AI & ML Workshop

What is all this hype about Embedded AI & ML?

Muthukumar

MACHINE LEARNING & AI?

Human Intelligence

- Solve problems
- Achieve goals
- Analyze & reason
- Communicate, collaborate & influence
- Consciousness,

Emotions, Intuition, Imagination

Artificial Intelligence

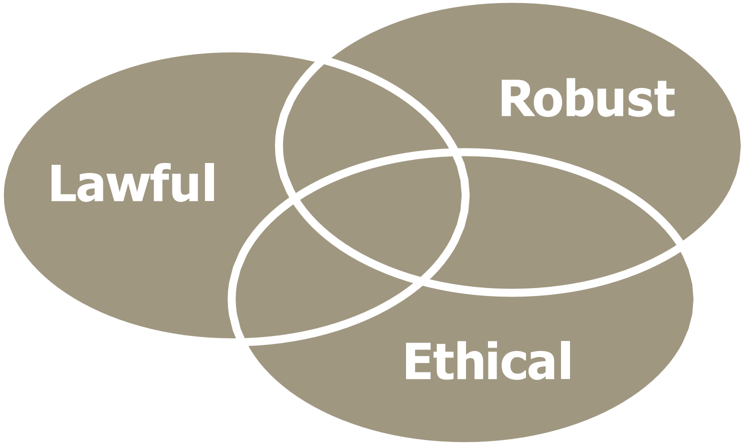
The ability for machines to simulate & enhance (human) intelligence

Trust in Humans

Selection of components

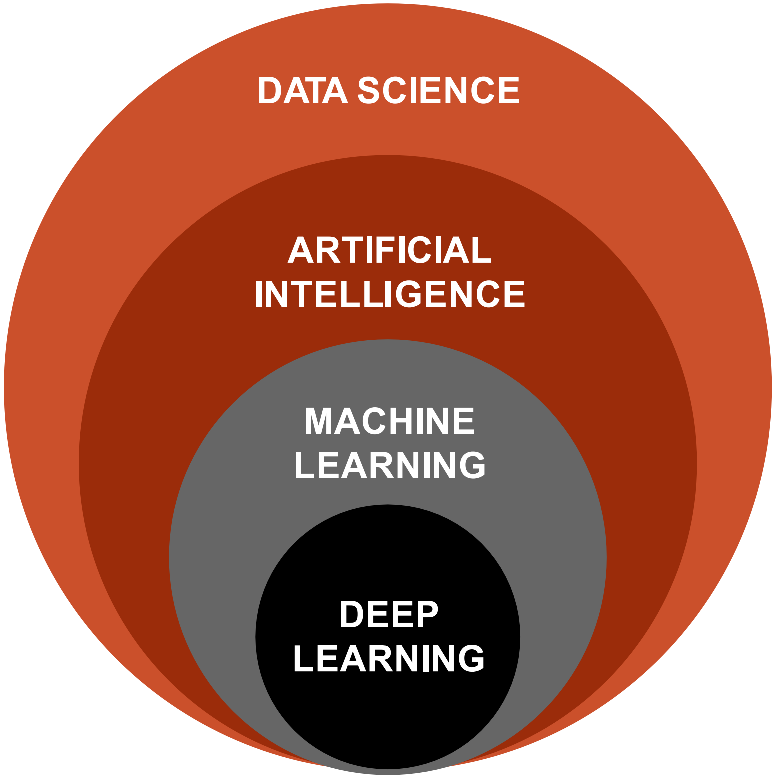
- Morals & Ethics
- Character
- Societal Laws
- Cultural Laws
- Compassion
- ….

Trust in AI



What is (Deep) Machine Learning?

1. Machine Learning is a subfield of Artificial Intelligence focused on developing algorithms that learn to solve problems by analyzing data for patterns
2. **Deep Learning** is a type of Machine Learning that leverages **Neural Networks** and **Big Data**



AI is **not new**, it's been around for a loong time

Mathematical Statistics

Artificial Intelligence

Machine Learning

From: AAAI Technical Report WS-98-05. Compilation copyright © 1998, AAAI (www.aaai.org). All rights reserved.

A Bayesian Approach to Filtering Junk E-Mail

Mehran Sahami* Susan Dumais† David Heckerman† Eric Horvitz†

*Gates Building 1A
Computer Science Department
Stanford University
Stanford, CA 94305-9010
sahami@cs.stanford.edu

†Microsoft Research
Redmond, WA 98052-6399
{sdumais, heckerma, horvitz}@microsoft.com

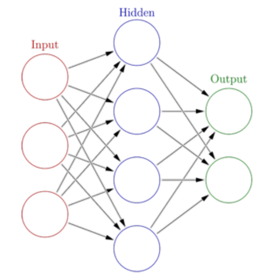
Abstract

In addressing the growing problem of junk E-mail on the Internet, we examine methods for the automated construction of filters to eliminate such unwanted messages from a user's mail stream. By casting this problem in a decision theoretic framework, we are able to make use of probabilistic learning methods in conjunction with a notion of differential misclassification cost to produce filters which are especially appropriate for the nuances of this task. While this may appear, at first, to be a straight-forward text classification problem, we show that by considering domain-specific features of this problem in addition to the raw text of E-mail messages, we can produce much more accurate filters. Finally, we show the efficacy of such filters in a real world usage scenario, arguing that this technology is mature enough for deployment.

contain offensive material (such as graphic pornography), there is often a higher cost to users of actually viewing this mail than simply the time to sort out the junk. Lastly, junk mail not only wastes user time, but can also quickly fill-up file server storage space, especially at large sites with thousands of users who may all be getting duplicate copies of the same junk mail.

As a result of this growing problem, automated methods for filtering such junk from *legitimate* E-mail are becoming necessary. Indeed, many commercial products are now available which allow users to hand-craft a set of logical rules to filter junk mail. This solution, however, is problematic at best. First, systems that require users to hand-build a rule set to detect junk assume that their users are savvy enough to be able to construct robust rules. Moreover, as the nature

Deep Learning



1700's | 1950's | 1960's | 1970's | 1980's | 1990's | 2000's | 2010's | Today | Future

1943 – The first ANN

1955 – Official term and academic recognition

1958 – Rosenblatt's Perceptron

1969 – Backpropagation

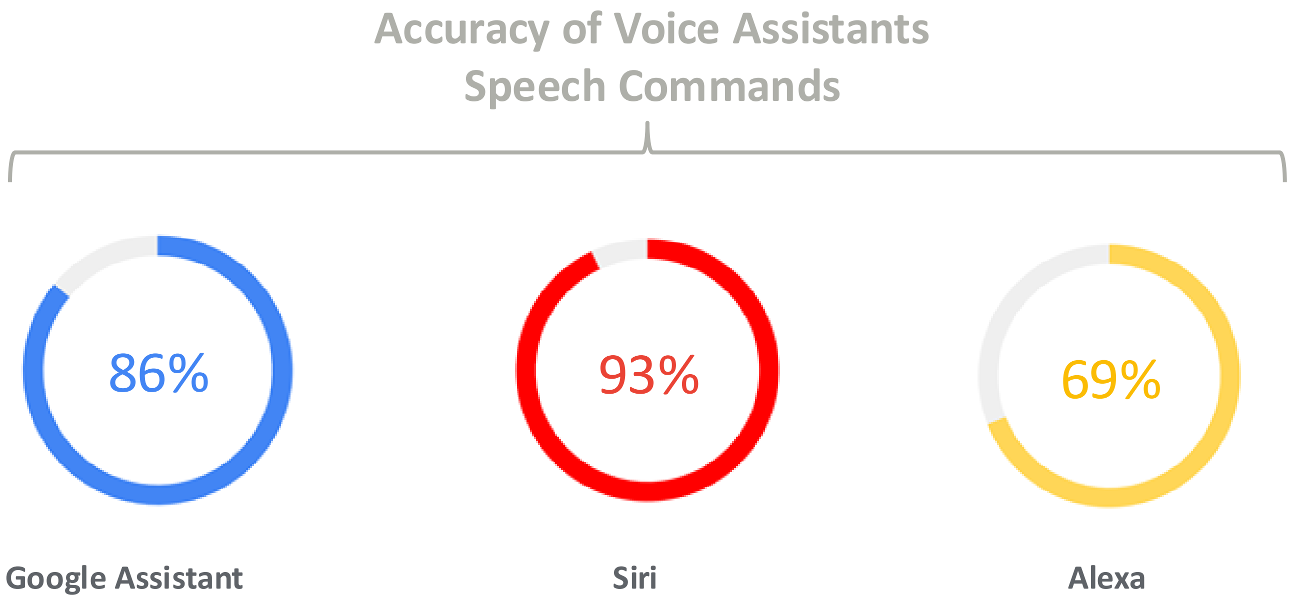
1985 – Rediscovery of Backprop

1996 – Chess victories – defeating the world champion

2012 – AlexNet wins ImageNet

2013 - Today: Deep Learning is applied almost everywhere!

AI isn’t perfect



Source: https://loupventures.com/annual-digital-assistant-iq-test/

Traditional Programming

Pros

- Quicker to build
- Easier to explain
- Easier to debug
- Easier to maintain
- More consistent/stable

```
if (speed < 4) {
    status = WALKING;
}
```

Cons

- Does not scale
- Does not adapt to changes
- Does not work for complex tasks

1

Machine Learning

Pros

- Complex problems
- Scale
- Adaptable
- Personalization
- Improves over time

```
0101001010100101010
1001010101001011101
0100101010010101001
0101001010100101010
```

Label = WALKING

Cons

- Slower to build
- Harder to explain/interpret
- Harder to debug

Answers → Machine learning → Rules — Training phase

Data →

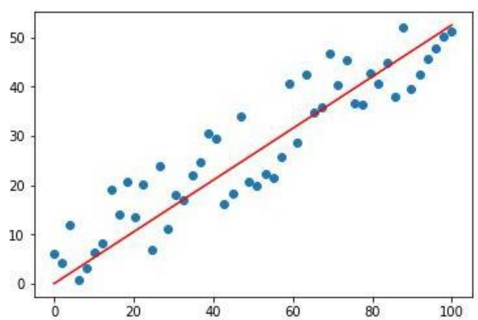
2

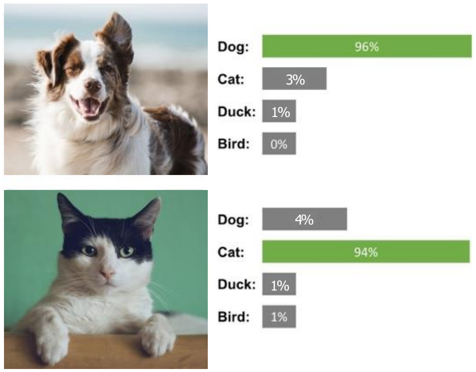
Data → Rules → Predictions — Inference phase

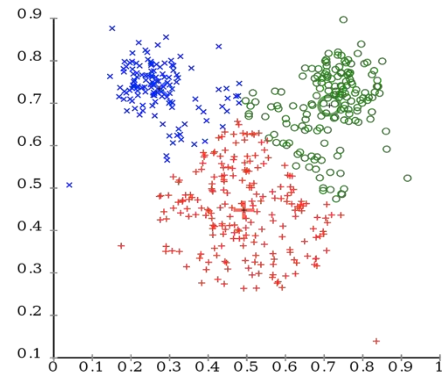
The Categories of Machine Learning

Machine Learning

- Supervised learning
  - Regression
  - Classification
- Unsupervised learning
  - Clustering, Segmentation etc
- + Reinforcement Learning







What is different? - 4 key enablers

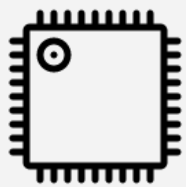
1. Data availability
2. Computational power
3. Algorithm advancements
4. Broad public interest

The flipsides of the 4 key AI enablers...

1

Data availability

Violate privacy & data integrity

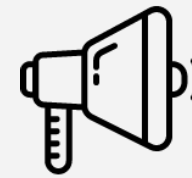
2

Computational power

Energy & capital intensive

3

Algorithm advancements

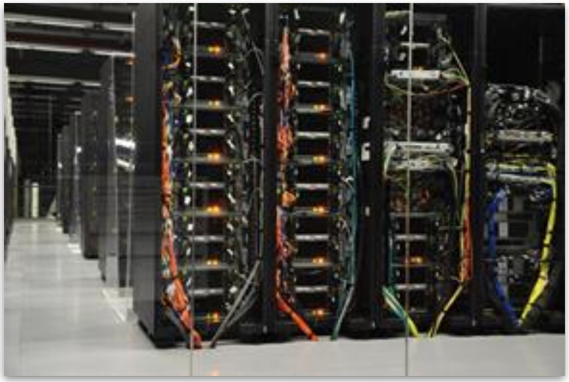
Introduction of biases & opacity

4

Broad public interest

Hype vs reality

PERSONAL HARDWARE

Cloud Vs Edge based ML



High power
High bandwidth
High latency

Low power
Low bandwidth
Low latency

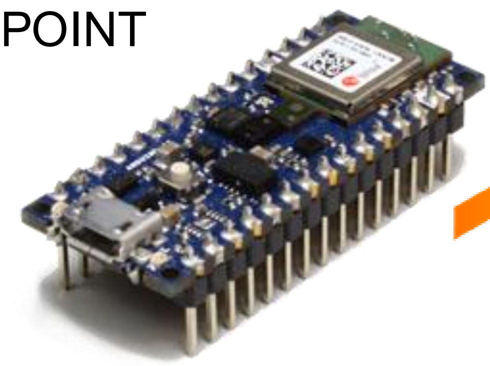
Edge & Endpoint Hardware

EDGE

Single Board Computer

More powerful (faster processor, more memory)
Runs full, general purpose operating system (OS)
Can provide full command line or graphical user interface
Requires more power

ENDPOINT



Microcontroller

Less powerful
Bare-metal (superloop) or real-time operating system (RTOS)
Limited or no user interface
Requires less power

Edge & Endpoint Devices

Node Devices

Edge Devices

Endpoints Have **Sensors**, Tons of Sensors

Motion Sensors
Gyroscope, radar, magnetometer, accelerator

Acoustic Sensors
Ultrasonic, Microphones, Geophones, Vibrometers

Environmental Sensors
Temperature, Humidity, Pressure, IR, etc.

Touchscreen Sensors
Capacitive, IR

Image Sensors
Thermal, Image

Biometric Sensors
Fingerprint, Heart rate, etc.

Force Sensors
Pressure, Strain

Rotation Sensors
Encoders

Data: No Good Data Left Behind

5 Quintillion

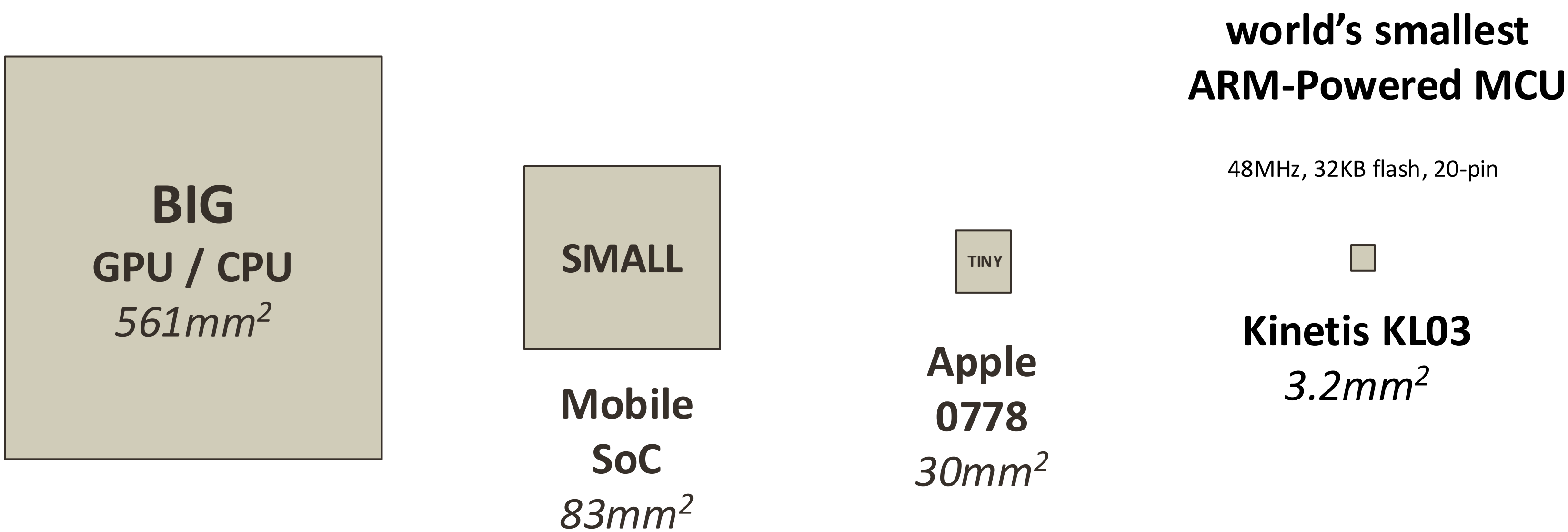
bytes of data produced every day by IoT

<1%

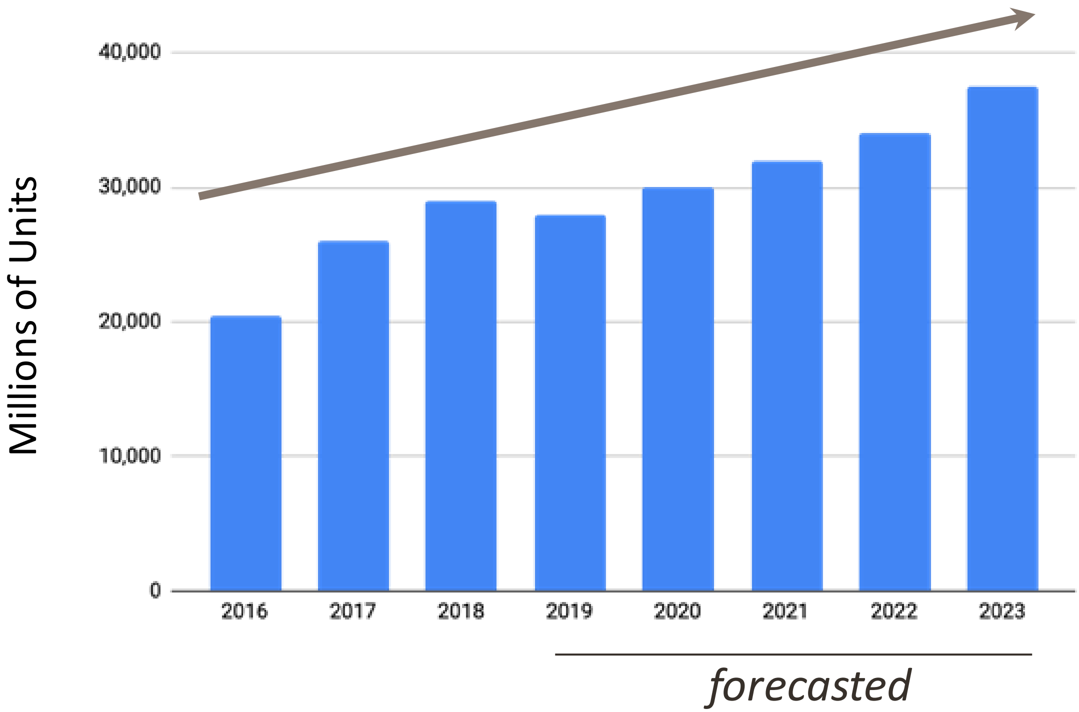
of unstructured data is analyzed or used at all

Source: Harvard Business Review, What's Your Data Strategy?, April 18, 2017
Cisco, Internet of Things (IoT) Data Continues to Explode Exponentially. Who Is Using That Data and How?, Feb 5, 2018

Shrinking Size of hardware

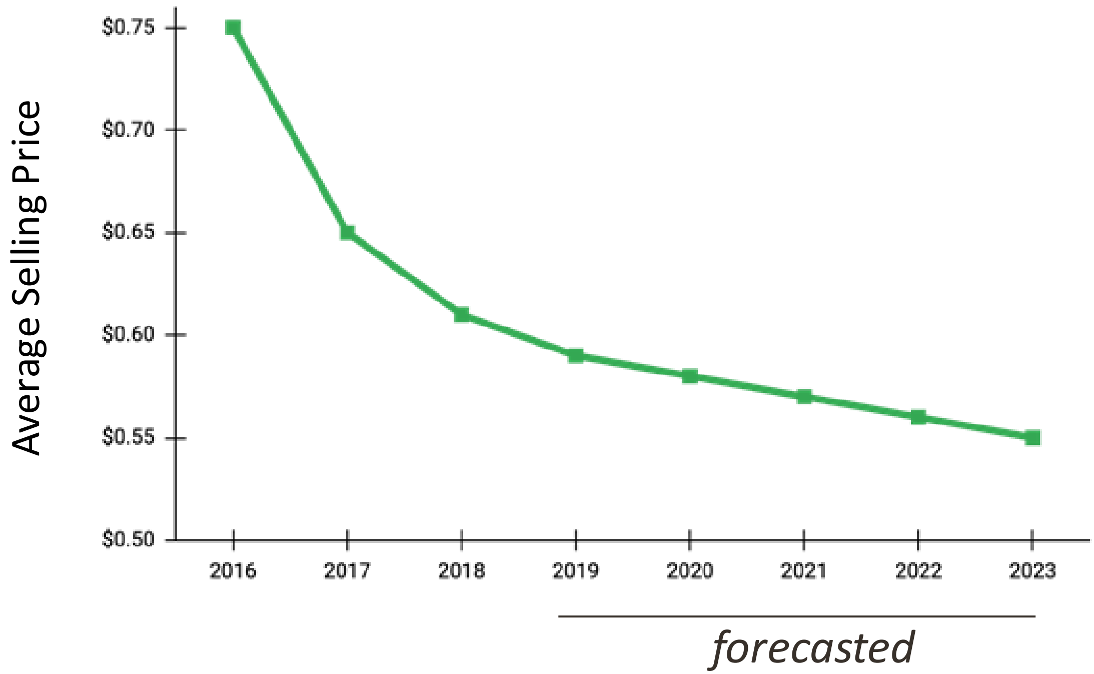


End device Usage: MCU Demand Forecast



Source: IC Insights

End device Pricing: MCU Pricing Forecast



Source: IC Insights

Hardware Power

BIG
GPU / CPU

300W
NVIDIA Tesla K80

SMALL

3.64W
Apple A12

Neural Decision Processor

Always-on deep learning speech/audio recognition

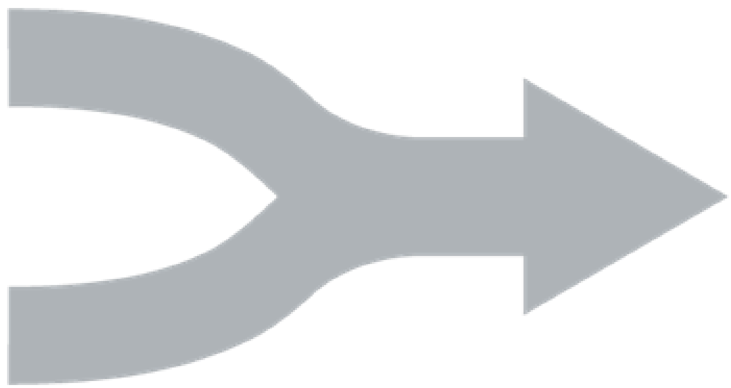
Ultra low power, 128KB SRAM, 12-pin, 2.52mm^2

140 µW
Syntiant NDP100

Birth of **EmbeddedML** or **TinyML**?

Embedded Systems

Machine Learning



TinyML
EmbeddedML

Window of Opportunity with **TinyML**

- Learning with limited memory and computation
- Battery-operated
- On-device computing
- Low latency
- Low cost
- Small size