

# Infrastructures for machine learning

Antonio Cisternino (@cisterni)  
IT Center @Unipisa



# Machine Learning hype in the news

---

WIRED

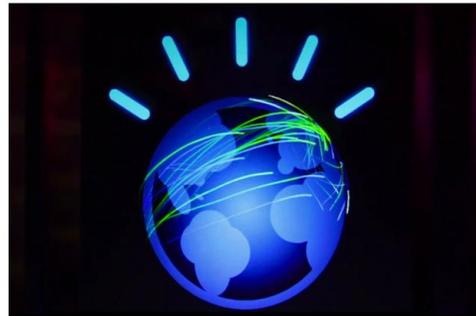
Technology | Science | Culture | Video | Reviews | Magazine | More ▾

Artificial Intelligence

## IBM's Watson is better at diagnosing cancer than human doctors



By IAN STEADMAN  
Monday 11 February 2013



Credit: IBM

IBM's [Watson](#) -- the language-fluent computer that beat the best human champions at a game of the US TV show *Jeopardy!* -- is being turned into a tool for medical diagnosis. Its ability to absorb and analyse vast quantities of data is, IBM claims, better than that of human doctors, and its deployment through the cloud could also reduce healthcare costs.

# Machine Learning hype in the news

WIRED

Technology | Science | Culture | Video | Reviews | Magazine | More ▾

Artificial Intelligence

## IBM's Watson is diagnosing cancer doctors



By IAN STEADMAN  
Monday 11 February 2013



Credit: IBM

IBM's Watson -- human champion -- being turned into an AI that can read and analyze vast amounts of human doctor data to reduce healthcare costs.

## Elon Musk leads 116 experts calling for outright ban of killer robots

Open letter signed by Tesla chief and Alphabet's Mustafa Suleyman urges UN to block use of lethal autonomous weapons to prevent third age of war



 A killer robot from the 2014 remake of RoboCop. The open letter read: 'lethal autonomous weapons will permit

# Machine Learning hype in the news

WIRED

Technology | Science | Culture | Video | Reviews | Magazine | More

Artificial Intelligence

## IBM's Watson is diagnosing cancer for doctors



By IAN STEADMAN  
Monday 11 February 2013

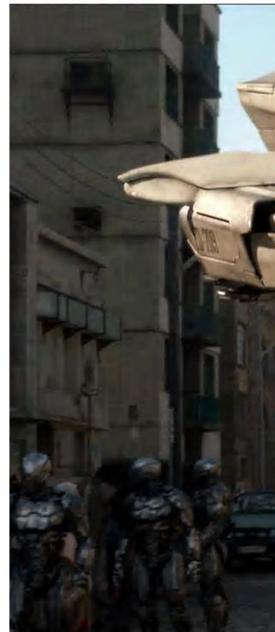


Credit: IBM

IBM's Watson -- human champion -- is being turned into a machine that can learn and analyse vast amounts of data to reduce healthcare costs.

## Elon Musk's open letter is an outright ban on

Open letter signed by Musk and others to block use of lethal autonomous weapons



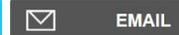
## 'Machine learning' is a revolution as big as the internet or personal computers



Drake Baer

Apr. 1, 2016, 2:34 PM

5,109



We're in the middle of a historic moment.

It used to be the case that you had to program a computer so that it knew how to do things. Now computers can learn from experience.

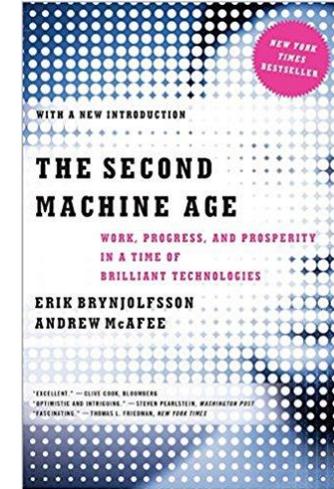
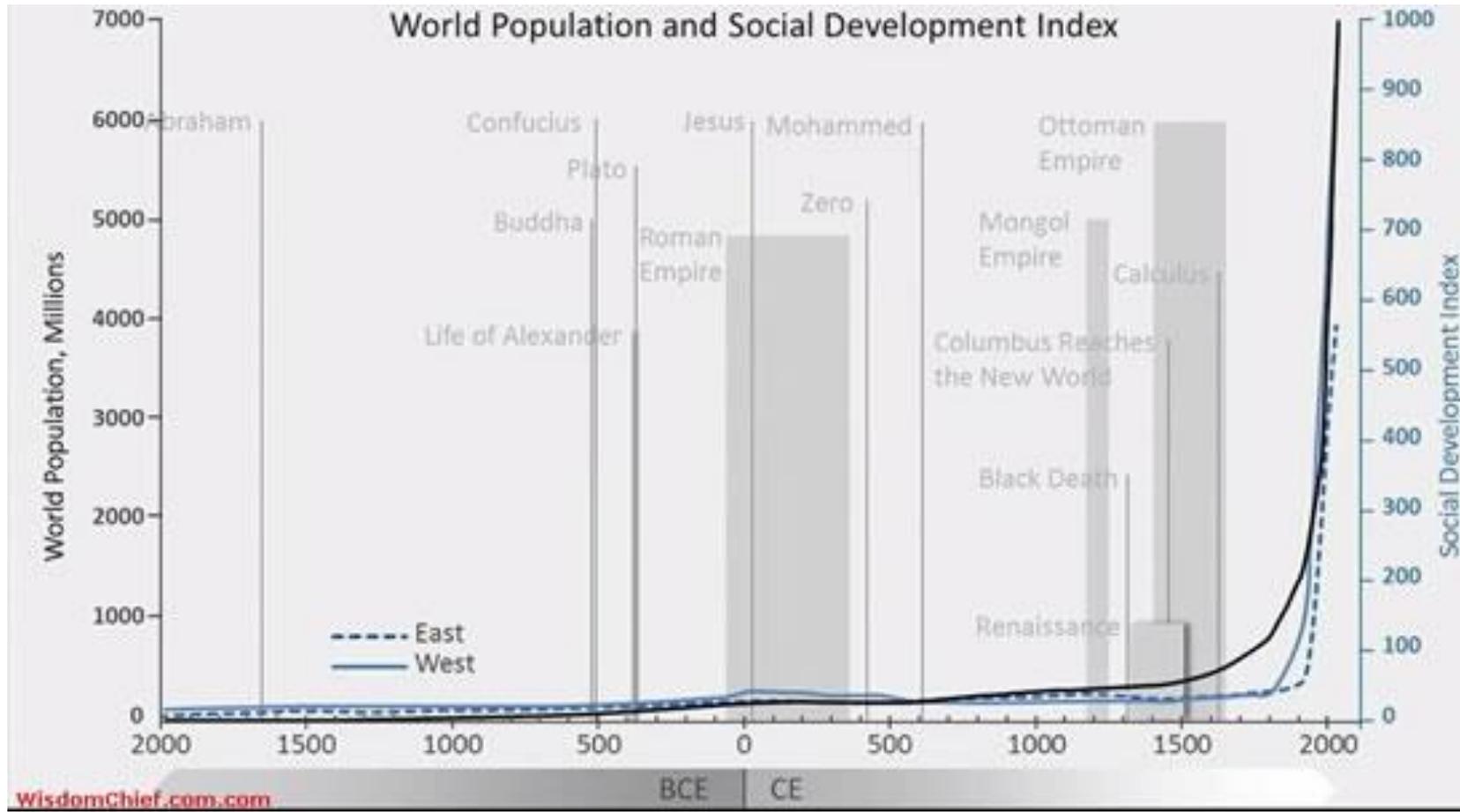
The breakthrough is called "machine learning." It's



Don't worry, the machines are your friend. Sean Gallup / Getty

A killer robot from the 2014 remake of RoboCop. The open letter read: 'lethal autonomous weapons will permit

# The second machine age: the automation of control



## Generating tables

---

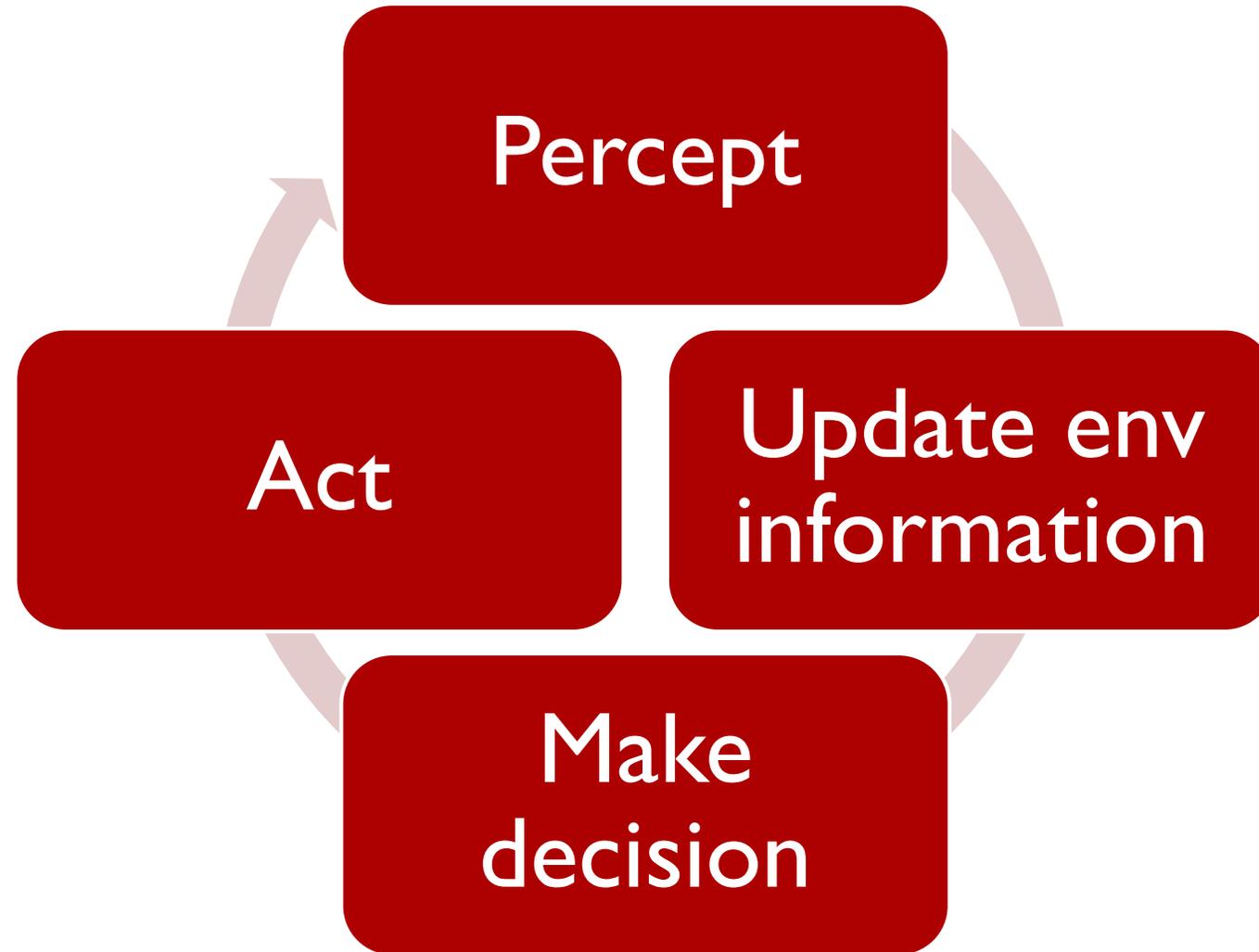


We will *initially* need trained mathematicians writing tables (i.e. computer programs) but eventually computers will generate tables (programs) automatically

Alan Turing working on Colossus (not literal citation)

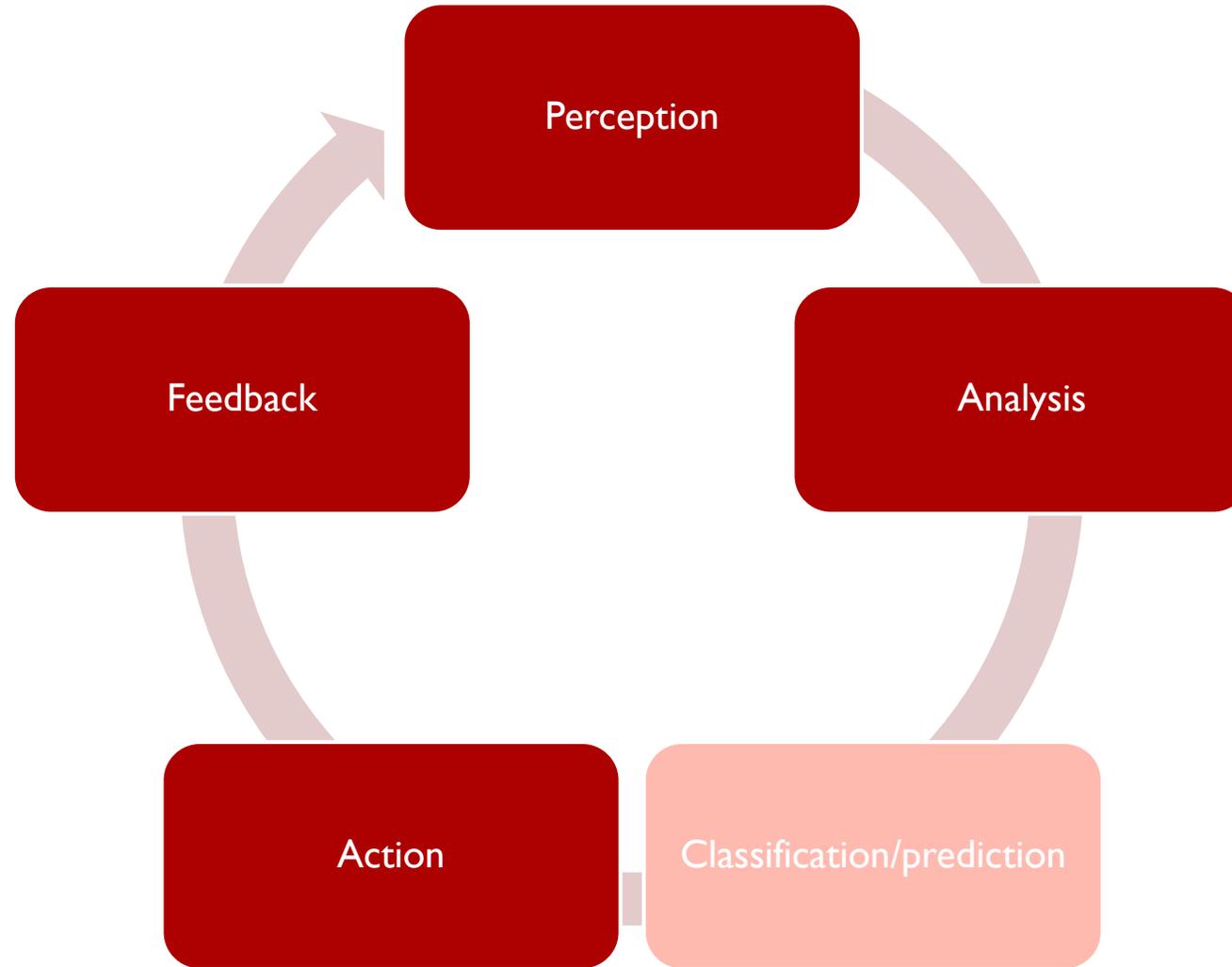
## A «functional» element

---



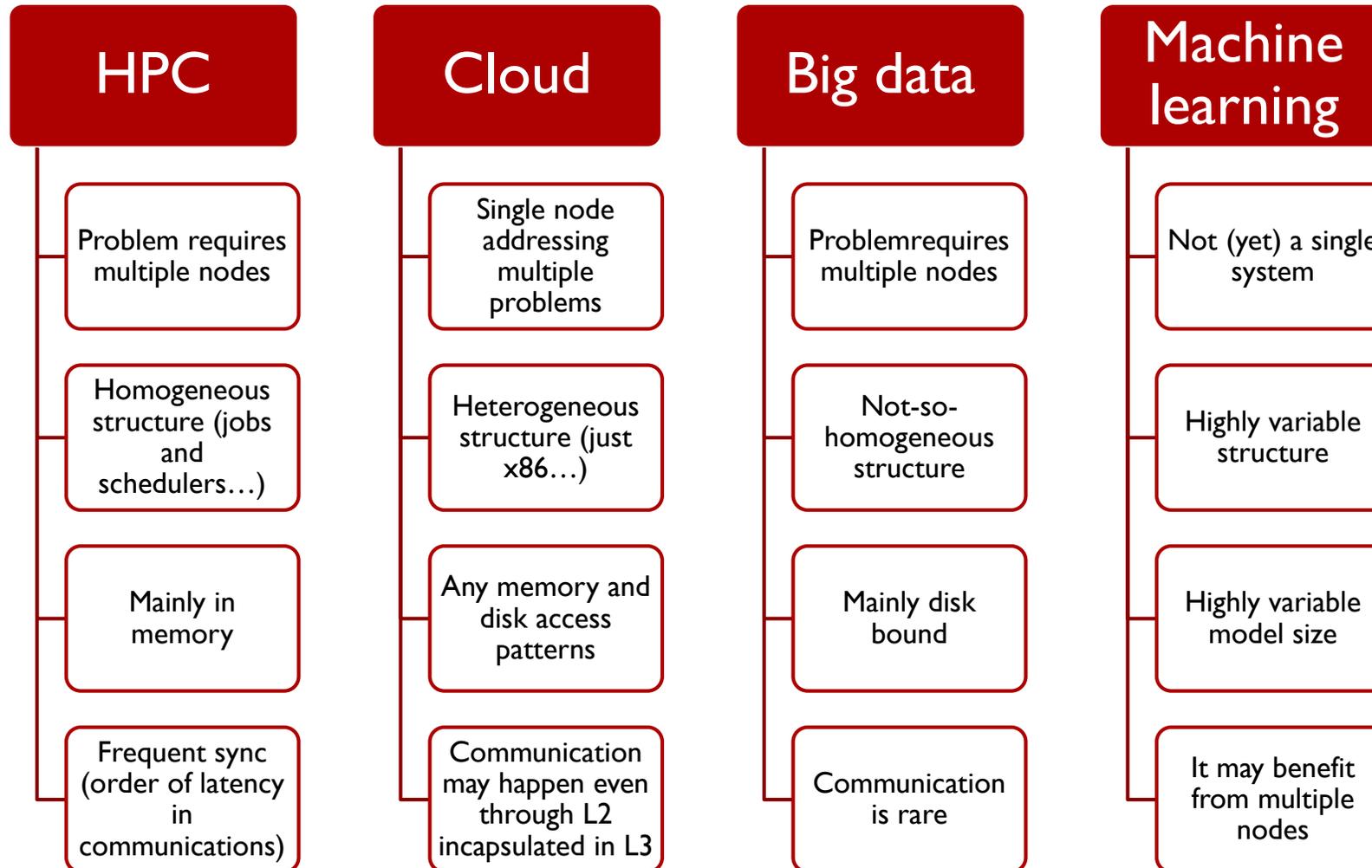
# A «functional» element

---



# Cloud, HPC and the problem size

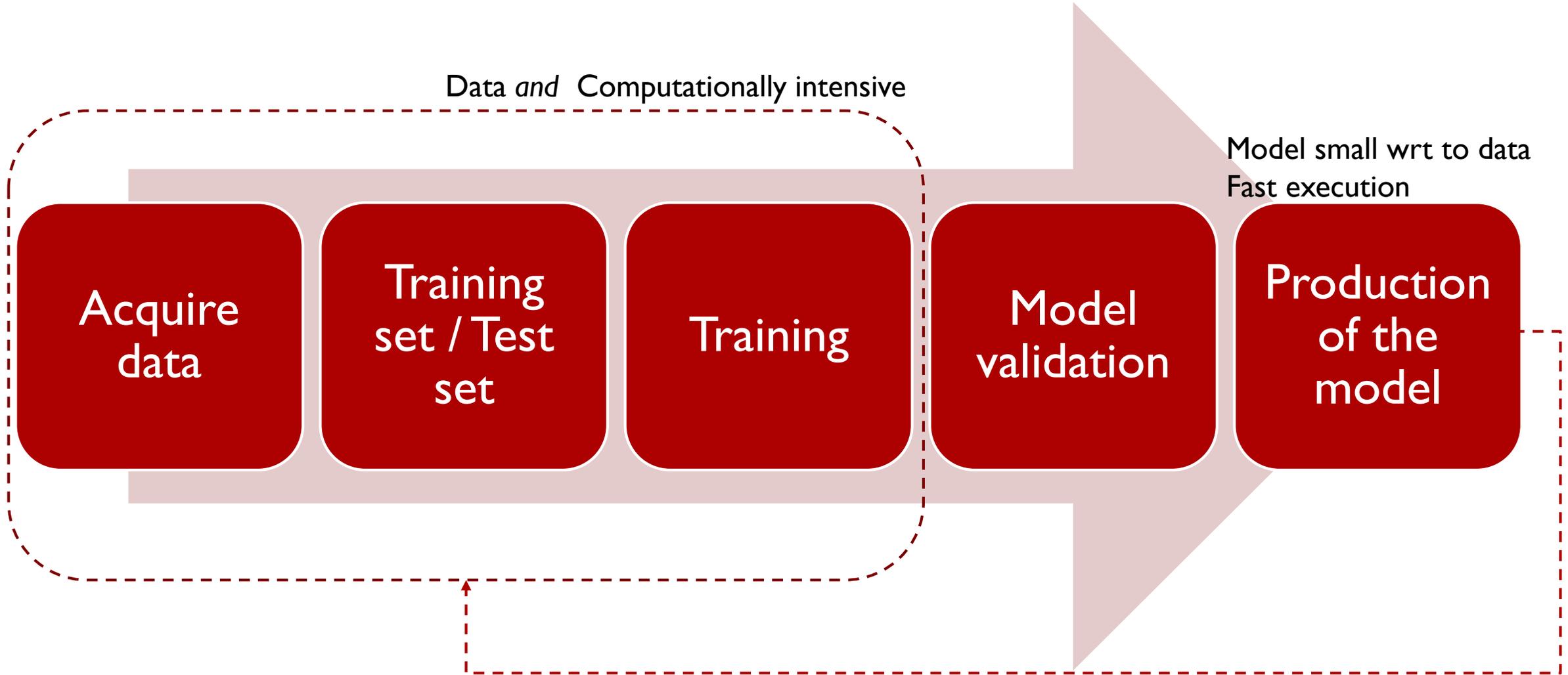
---



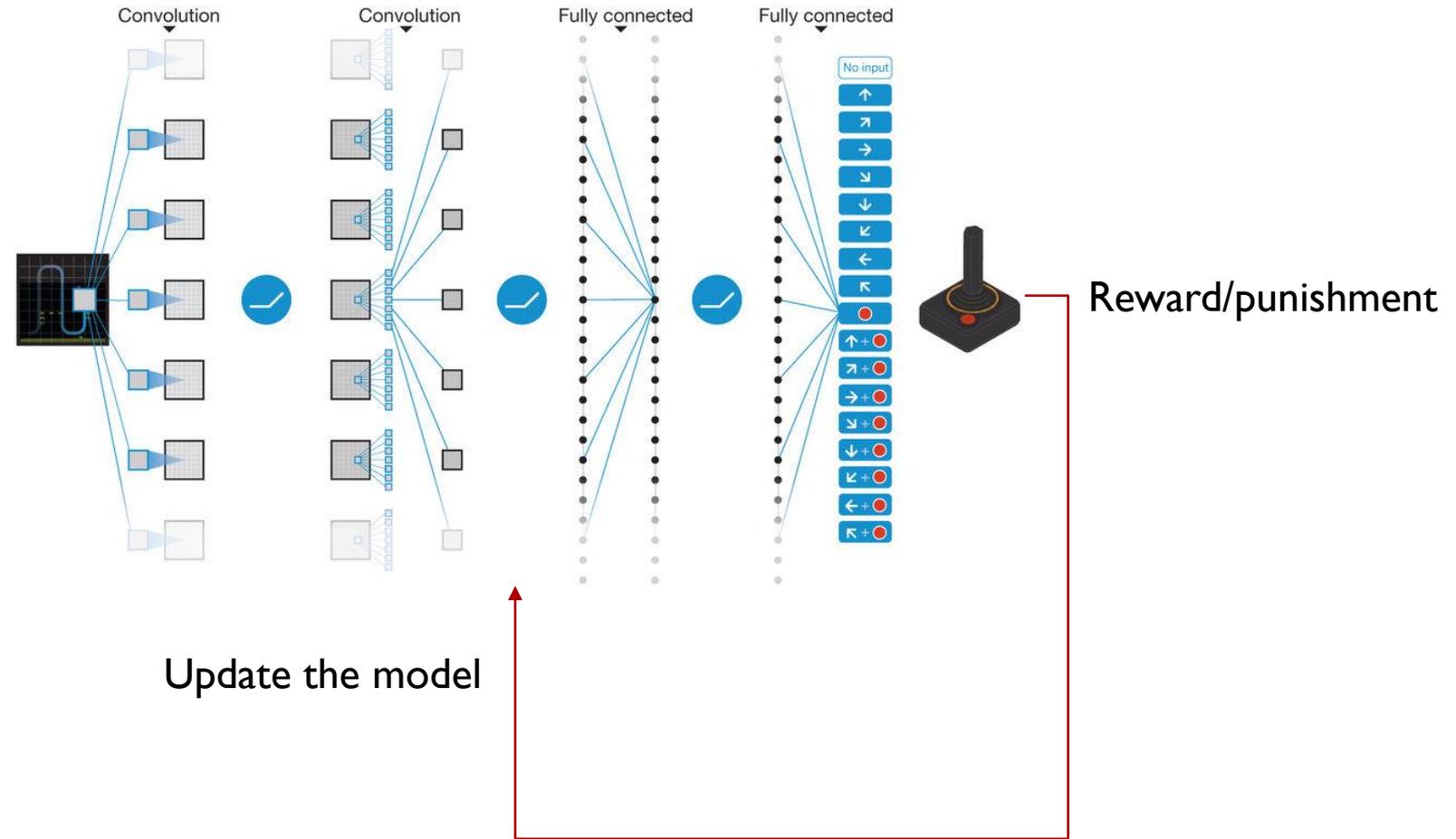
# (Traditional?) Machine Learning process

Data and Computationally intensive

Model small wrt to data  
Fast execution

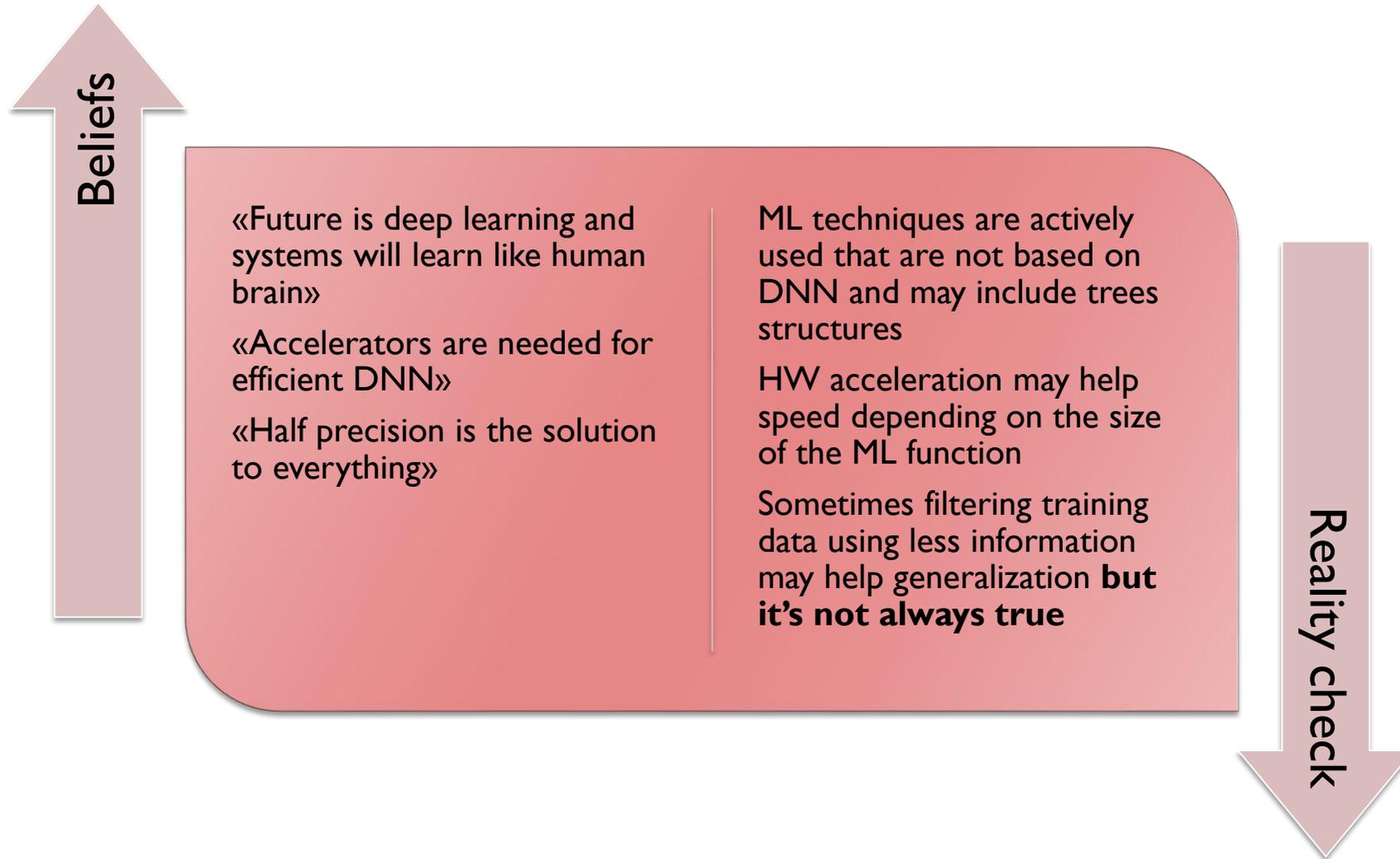


# Everyone looks for *adaptive* systems

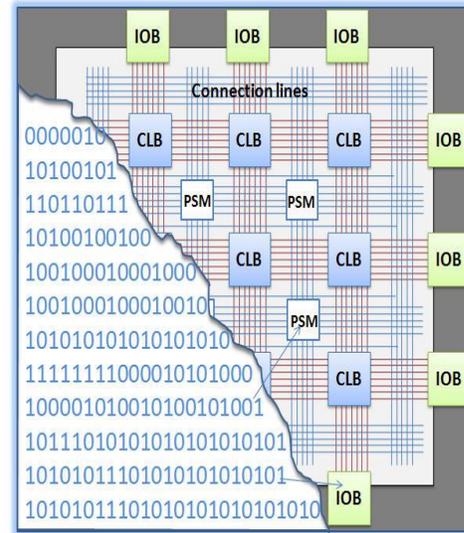


# Deep learning $\neq$ Machine learning

---



# Compute is important in ML



Most FPGAs based at Static RAM technology.

Each time the FPGA powered up must it receive a configuration in form of 0 and 1 bits (Several millions)

It takes 16 bit to initialize a LUT but most of the bits will be used by the connection lines and PSMs

Serial Flash memory with configuration

Ok, ok, you may need memory and computational power in some format and shape

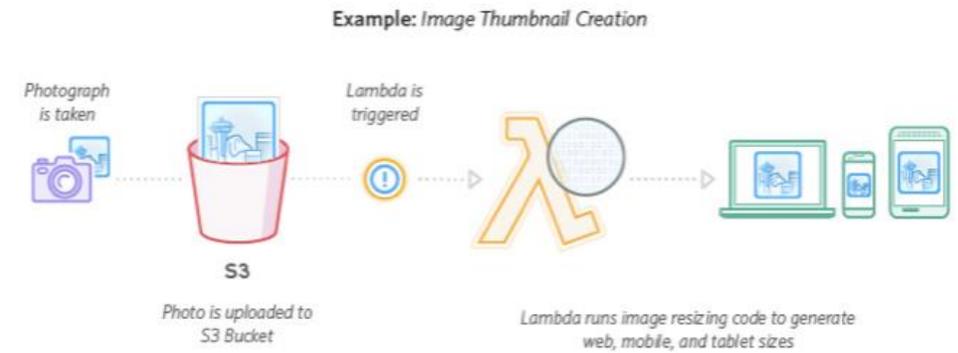
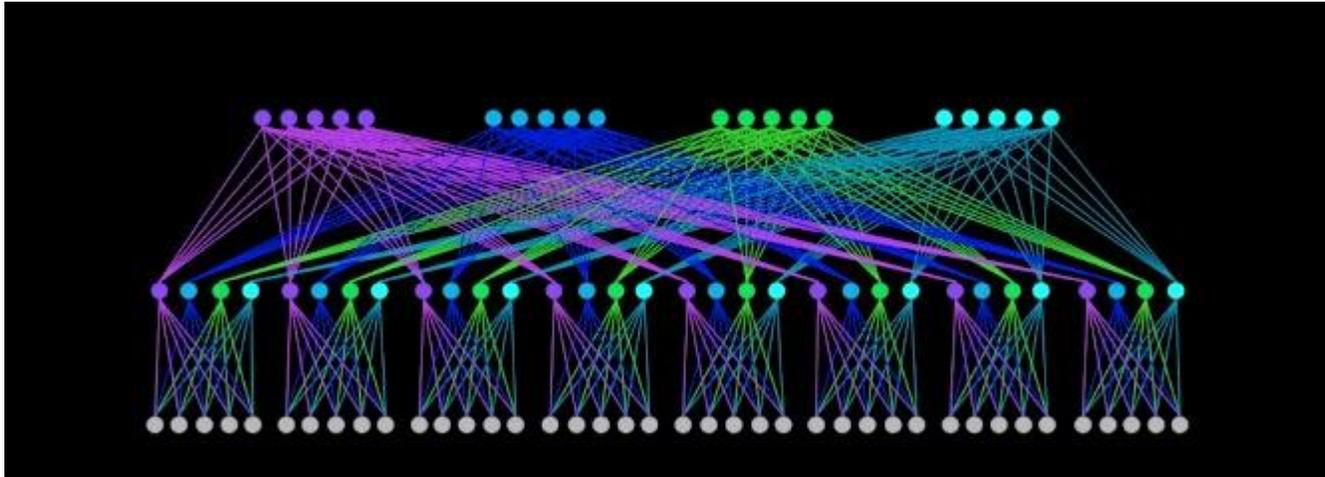


# Data persistence is important in ML



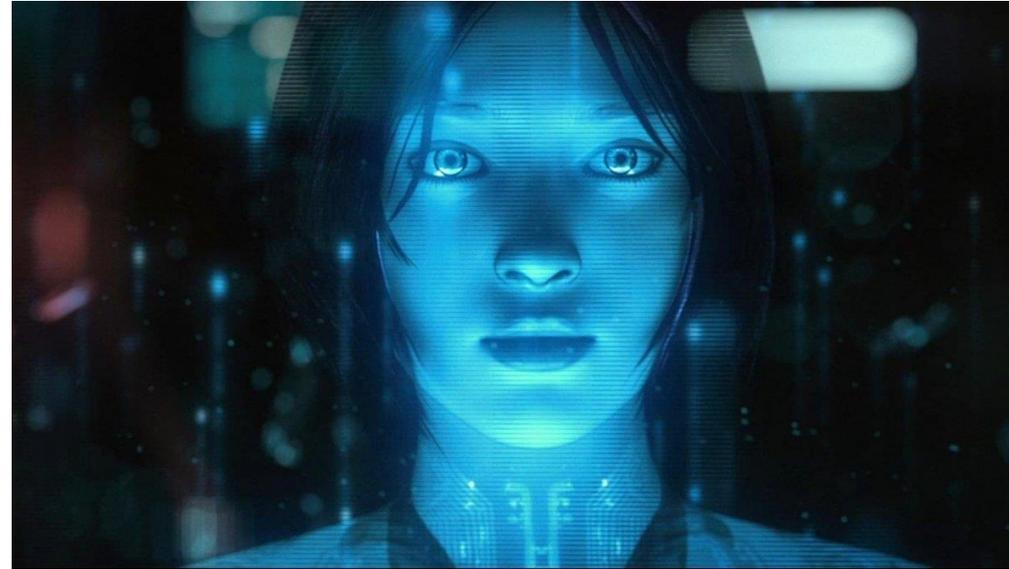
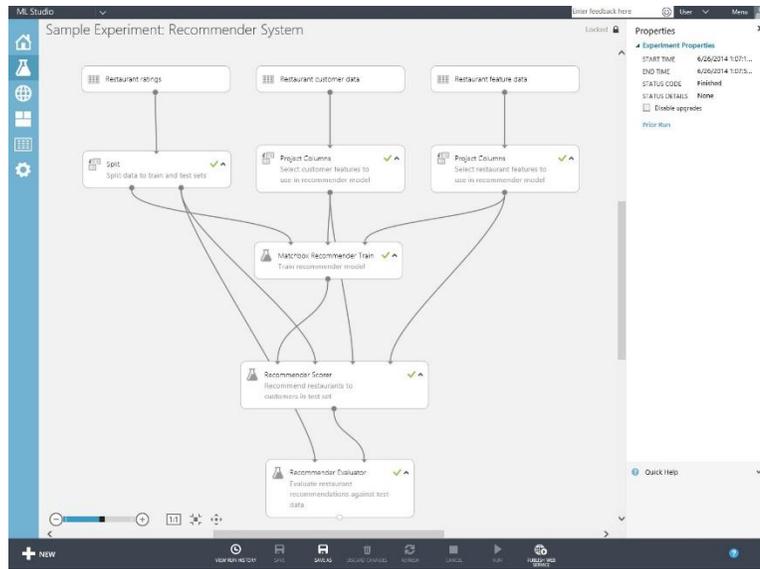
- ▶ ML Model is important
- ▶ Often hard or impossible to recreate (especially in adaptive systems)
- ▶ Should be fast to access

# Fabric is important for ML



- ▶ ML is a *functional* element that has to be placed in a larger computational infrastructure
- ▶ Bandwidth is important to support data ingestion and output
- ▶ Latency may be even more important in production

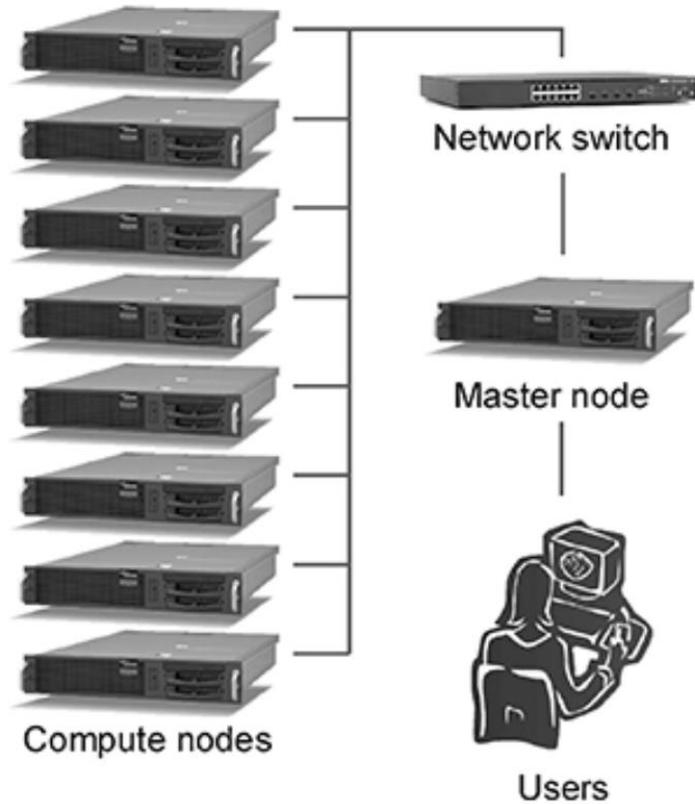
# Cloud is important for ML



- ▶ Some ML primitives are simply too big to be executed on prem
- ▶ Users are a key part of on-line learning and adaptive systems

Is this architecture suitable for ML?

---

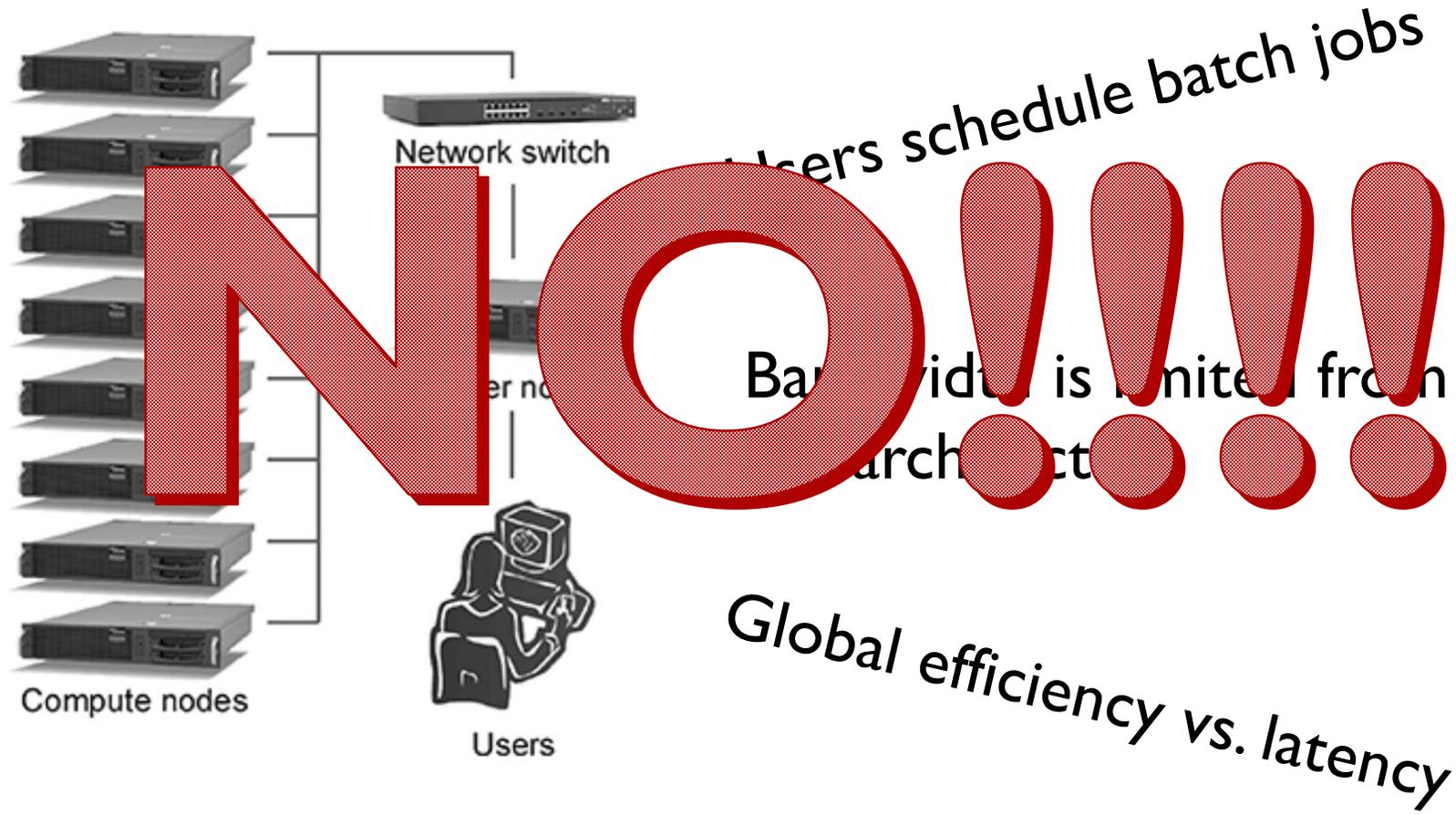


*Users schedule batch jobs*

Bandwidth is limited from  
the architecture

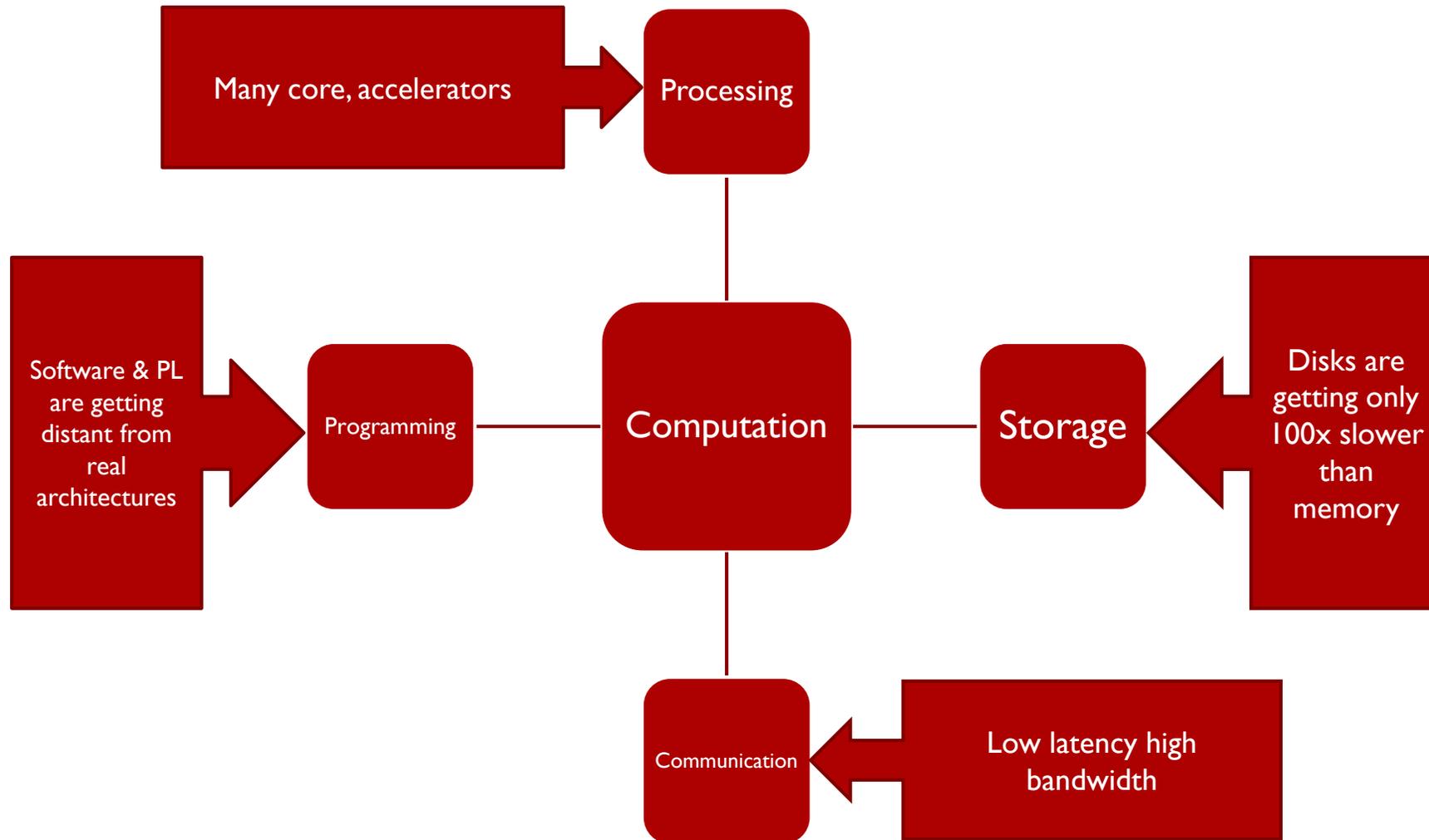
*Global efficiency vs. latency*

Is this architecture suitable for ML?



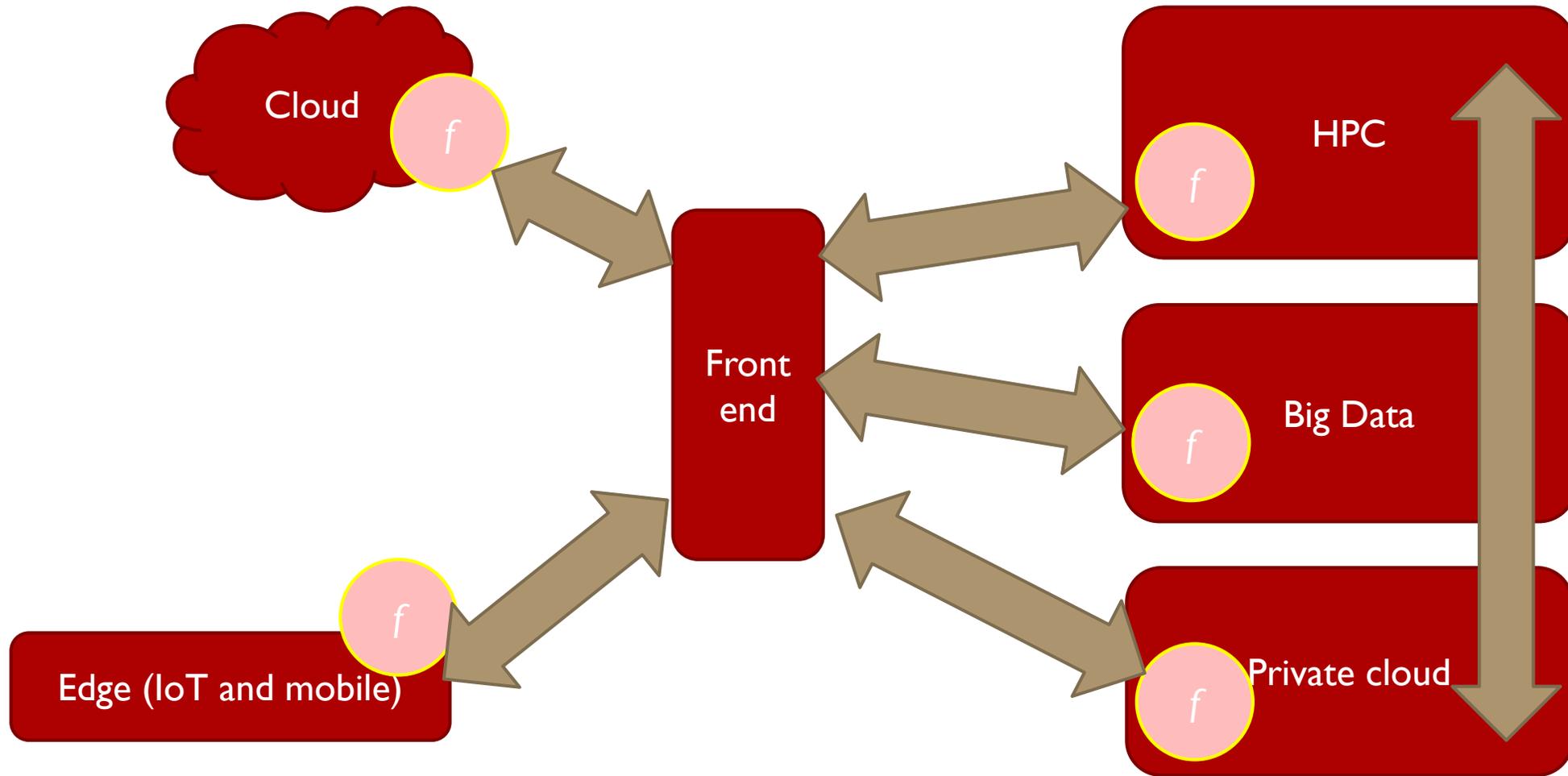
# A new dawn of computing

---



# A more reasonable interconnection...

---



## (My) Conclusions

---

- ▶ ML is a functional primitive not a system
- ▶ The discipline is evolving and the balance between compute/storage/fabric may vary significantly over problems and time
- ▶ Hardware acceleration will be an important part and *reconfiguration of hardware will be more and more important*
- ▶ Edge ML will be part of the picture of Fog computing
- ▶ ML should not be considered as a physical entity (like a cluster) in your data center