G/5G, increasing speed, stability, but most of all capacity (number of connected devices)



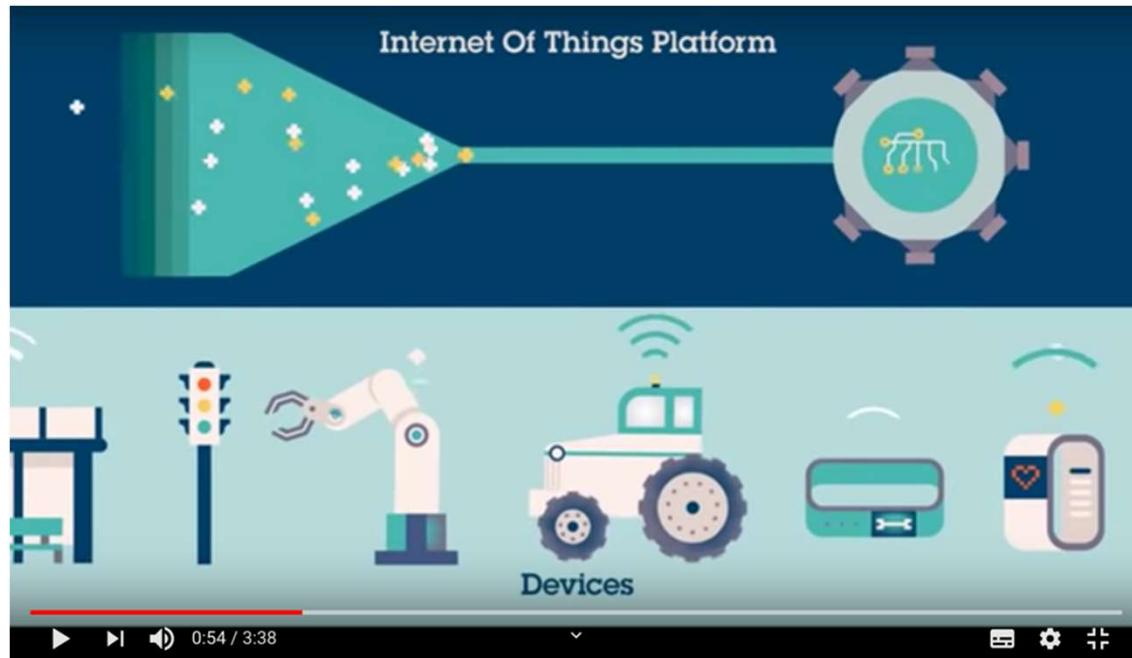
Speed	Capacity	Latency/ Response time	Connection	Mobility	Battery life
10 to 20 Gbps	10 TB/s/km <sup>2</sup>	Ultra-reliable low-latency communication (URLLC): 1 millisecond	1,000,000 devices/km <sup>2</sup>	500+ km/hour	15+ years
10-100x of 4G	1000x of 4G	Enhanced Mobile Broadband (eMBB): 4 millisecond	1/10 of 4G	100x of 4G	1.5x of 4G
					10x of 4G

Source: GSMA, EY

Graphic© Asia Briefing Ltd.

# From Smart Systems to Smart Ecosystems

IoT = Smart Systems Connected

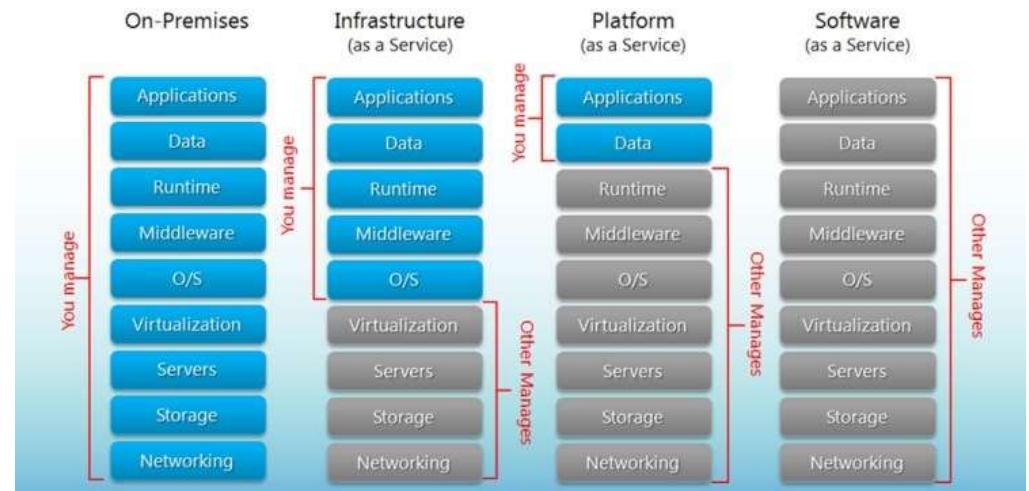


Video at: <https://www.youtube.com/watch?v=QSIPNhOjMoE>

- Forecast trends
  - Anticipate social needs before they materialize
- Minimize resource waste
  - Produce only what is really needed
- Improve quality of life
  - Services get custom
- Reach more people
  - High-technology to lower-cost
- Maximize efficiency
  - Distributed services

# Enabling technology: Cloud computing

- Cloud infrastructure
  - High computing power thanks to massively parallel computing: GP-GPUs
  - Massive storage technologies
  - Distributed SW stacks & virtualization
  - Advantages
    - Reduced time-to-market/money
      - Fast prototyping
    - Cheap
      - accessible to everyone
    - Scalable
- Analytics tools based on ML
  - Collect, store, and process raw-data
  - Generate accessible, highly-informative data

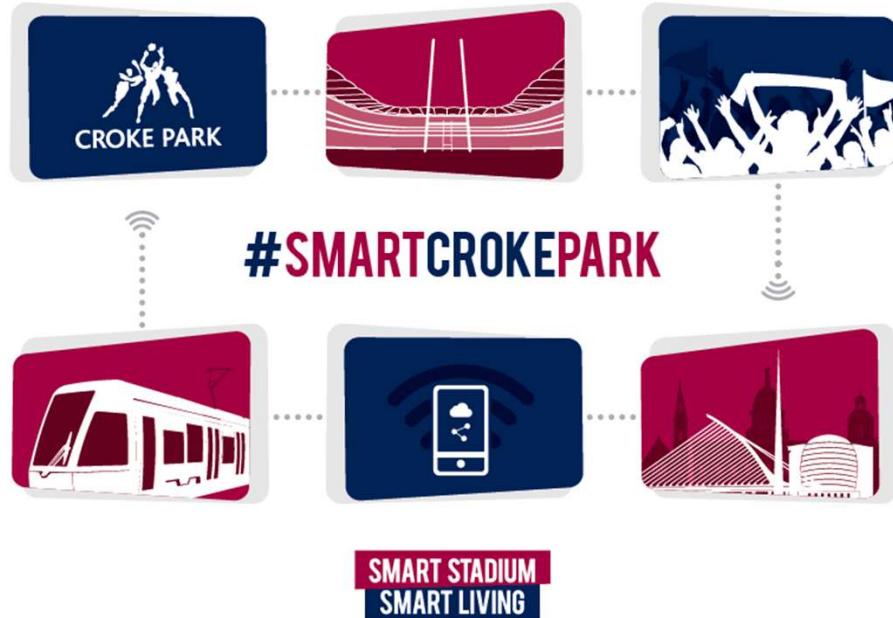


# Top 10 IoT Cloud Platforms

---



# Smart City



[https://www.youtube.com/watch?v=4irsXI\\_phCs](https://www.youtube.com/watch?v=4irsXI_phCs)



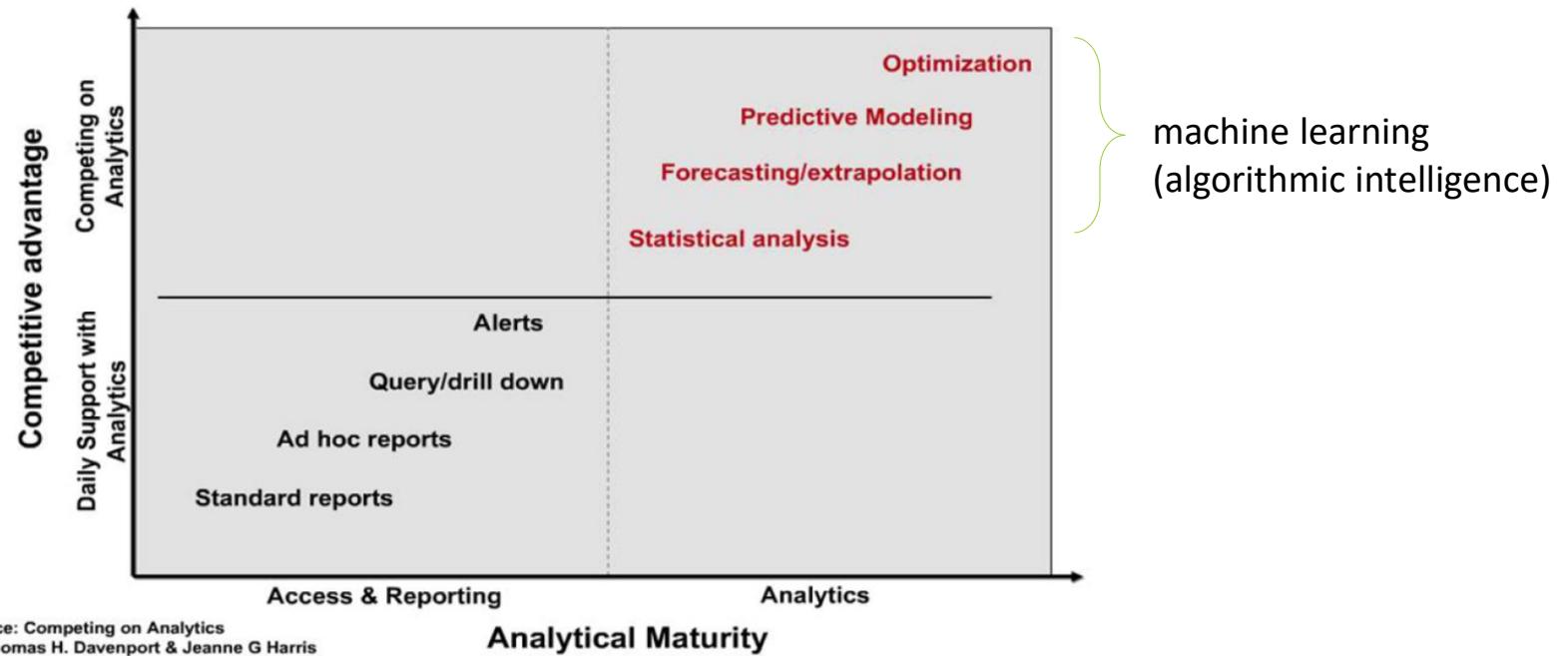
<https://www.youtube.com/watch?v=49FEwpVLgPo>

# IoT Challenges



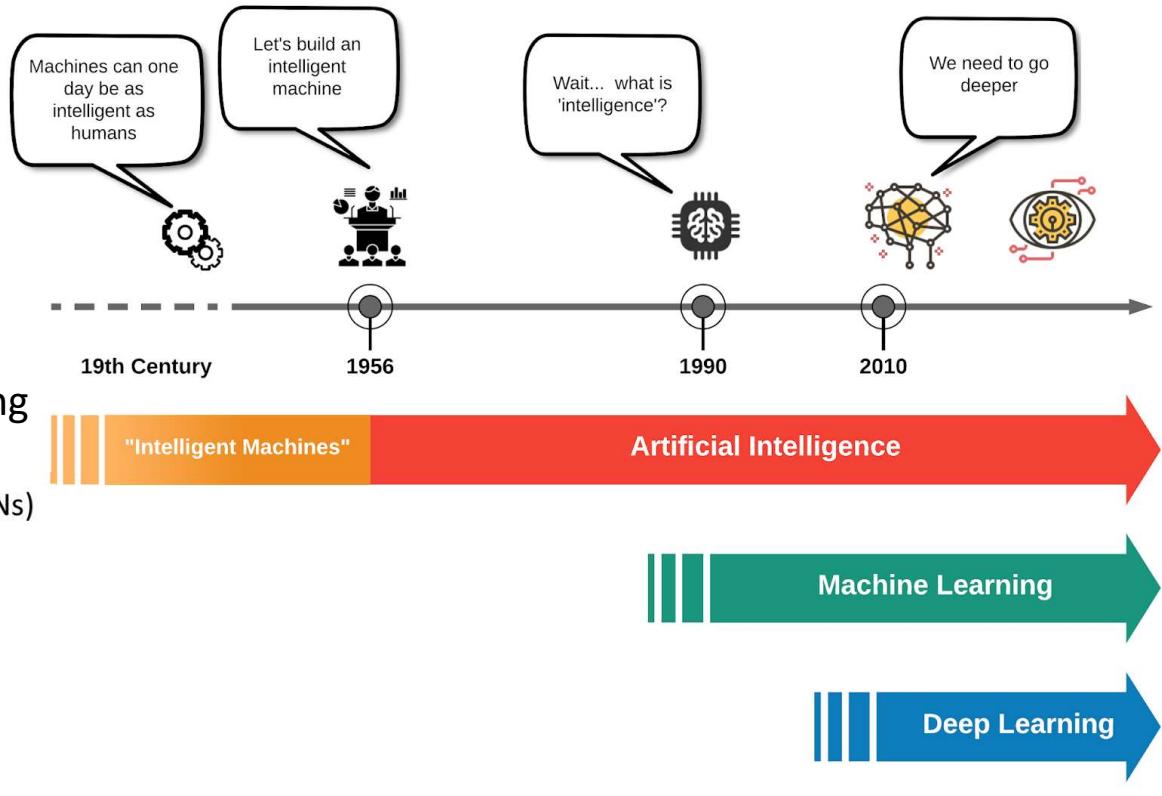
<https://www.youtube.com/watch?v=XYD-rHkCfyM>

# Enabling Technology: data analytics



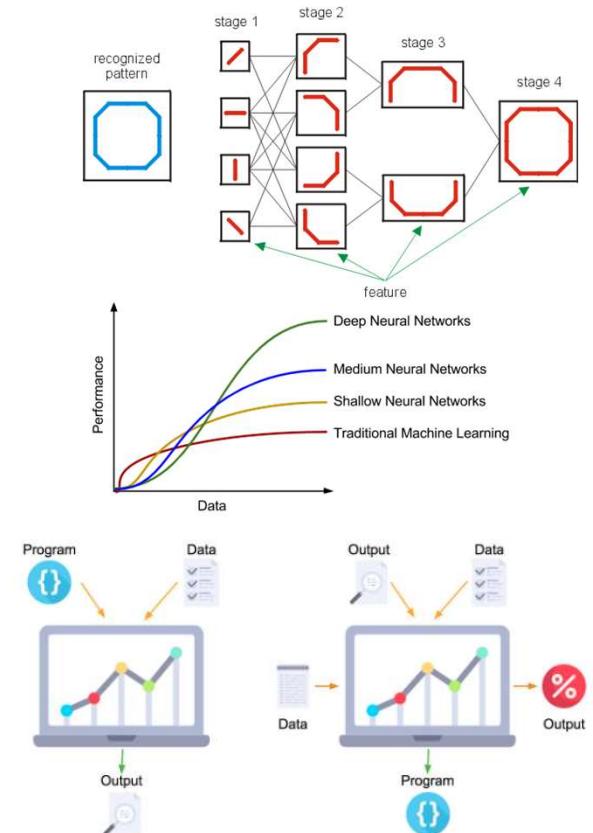
# Enabling technology (for ML)

- Computing power
  - GP-GPUs
- Availability of data
  - IoT technologies
    - Ubiquitous sensors
    - Data connectivity
    - Cloud and storage
- Deep Learning as the new AI spring

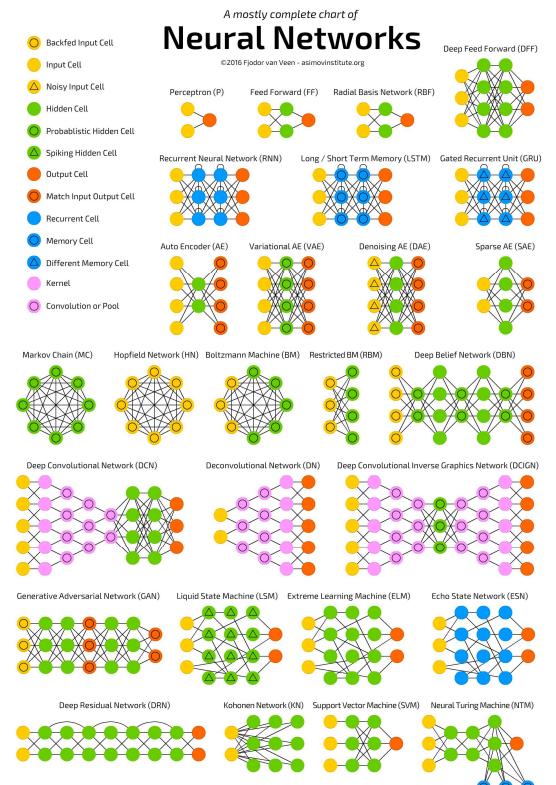


# Why Deep Learning and why now

- It works
  - Human-level accuracy in complex tasks
  - Hierarchical features extraction, like the brain (?)
- Enough resources for training
  - Data (sensors + connectivity)
  - GPUs (massively parallel)
- End-to-end learning (paradigm shift)
  - Rule-free programming (Re-coding = re-training)
- Why in this course?
  - For the main reasons above
  - It's a nice computing problem (personal view)



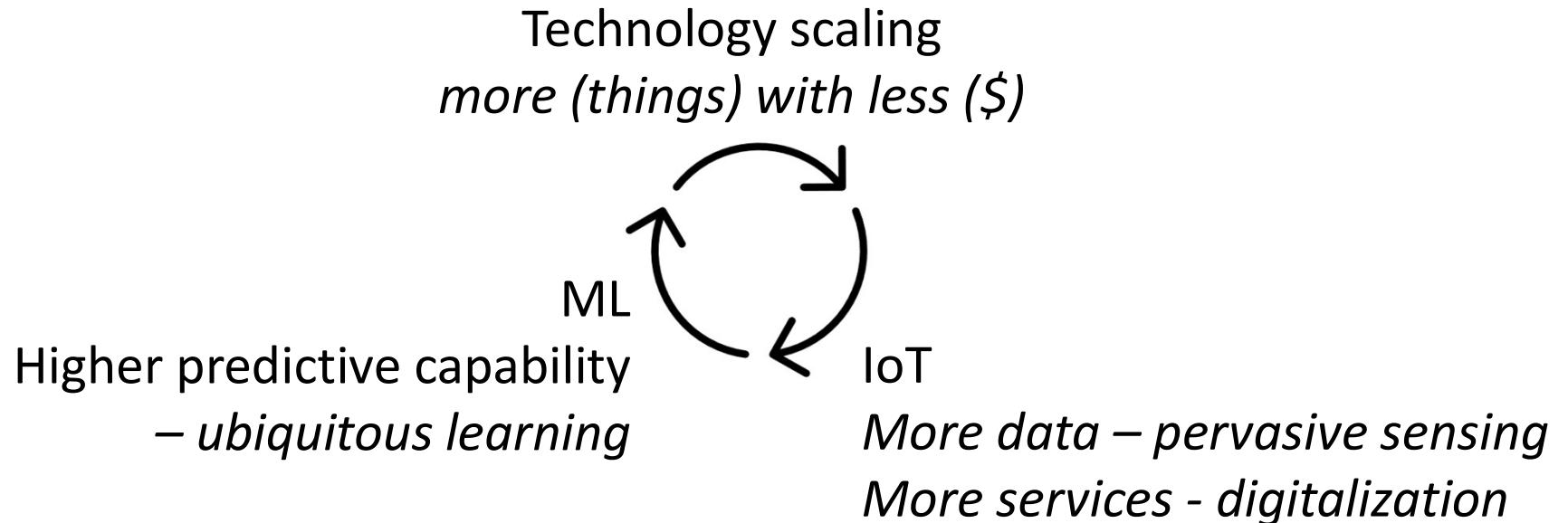
# Plenty of DNN Models Out There



<https://www.asimovinstitute.org/author/fjodorvanveen/>

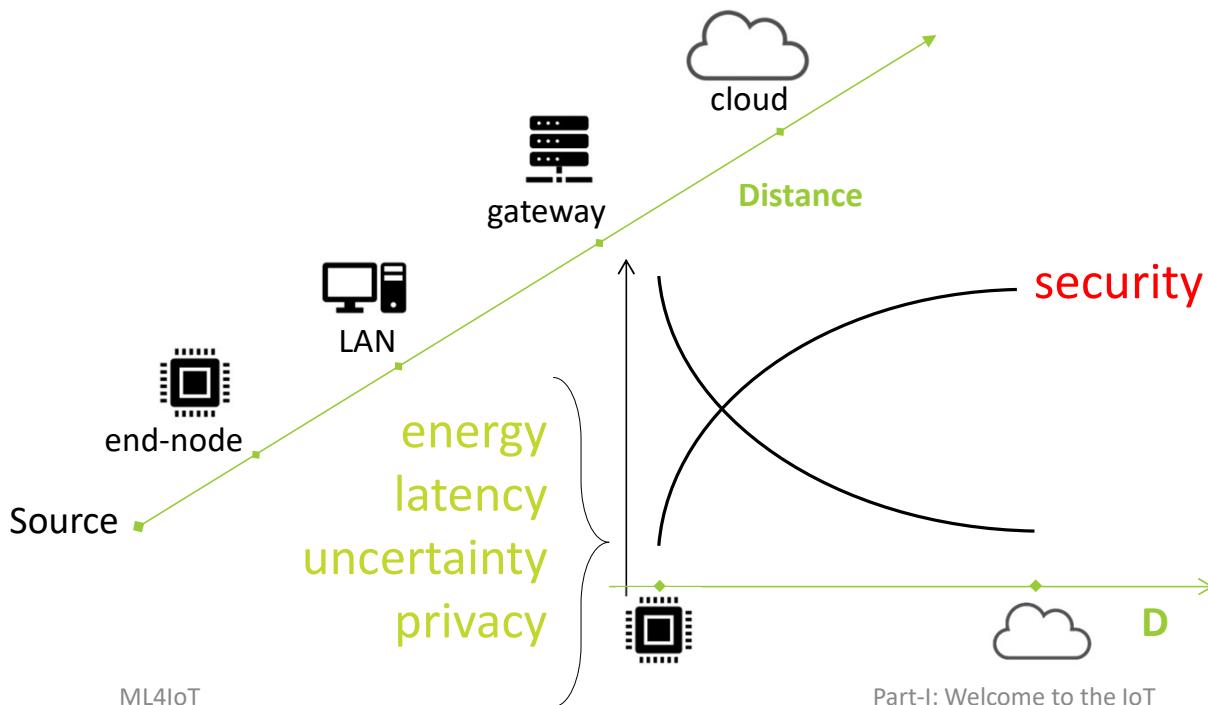


# ML+IoT: Selfsustainable cycle (AIoT for marketers)

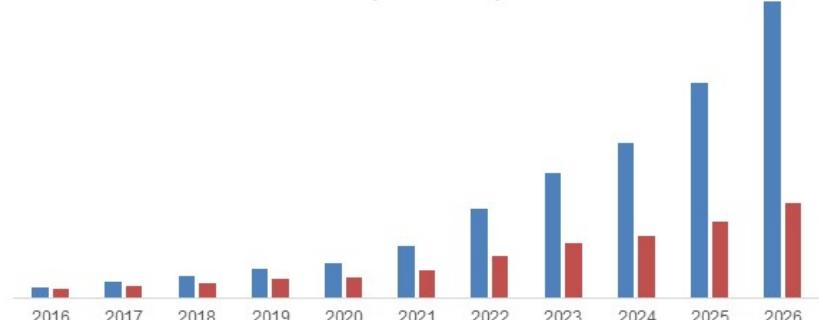


# What's next? Inference on the edge (a.k.a. edge-AI)

- Today: both training and inference in the cloud
- Very near feature: inference on the edge

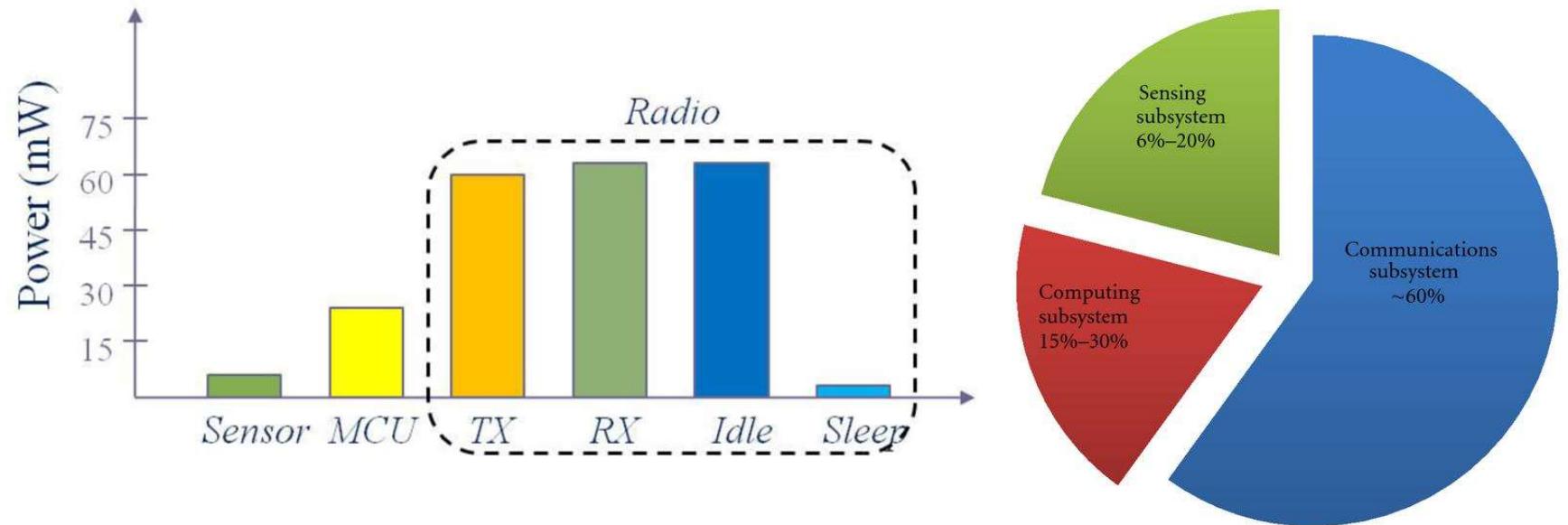


North America AI Chipsets Market, By Processing Type, 2016 - 2026 (USD Million)



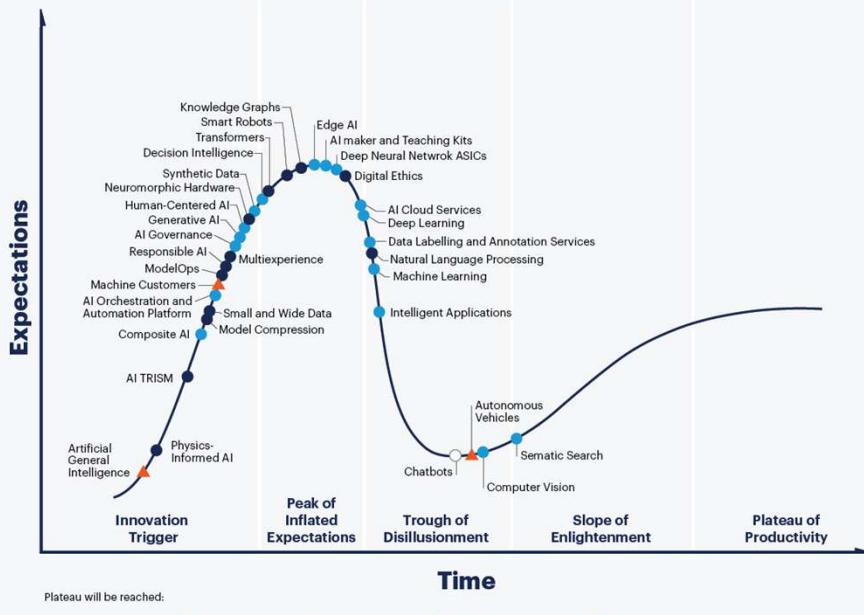
Source: Global Market Insights

# Power breakdown in a tiny sensor node



# Gartner Hype Cycle

## Hype Cycle for Artificial Intelligence, 2021



gartner.com

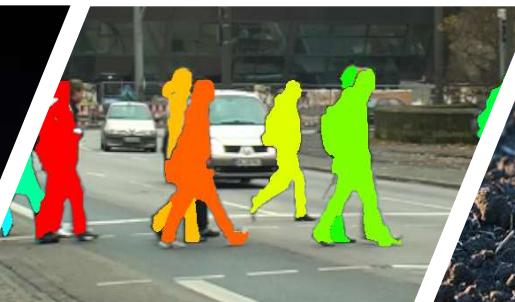
Source: Gartner  
© 2021 Gartner, Inc. and/or its affiliates. All rights reserved. Gartner and Hype Cycle are registered trademarks of Gartner, Inc. and its affiliates in the U.S. 1482644

Gartner®

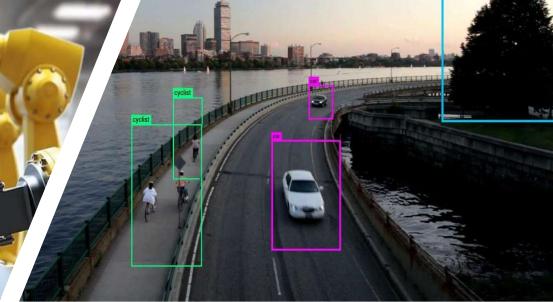


**Edge-AI:**  
Most of the players understood the value. To make it working in a real business is still concerning.

# Killer applications



Reduce the volume of data

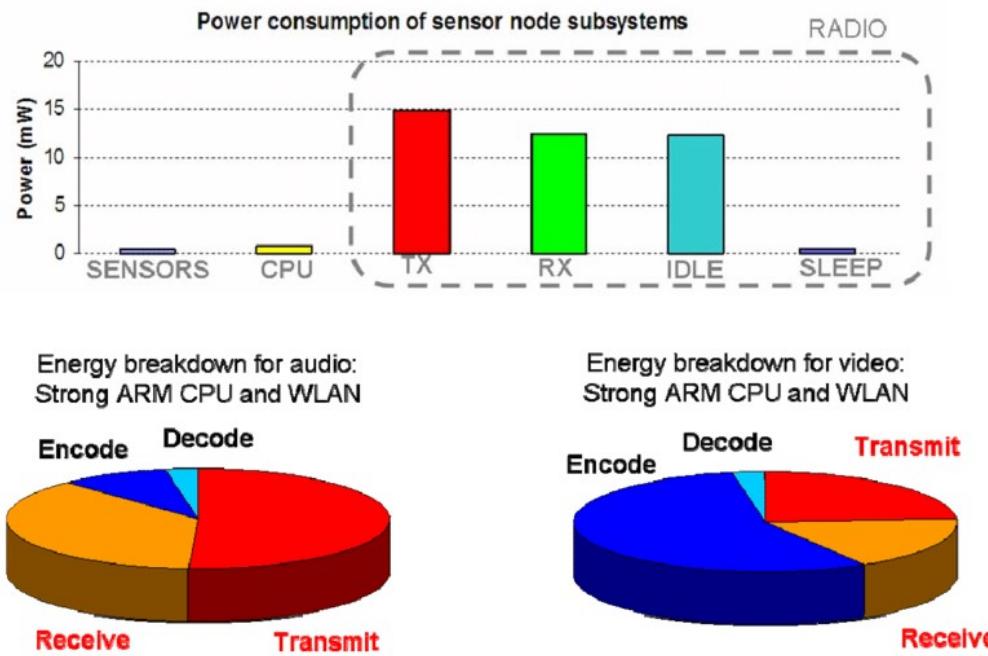


Reach remote geographical areas

Meet real-time constraints

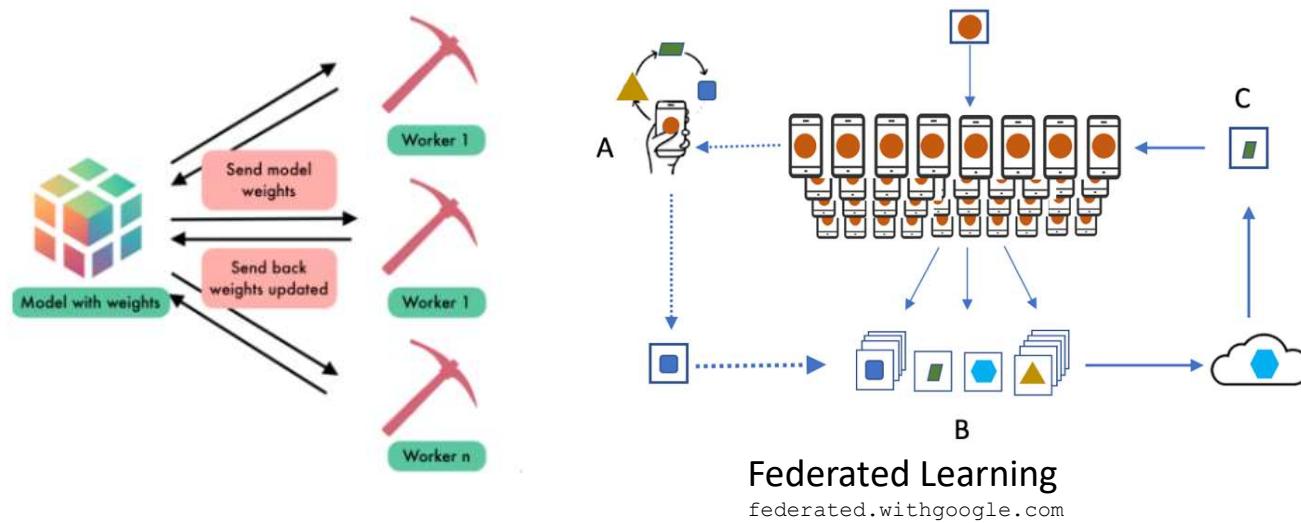
M2M information commutation

# Life is a tradeoff!

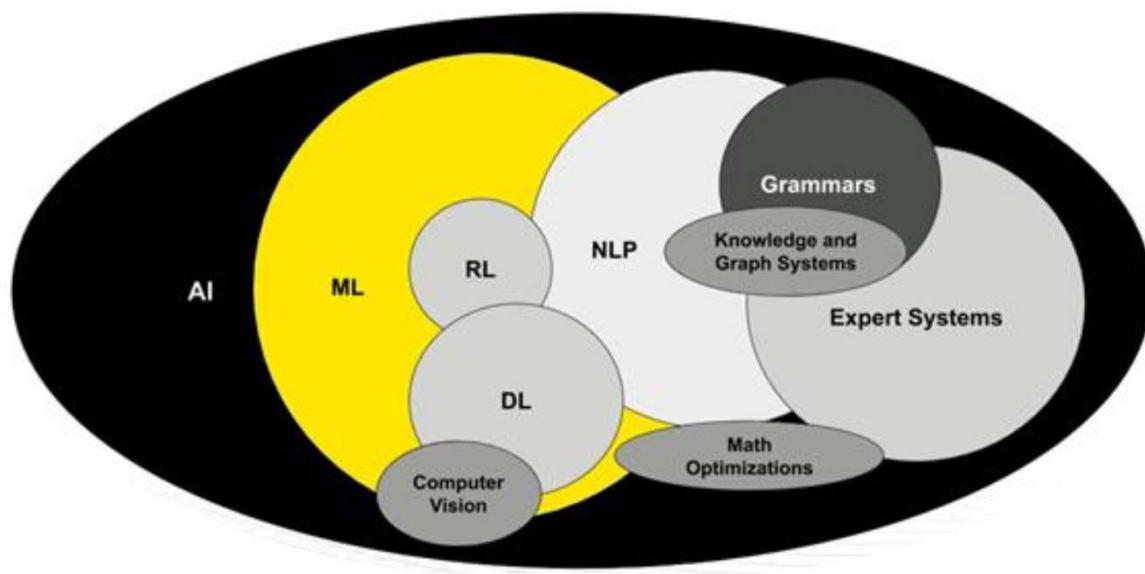


# And then? Training on the edge

- Distributed learning
  - Each device contribute to the overall training with locally computed samples



# Towards general AI (?)



# (A)IoT review

---

- Built to expand interdependence of humans and machines
  1. Interact
  2. Contribute
  3. Collaborate
- Benefits
  1. Efficient resource utilization (infer & control)
  2. Minimize human effort (better use of time)
  3. Save time and achieve higher productivity (make only what you need)
  4. Cross-field advancement (e.g. AI)
- Costs
  1. Energy to store/process massive amount of data
  2. Emission/CO2/waste new physical infrastructure needed
  3. Privacy

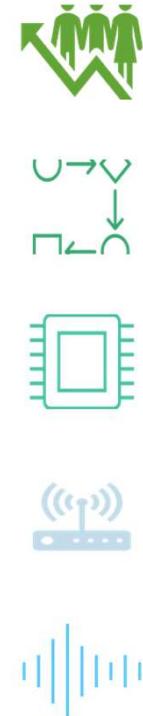
# (A)IoT summary (I)

---

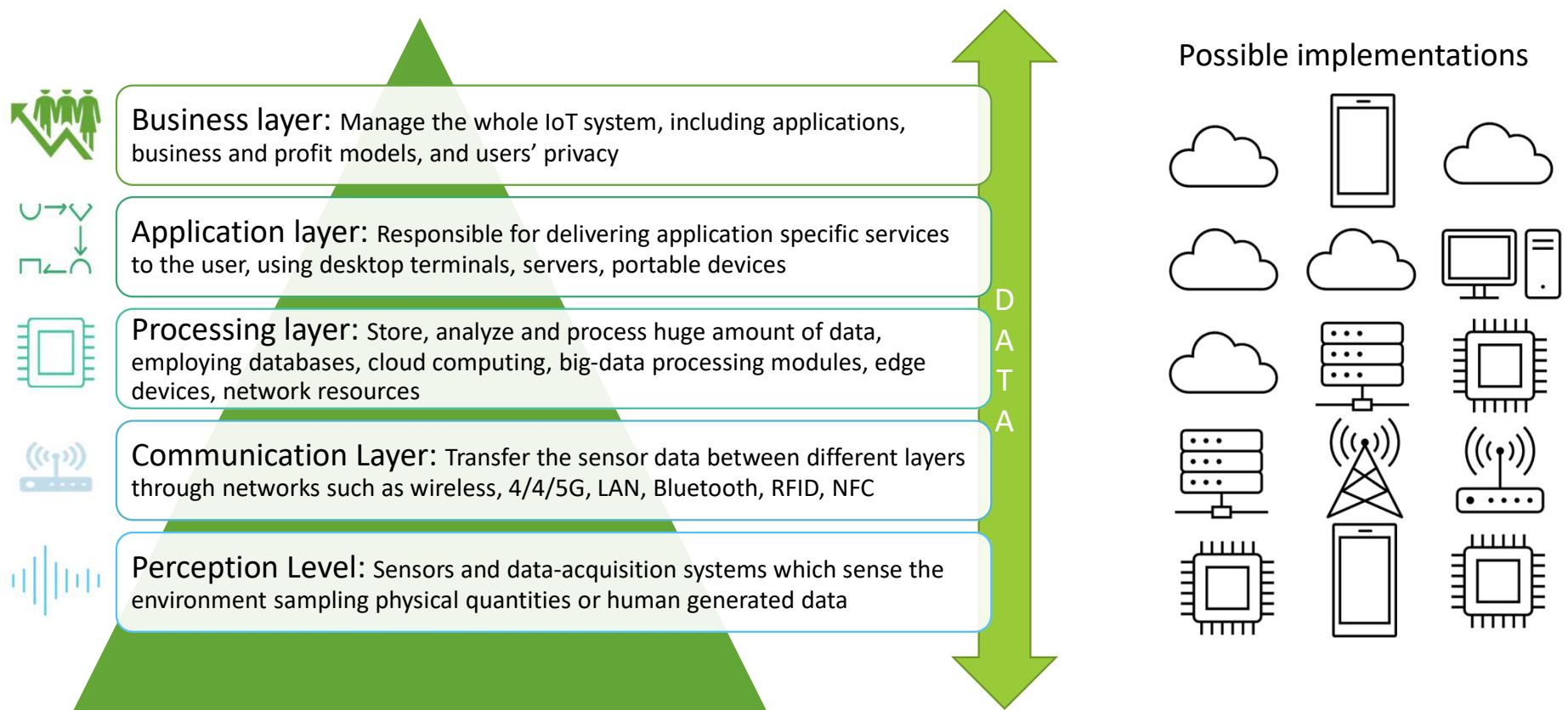
- Key features
  - 1. Integrate
    - A transparent multi-layer abstraction: let business managers send/receive informative data, from mobile apps and enterprise, independent of device connectivity
    - Event Store: query and visualize massive amount of data in a compact form for easy business management
  - 2. Virtualize:
    - device management: create objects identity and let them be controlled easily, following standard protocols
    - high-speed messaging: bi-directional communication among devices, remote servers, cloud resources and humans for dispatching useful information
  - 3. Analyze
    - Stream processing: real-time analysis of incoming data streams with event aggregation, filtering and correlation
    - Data enrichment: enrich raw data streams with contextual information and generate composite streams of information

# IoT architecture

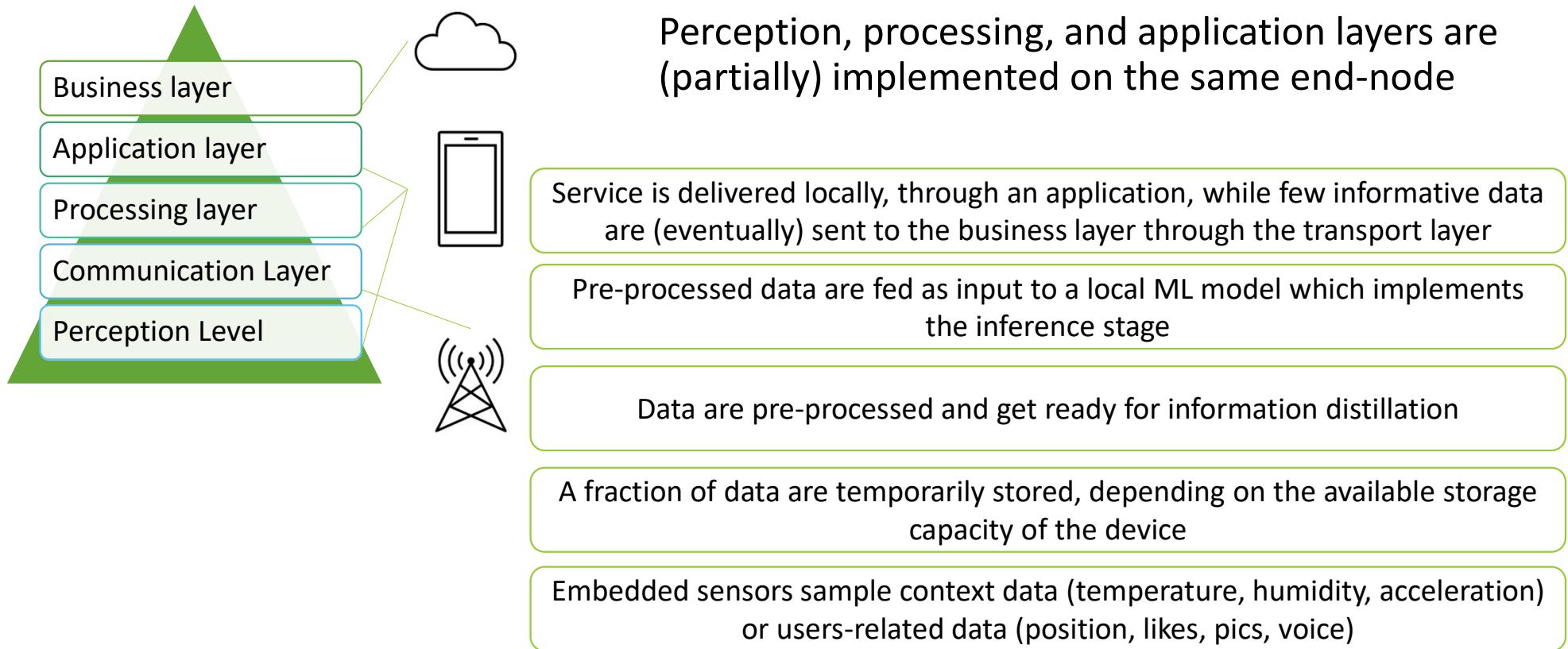
- There is no single consensus on a universal IoT implementation
  - may differ on
    - the context and application domain
    - the implemented functions and the requirements
    - the company organization
- It is however possible to generalize any IoT ecosystem using a basic multi-layer architectural template:
  - Each layer implements a specific function
  - The flow of data/information is vertical
  - Layers are orthogonal to the technologies in use
  - Physical place where layers are implemented is irrelevant (HW agnostic)



# A generic 5-layer architecture



# The case of edge-inference



# Basic (HW/SW) IoT components

Perception SENSING	Processing STORAGE/ANALYTICS	Communication SEND/RECEIVE	Application USERS-INTERFACE
<p>HW sensors</p> <ul style="list-style-type: none"><li>• Different types</li><li>• Small form factor</li><li>• Low-cost &amp; Low-energy</li><li>• Speed &amp; accuracy</li></ul> <p>SW Drivers</p> <ul style="list-style-type: none"><li>• Programming interface</li><li>• Multi-platform</li></ul>	<p>HW processing units</p> <ul style="list-style-type: none"><li>• GPU, CPU, MCU, ASIC</li><li>• High-perf. or low-power</li><li>• MFLOPs</li></ul> <p>HW storage</p> <ul style="list-style-type: none"><li>• NVMe</li><li>• RAM</li><li>• Capacity/Speed</li></ul> <p>SW Analytics</p> <p>SW Compiler</p> <ul style="list-style-type: none"><li>• High-level description languages</li><li>• Cross platform</li></ul> <p>SW DBMSs (!)</p>	<p>HW units and protocols</p> <p>Low-range</p> <ul style="list-style-type: none"><li>• NFC, RFID, WSN</li></ul> <p>Long-distance</p> <ul style="list-style-type: none"><li>• Wifi/eth cards</li><li>• Routers, linkers, gateways</li><li>• Speed &amp; Power consumption</li></ul> <p>SW network management [OSI model]</p>	<p>Service dependent</p> <p>Many SW programming components</p> <p><b>SW middleware:</b> acts as a bridge between an operating system or database and applications, especially on a network, enabling the management of heterogenous resources</p> <ul style="list-style-type: none"><li>• Interoperability</li><li>• Programming abstraction</li><li>• Device abstraction</li></ul>

# Basic (HW/SW) IoT components

Perception SENSING	Processing STORAGE/ANALYTICS	Communication SEND/RECEIVE	Application USERS-INTERFACE
HW sensors <ul style="list-style-type: none"><li>• Different types</li><li>• ...</li><li>• ...</li></ul>	HW processing units <ul style="list-style-type: none"><li>• GPU, CPU, MCU, ASIC</li></ul>	HW units Low-range	Service dependent
It is not the design of each component that makes IoT challenging, is the integration and proper balancing under tight resources constraints			
Programming interface <ul style="list-style-type: none"><li>• Multi-platform</li></ul>	Capacity/Speed <ul style="list-style-type: none"><li>SW Analytics</li><li>SW Compiler<ul style="list-style-type: none"><li>• High-level description languages</li><li>• Cross platform</li></ul></li><li>SW DBMSs (!)</li></ul>	Speed & Power consumption <ul style="list-style-type: none"><li>SW network management [OSI model]</li></ul>	System of database and applications, especially on a network, enabling the management of heterogenous resources <ul style="list-style-type: none"><li>• Interoperability</li><li>• Programming abstraction</li><li>• Device abstraction</li></ul>
ML4IoT	Part-I: Welcome to the IoT		43

# The concept of «constrained» design

- The term constrained is an important concept in understanding IoT devices, data, and impacts on analytics.
- It refers to the limited amount of resources that must be considered in the design of IoT ecosystems.
  - End-nodes
    - Power and energy budget (battery)
    - Data bandwidth
    - Computing power
  - Server/cloud
    - Storage & Computing power (millions of users)
    - Maintenance costs
    - Energy consumption and Carbon footprint
- For many IoT use cases, one or more of these must be balanced with the need to collect/record/process useful data/information.



# Energy constraint in mobile devices

## The Phones With the Longest-Lasting Batteries

Phones with longest use duration\* and biggest battery capacity (in milliampere hours)

	Use duration	Battery capacity
Motorola Moto G7 Power	20h8min	5,000
Xiaomi Mi Max 2	17h22min	5,300
LG X Power	15h18min	4,100
Motorola Moto E5 Plus	15h8min	5,000
BLU Studio Energy	14h53min	5,000
Motorola Moto G8 Plus	14h29min	4,000
Huawei Mate 20	14h26min	4,000
Asus ROG Phone II	14h11min	6,000
Motorola Moto Z Play Droid	13h43min	3,510
Xiaomi Redmi 3S	13h39min	4,100
Apple iPhone 11	13h29min	3,110

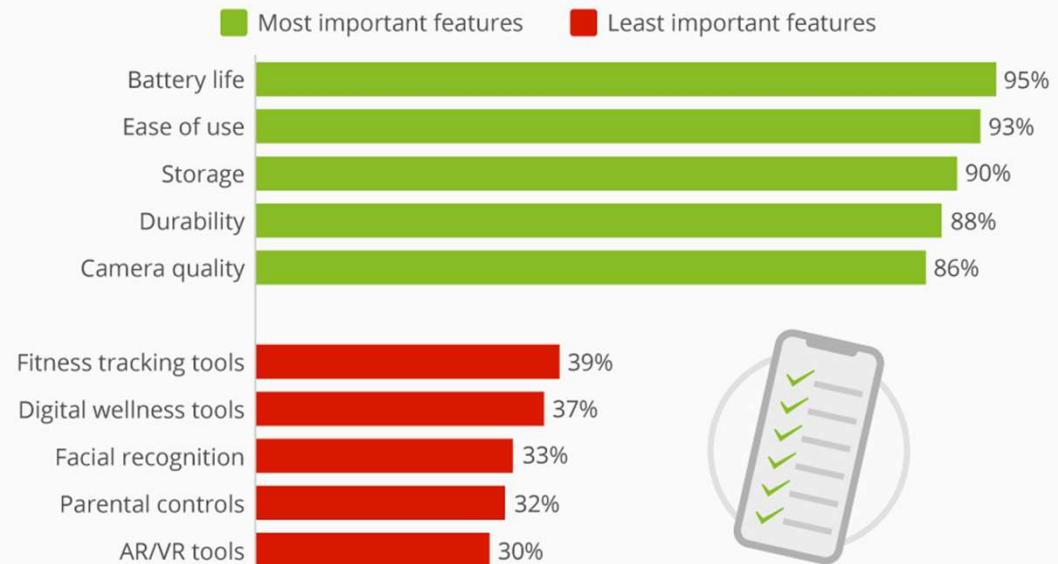
\* duration measured by custom web script replicating real-life use (Phone Arena)  
Source: Phone Arena



statista

## What Smartphone Buyers Really Want

Features considered somewhat/very important when deciding which smartphone to buy



CC BY SA  
@StatistaCharts

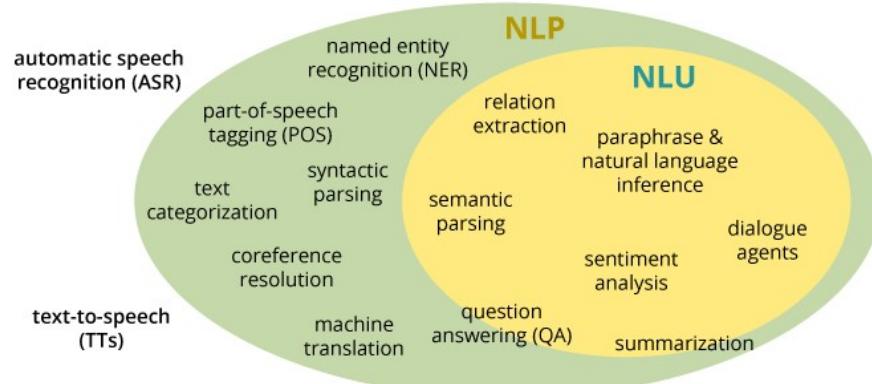
Based on a survey among 1,894 U.S. smartphone owners conducted in November 2018

Source: Morning Consult

statista

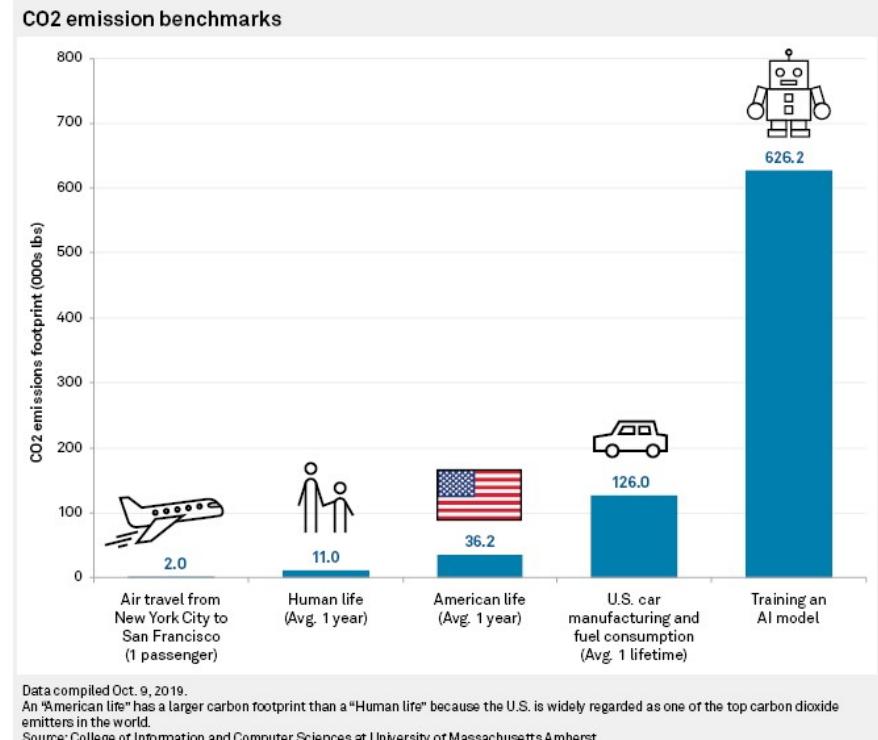
# Energy/Emission (CO2) constraints in cloud

- Training Natural Language Processing models

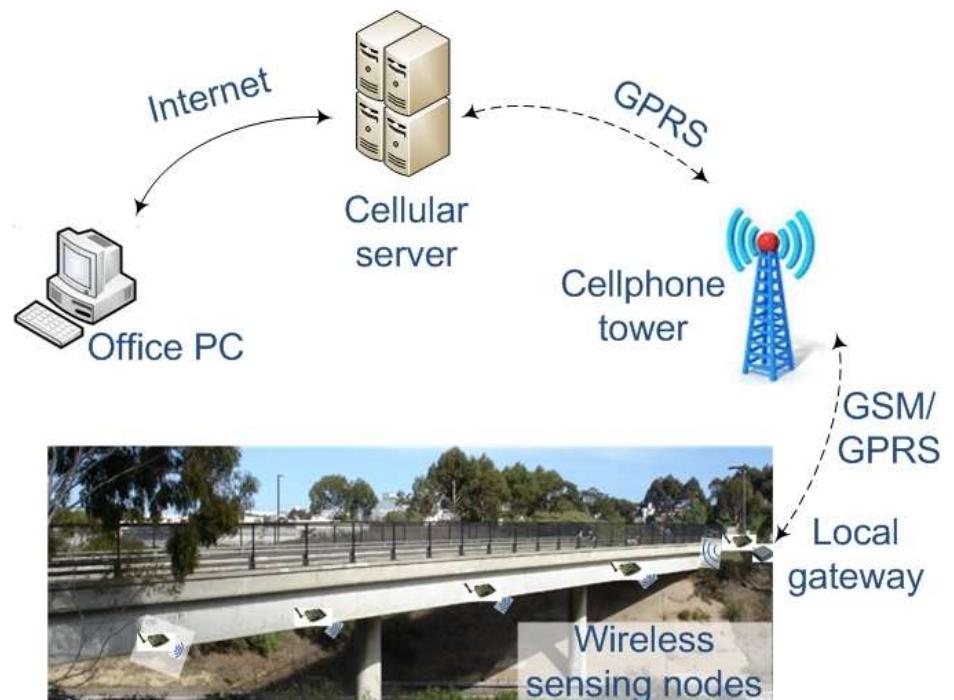
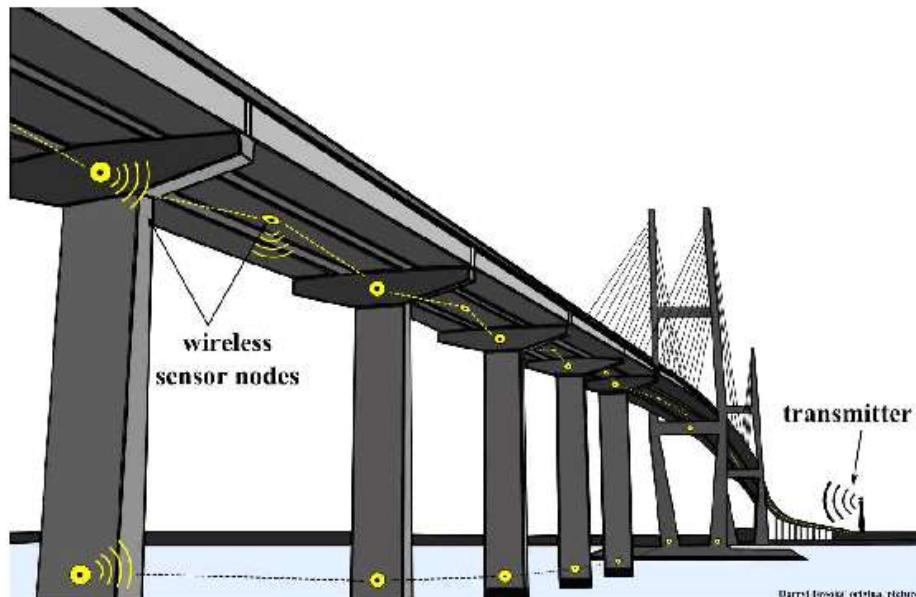


Model	Hardware	Power (W)	Hours	kWh-PUE	CO <sub>2</sub> e	Cloud compute cost
Transformer <sub>base</sub>	P100x8	1415.78	12	27	26	\$41–\$140
Transformer <sub>big</sub>	P100x8	1515.43	84	201	192	\$289–\$981
ELMo	P100x3	517.66	336	275	262	\$433–\$1472
BERT <sub>base</sub>	V100x64	12,041.51	79	1507	1438	\$3751–\$12,571
BERT <sub>base</sub>	TPUv2x16	—	96	—	—	\$2074–\$6912
NAS	P100x8	1515.43	274,120	656,347	626,155	\$942,973–\$3,201,722
NAS	TPUv2x1	—	32,623	—	—	\$44,055–\$146,848
GPT-2	TPUv3x32	—	168	—	—	\$12,902–\$43,008

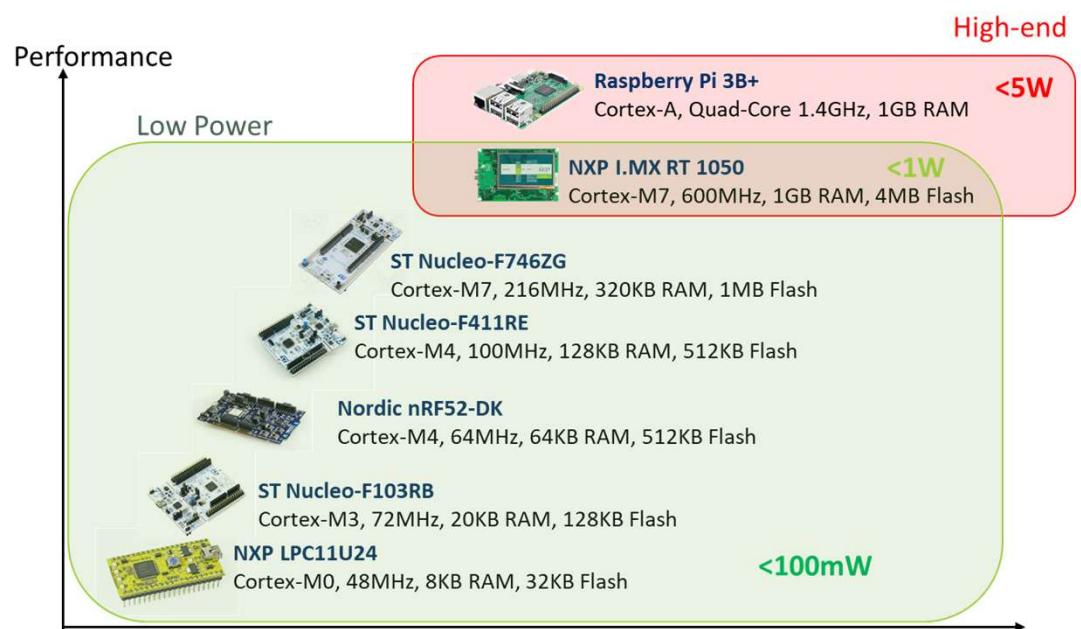
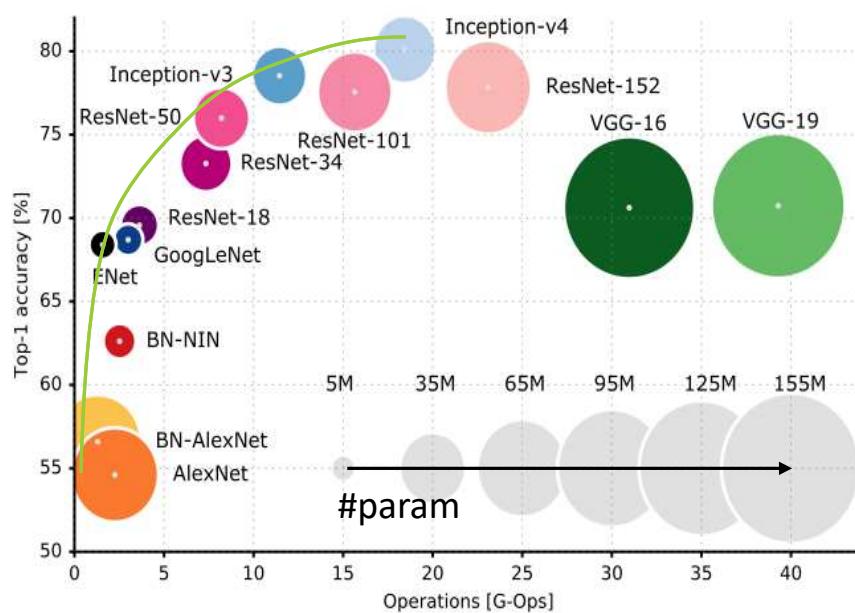
Energy and Policy Considerations for Deep Learning in NLP



# Even more challenging for WSNs

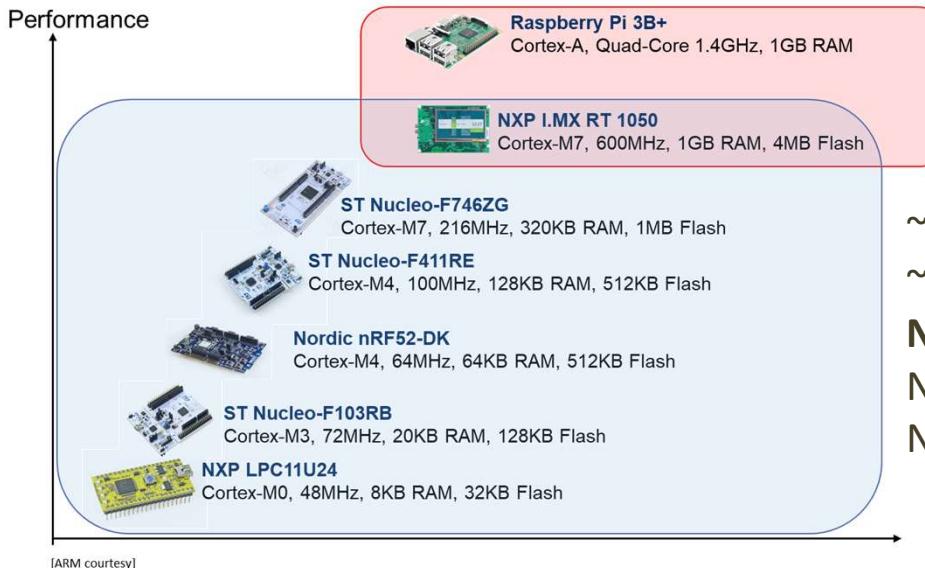
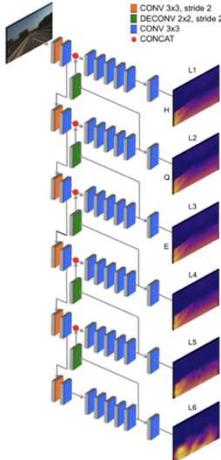


# The edge-AI case: compute&memory constraints



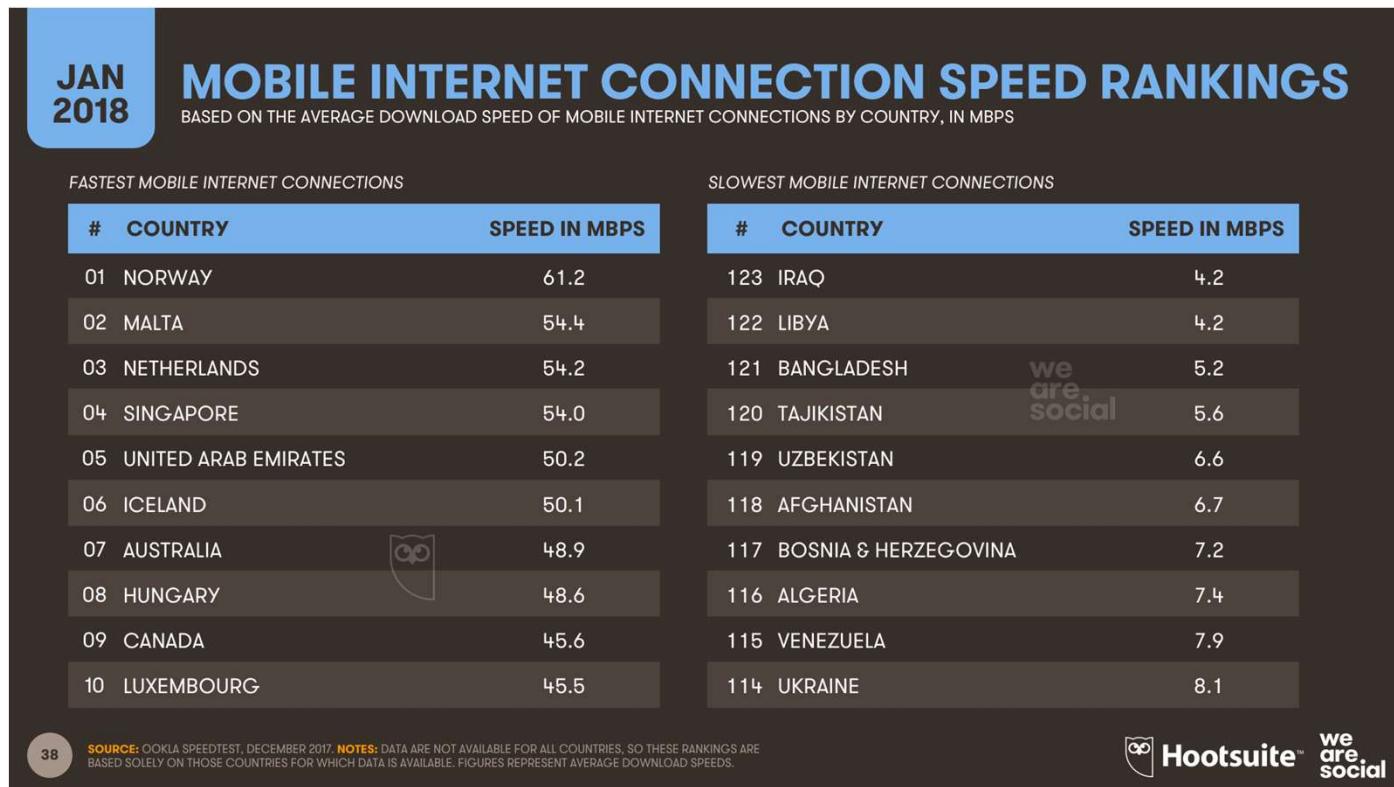
# Monocular depth perception

- Pyramidal Deep Net (PyD-Net): Lightweight CNN architecture for mobile CPUs
  - 10x smaller than its predecessors, almost same accuracy
  - FLASH=7.9MB (32-bit FP, 1.9 Mparams), RAM=264MB (active memory)



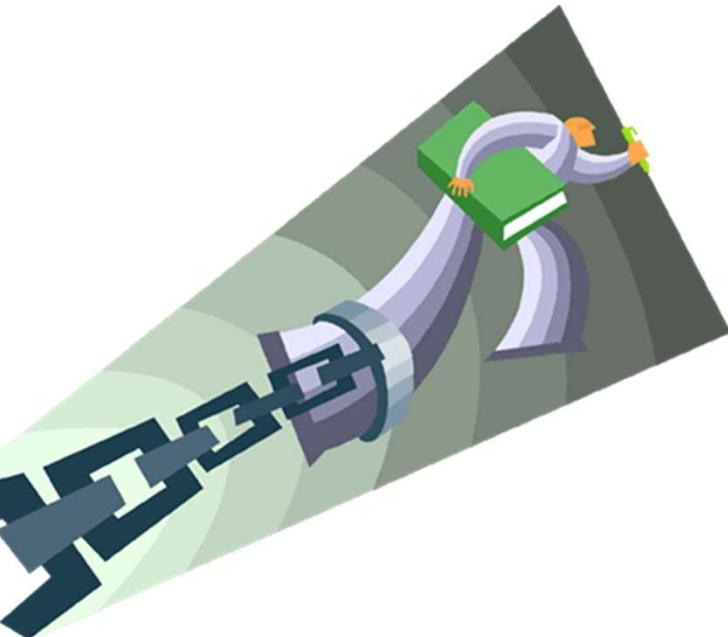
**~100 kB of memory**  
**~100 MHz of speed**  
**No floating point**  
No cache  
No media accelerator

# Latency constraint (edge<<>cloud)



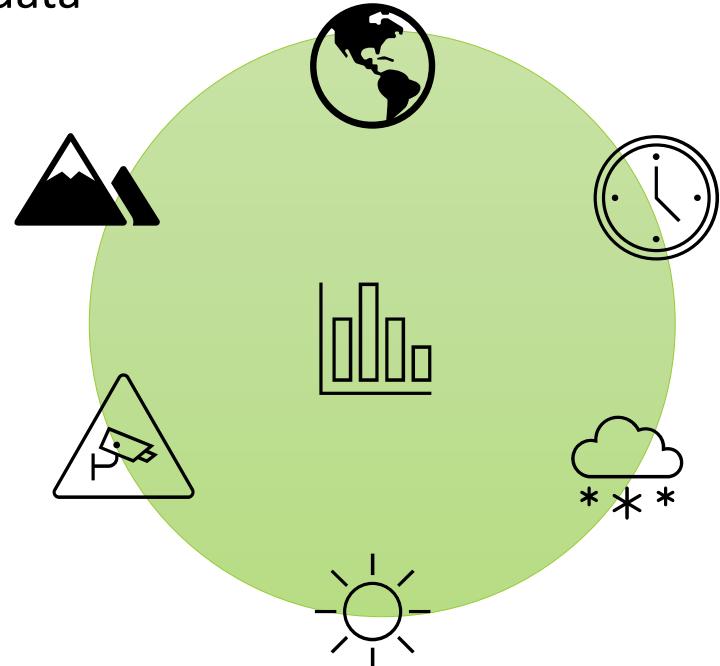
# Constrained designs

- Extra-functional properties are key
  - Latency
  - Energy
  - Power
  - Cost
  - Emissions
- Multi-objective optimization
  - For hardware: performance vs power vs energy
  - For software: model accuracy vs hardware-related metrics
  - HW/SW co-design
    - What and how to compute and where



# IoT analytics challenges (for a data scientist)

- IoT data are a nightmare for data scientist
- First, you know very few about the source of data
  - Their actual meaning and importance
  - The environment where they were collected
  - The conditions where they were
- Second, data lakes are
  - Hard to be handled due to their size and format
  - Noisy and incomplete
  - Subject of many (unknown) variations



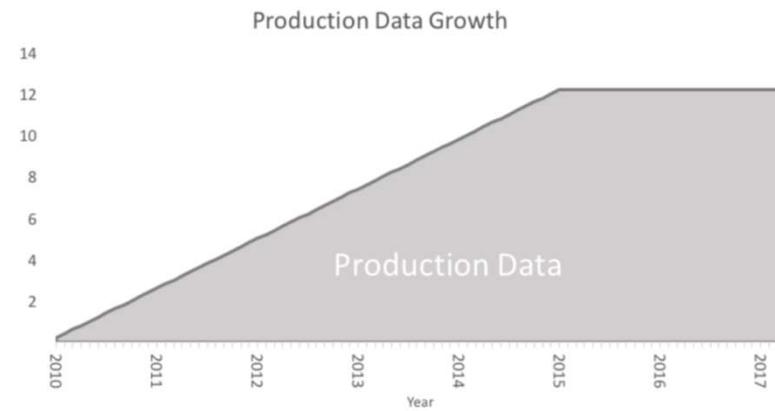
# Volume grows fast

---

- A company can easily have thousands to millions of IoT devices with several sensors on each unit, each sensor reporting values on a regular basis.
- The inflow of data can grow quite large very quickly, reaching volumes that many companies (not just the smallest one) can't afford.
- The sources of data are many
  - Users
    - The actual data sensed to deliver a specific application
  - Production:
    - Process parameters, manufacturing line
  - Supply-chain:
    - Logistics, transportation, delivery

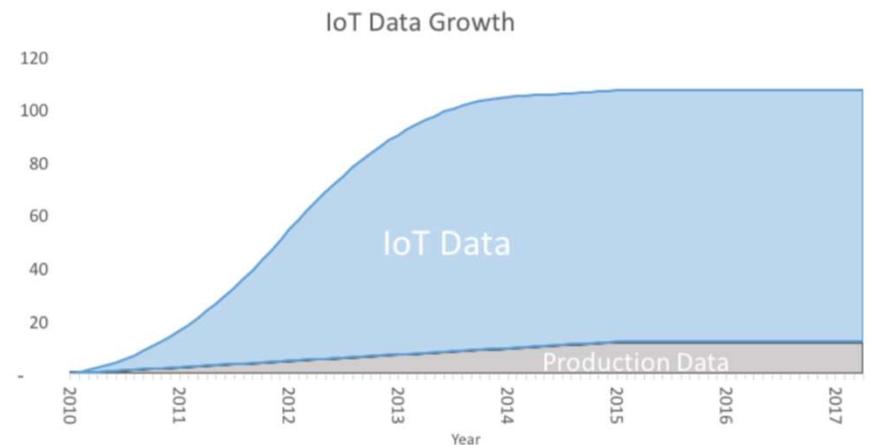
# Production vs. user data

- Imagine a manufacturing company producing small monitoring devices (temperature)
  - Production started in 2010 when the product was launched
  - It produces 12000 devices a year (1k units a month) – btw, in 2020 the number of temperature sensors sold worldwide is in the order of tens of billions
  - Each device is tested at the end of assembly
  - The values reported by the sensors on the device (200KB) are kept for analysis for 5 years
    - 2.4 GB/year



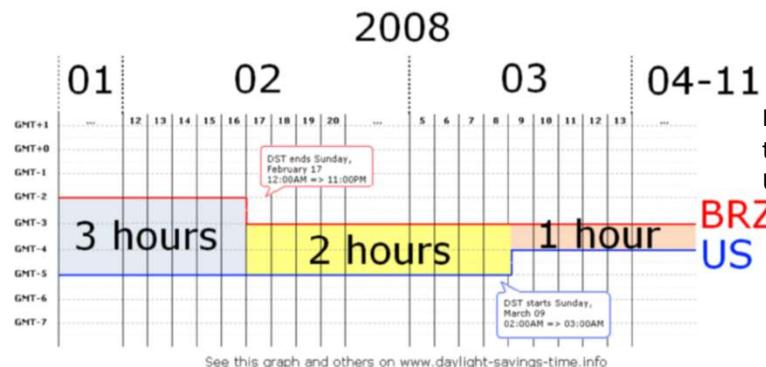
# Production vs. user data

- Imagine the device also had internet connectivity to track sensor values, and each one remains connected for two years.
- Each sensor sends 24 messages per day (1/h)
  - The message is 10KB large
  - 240KB/day/device  $\sim$  90MB/yr/device
  - 12000devices \* 90MB/yr/year  $\sim$  1.1 TB/yr
- And this is for a single type of sensor
- And we're not considering logistics
- And many other production data



# Timing issues

- IoT devices are spread out across the globe. Events that happen at the absolute same time do not happen at the same local time.
  - Time-zone and Time-standard (e.g. daylight savings time)
    - Different countries have different policies



In early 2008, Central Brazil was one, two, or three hours ahead of eastern U.S., depending on the date

- Clock synchronization issues
  - A device could be using the wrong time zone due to a configuration error.
  - It could also get out of sync due to a communication problem with the time standard source

# Timing issues (I)

---

- These issues may not be apparent to the engineer creating the device. His/her task is to design a device that determines if the spot is open or not, that's it. He/she may not appreciate the importance of writing code that captures a time value that can be aggregated across multiple time zones and locations.
- Moreover, the time available for analytics can be
  - the time the event occurred
  - or the time the IoT device sent the data
  - or the time the data was received
  - or the time the data was added to your data warehouse
  - ... many times you do not know anything about that, and you don't know who to ask

# Ooops!



## Mars Probe Lost Due to Simple Math Error

BY ROBERT LEE HOTZ

OCT. 1, 1999 12 AM PT

TIMES SCIENCE WRITER

NASA lost its \$125-million Mars Climate Orbiter because spacecraft engineers failed to convert from English to metric measurements when exchanging vital data before the craft was launched, space agency officials said Thursday.

A navigation team at the Jet Propulsion Laboratory used the metric system of millimeters and meters in its calculations, while Lockheed Martin Astronautics in Denver, which designed and built the spacecraft, provided crucial acceleration data in the English system of inches, feet and pounds.

As a result, JPL engineers mistook acceleration readings measured in English units of pound-seconds for a metric measure of force called newton-seconds.

In a sense, the spacecraft was lost in translation.

"That is so dumb," said John Logsdon, director of George Washington University's space policy institute. "There seems to have emerged over the past couple of years a systematic problem in the space community of insufficient attention to detail."

# Problems with space

---

- IoT devices are located in multiple geographic locations:
  - different environmental conditions can affect sensor accuracy.
    - External temperature induce sensors drifts. You could have less/more accurate readings depending on the country the sensor is placed (Calgary, Canada vs Singapore).
    - the available energy budget on the sensors can affect the frequency of data reporting.
      - For instance, many IoT devices are solar powered. A device in a often cloudy and rainy country will be more impacted than the same device located in a sunny region
  - elevation can affect equipment
    - Think about gas engines, which consume more at higher altitude; if location and elevation is not taken into consideration, you may falsely conclude from IoT sensor readings that a Denver-based fleet (1700m) of delivery trucks is poorly managing fuel economy compared to a fleet in Indiana (100m)
  - remote locations may have weaker network access.
    - The higher data loss could cause data values for those locations to be underrepresented in the resulting analytics.

# Problems with space (I)

---

- Data privacy and its regulation - General Data Protection Regulation
  - Privacy laws in Europe affect how the data from devices can be stored and what type of analytics is acceptable.
  - You may be required to anonymize the data from certain countries, which can affect what you can do with analytics
  - This also means a software differentiation depending on the destination of your product

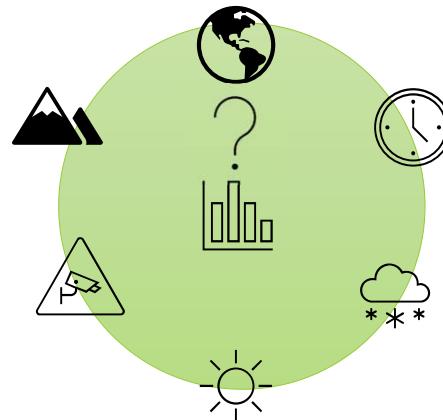
# Data quality

---

- Missing or inconsistent data lakes due to constrained devices
  - Messages lost in translation or never sent due to dead batteries or lossy networks
  - The conservation of power often means not all values available on the device are sent at the same time.
  - Software bugs can lead to garbled messages and data records.
  - Devices run on a software, called firmware, which may not be consistent across locations. This could mean differences in reporting frequency or formatting of values. It can result in mangled data
- Biased data
  - The missing data is often not random. For instance, it can be impacted by the location. Or other more complex social conditions
    - Young vs old users may affect the statistics significantly
    - Pandemic...
  - There can also be outside influences, such as environment conditions, that are not captured
    - Winter storms can lead to power failures affecting devices that are able to report back data.
  - Connectivity is a new thing for many devices, there is also often a lack of historical data to base predictive models on
    - newer products are overrepresented in the data simply because a higher percentage are now a part of the IoT

# What to do then?

As a data-scientist you must be aware of these things and take preventive or corrective actions (e.g. filling missing data, time alignment, etc) in order to avoid ending up with misleading analysis that impact your business value negatively.



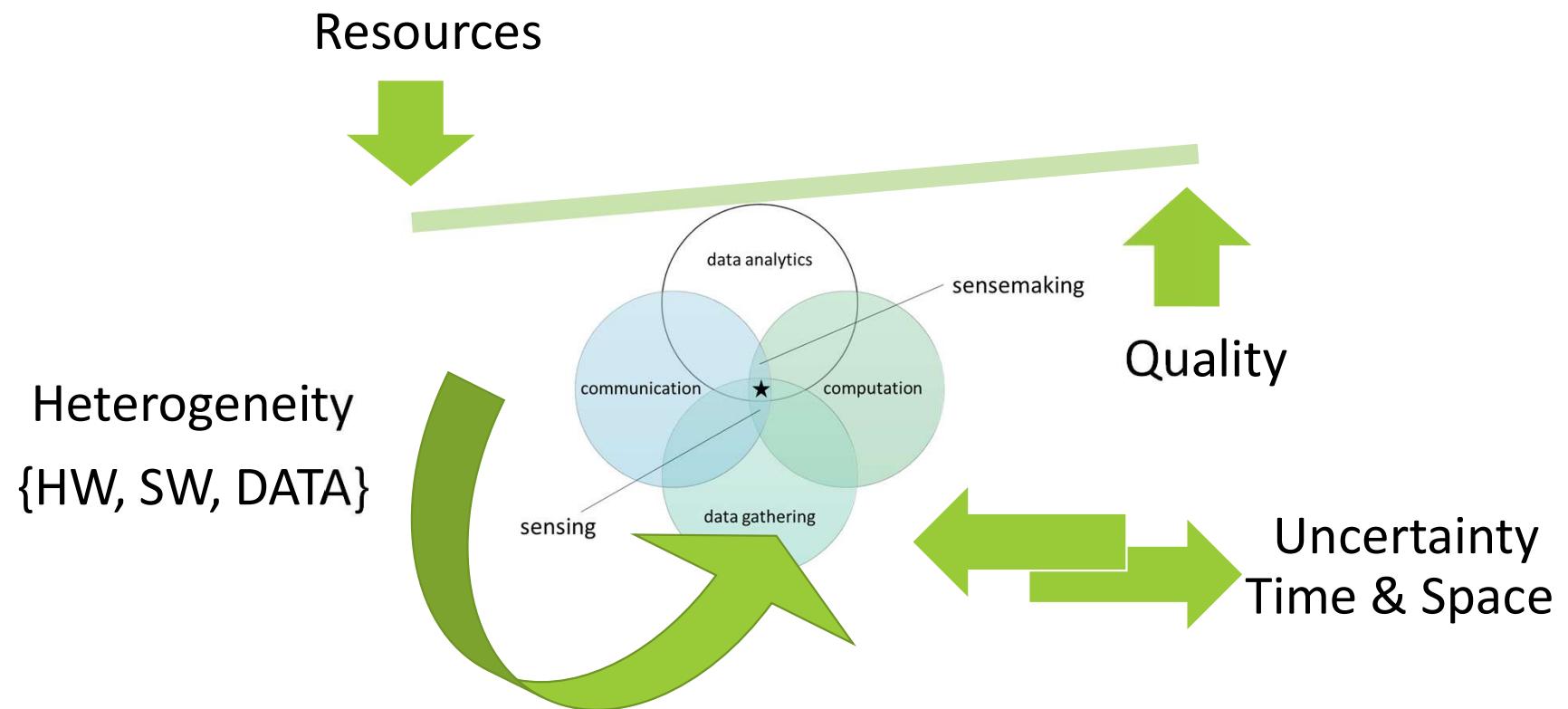
How? There is no recipe for that, but being wary and aware of your data is the first step

# Data exploration & visualization



<https://www.youtube.com/watch?v=VyhLRJV0lrl>

# The perfect storm



# Put hands on

- Software tools for implementing a typical data lifecycle in IoT
  - Data analysis
  - Sensors
    - Voice, Temperature, humidity, Images
    - Python libraries and their use on Raspberry-PI
  - Pre-Processing
    - Filtering, Time-frequency transformation
    - Run on end-nodes or servers
  - Communication
    - Device-to-remote server communication protocols
    - Raw-data and/or information
  - Training & Inference
    - Deep Convolutional Neural Networks
    - Google TensorFlow
  - Deployment
    - DNN optimization techniques for memory/latency/energy
    - Google TensorFlow-Lite
  - Performance assessment
    - edge-vs-remote computing

