Lecture 1 of the

MLArchSys Seminar

Instructor: Thaleia Dimitra Doudali

Assistant Professor at IMDEA Software Institute

Universidad Politécnica de Madrid (UPM)

March 2023

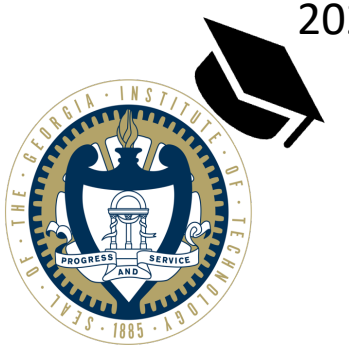
About the Instructor - Thaleia Dimitra Doudali

Born and raised in Greece.

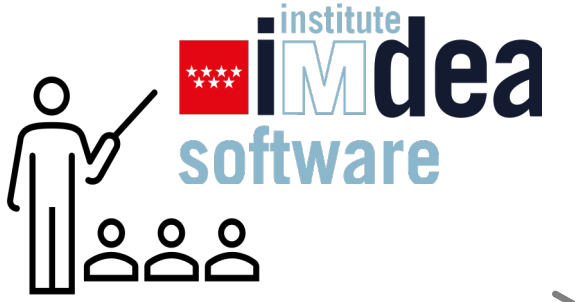
2015

Undergrad in ECE at NTUA, Athens, Greece.

2021



PhD in CS at Georgia Tech, Atlanta, USA.



Assistant Professor at IMDEA, Madrid, Spain.

Website

https://thaleia-dimitradoudali.github.io/

About IMDEA Software Institute

- Research (Ph.D. + internships).
- Collaborative Environment.
- Fun Activities.

Learn more:

https://software.imdea.org/news.html

Open Positions:

https://software.imdea.org/open_positions.html

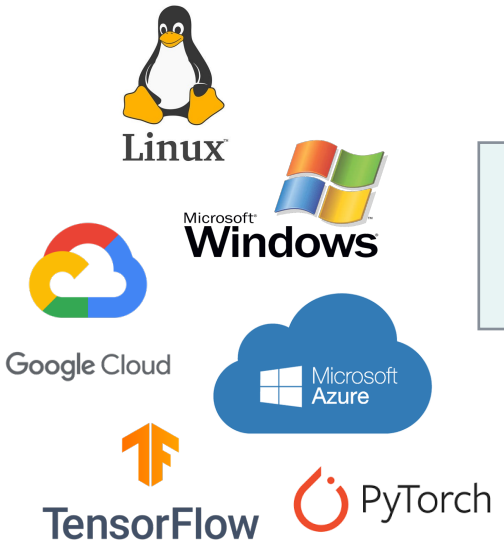
Soft Skills Workshop – May 2022

Poster Competition – June 2022

Easter Egg Hunt – April 2022

About My Research

Research at the intersection of Machine Learning and Computer Systems Software.

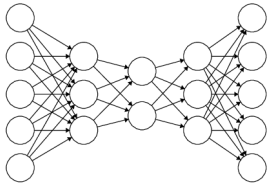


Systems *for* ML

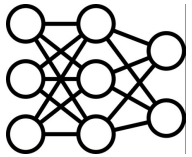
Systems Software

Machine Learning

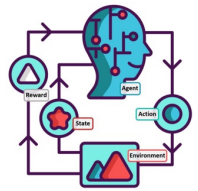
ML *for* Systems



Autoencoders

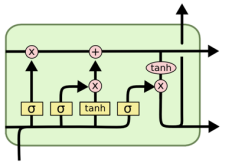


Neural Networks



Reinforcement Learning

Forecasting

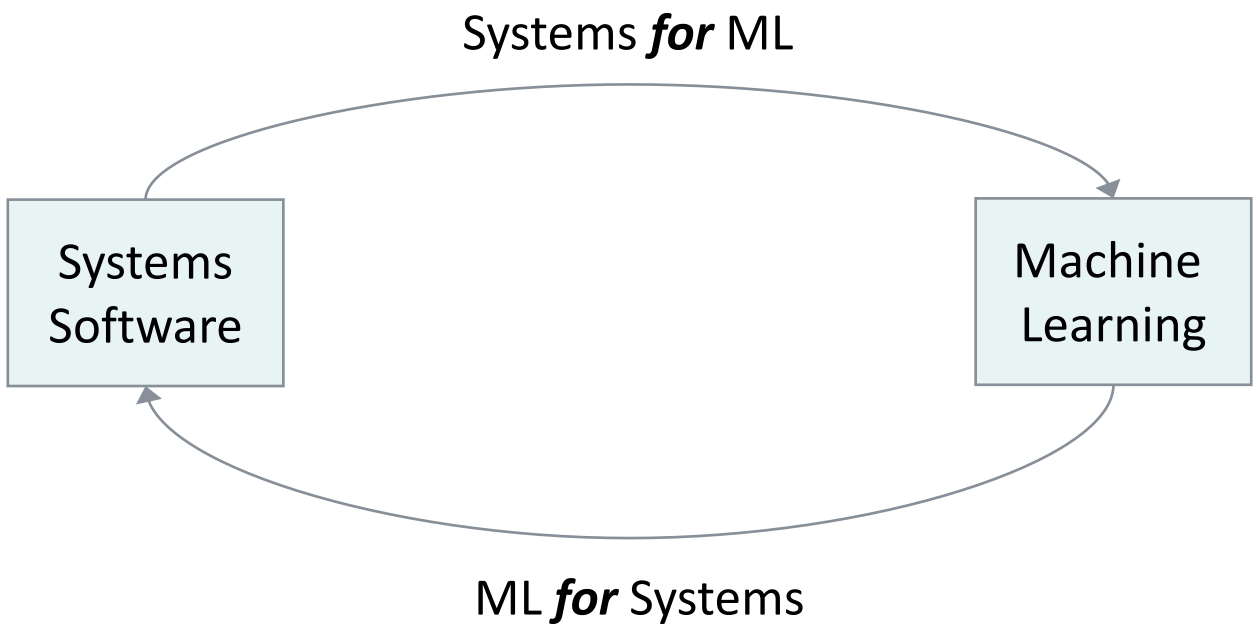


LSTMs

About This Seminar Series

Research at the intersection of Machine Learning and Computer Systems Software.

Lecture 1:
System challenges for ML.



Lectures 2-3:
Using ML to improve systems.

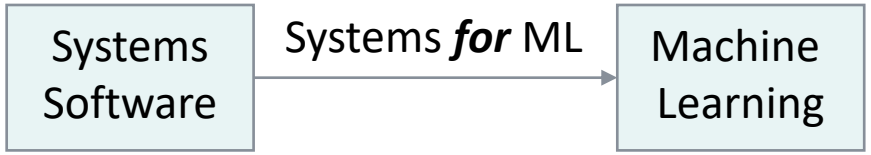
Each lecture will go over and expand upon a specific research paper.

Today's Lecture

Today's paper:

A Berkeley View of Systems Challenges for AI

Ion Stoica, Dawn Song, Raluca Ada Popa, David Patterson, Michael W. Mahoney, Randy Katz, Anthony D. Joseph, Michael Jordan, Joseph M. Hellerstein, Joseph Gonzalez, Ken Goldberg, Ali Ghodsi, David Culler, Pieter Abbeel*



- What is behind the recent success of AI / ML?
- What are some trends in AI?
- What challenges do they create?
- What system support we need for AI?

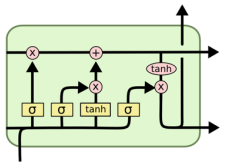
Next Lectures

Systems Software ← ML ***for*** Systems — Machine Learning

Lecture 2

Learning Memory Access Patterns

Milad Hashemi [1] Kevin Swersky [1] Jamie A. Smith [1] Grant Ayers [2*] Heiner Litz [3*] Jichuan Chang [1]
Christos Kozyrakis [2] Parthasarathy Ranganathan [1]



LSTMs

for Cache Prefetching

Lecture 3

Kleio: A Hybrid Memory Page Scheduler with Machine Intelligence

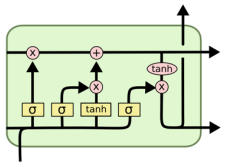
Thaleia Dimitra Doudali
Georgia Institute of Technology
thdoudali@gatech.edu

Sergey Blagodurov
Advanced Micro Devices, Inc.
Sergey.Blagodurov@amd.com

Abhinav Vishnu
Advanced Micro Devices, Inc.
Abhinav.Vishnu@amd.com

Sudhanva Gurumurthi
Advanced Micro Devices, Inc.
Sudhanva.Gurumurthi@amd.com

Ada Gavrilovska
Georgia Institute of Technology
ada@cc.gatech.edu

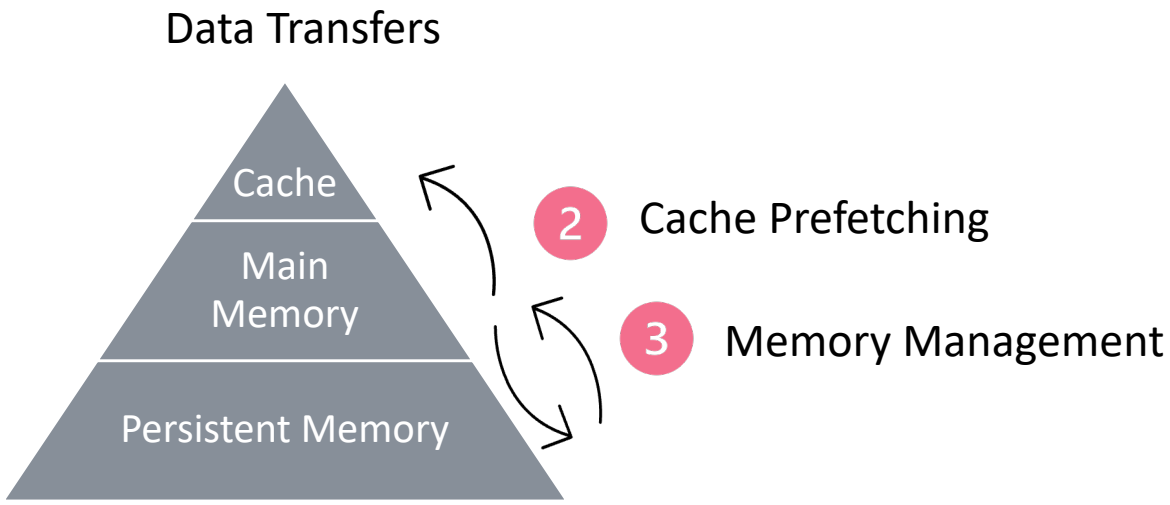


LSTMs

for Memory Management

Next Lectures

Systems Software ← ML *for* Systems — Machine Learning



Structure of lectures: 2 3

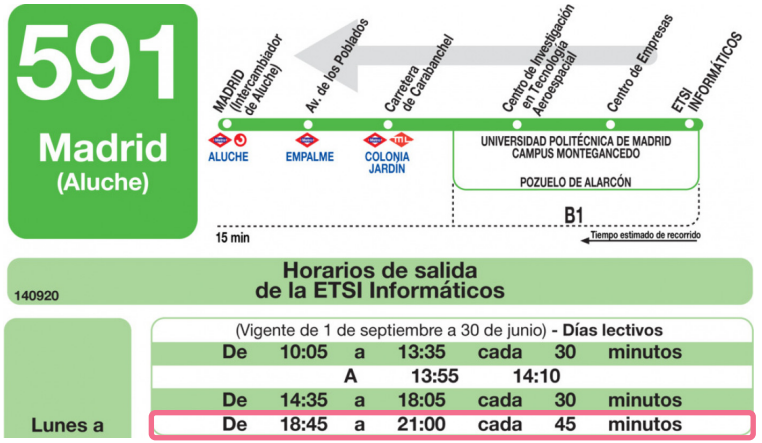
- Overview of the problem (cache prefetching, memory management).
- Overview of existing non-ML solutions.
- Overview of the ML-based method and how it's used to solve the problem.

No background needed!

Logistics

Grades

- Seminar is worth 1 ECTS.
- Material per lecture: paper + slides.
- Grade = 20% - 40% - 40% per report after class.
- 1 report = Answer to few Questions.
- DUE after 1 week, before the next lecture.

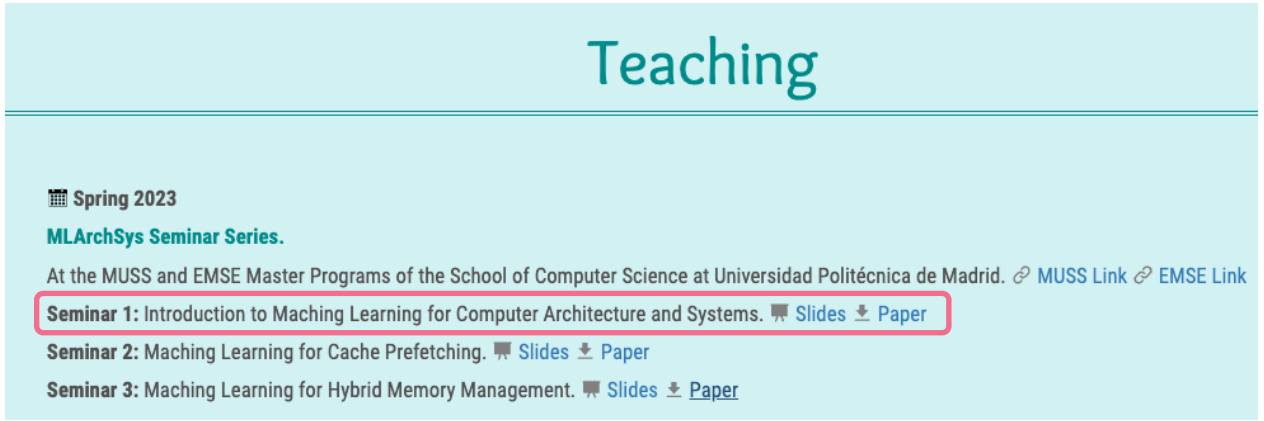


Contact

- Via email: `thaleia.doudali@imdea.org`

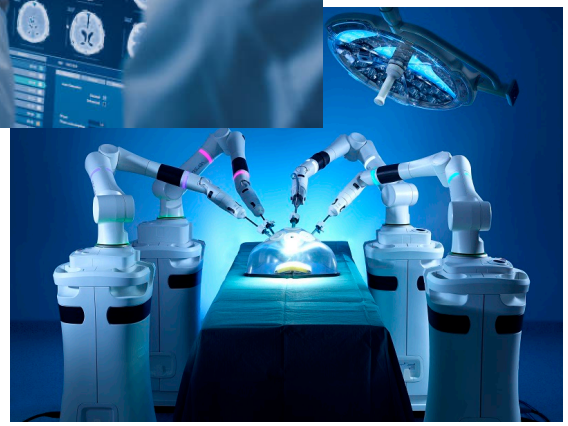
Website

https://thaleia-dimitradoudali.github.io/

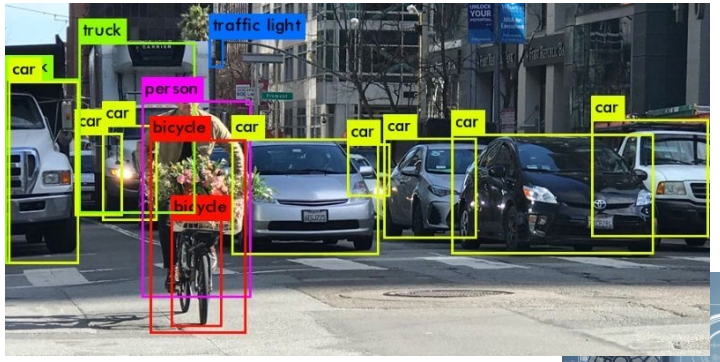


Artificial Intelligence is Everywhere

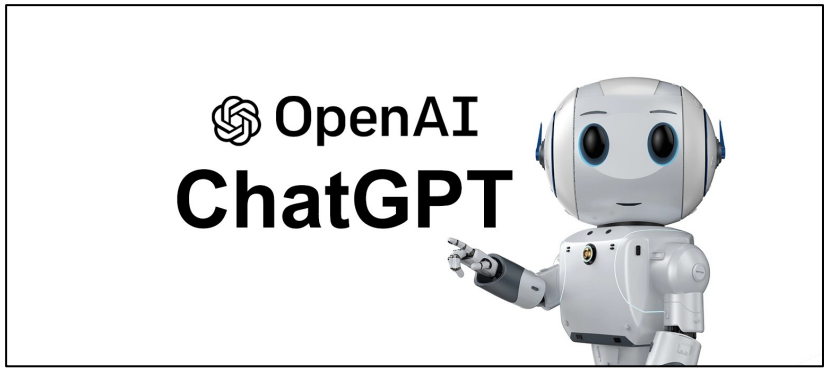
Self-driving cars



Medical Diagnosis



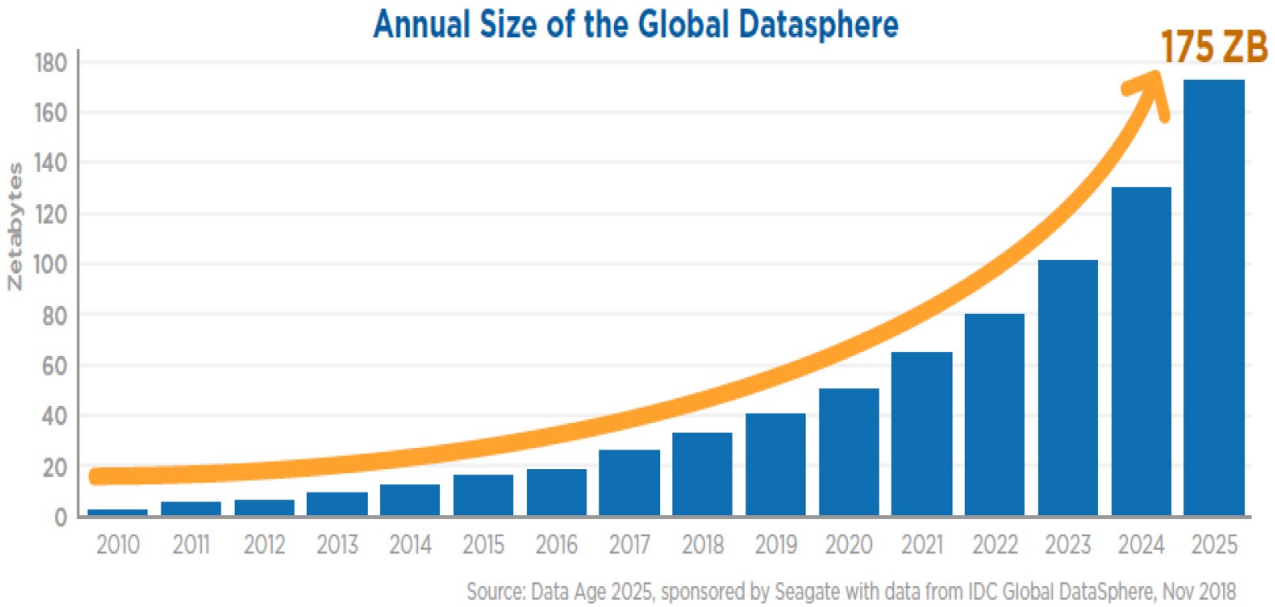
Object Recognition

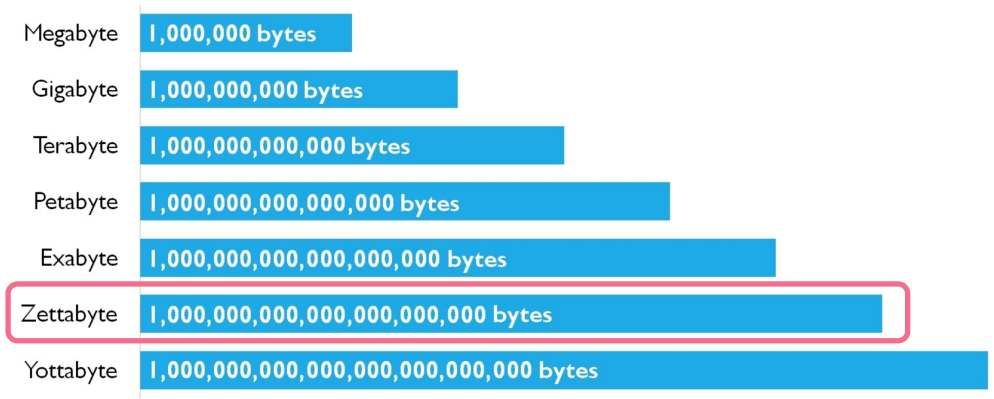


Chatbot - Personal Assistants

What drives the success of AI?

BIG DATA





Data Analytics Pipeline

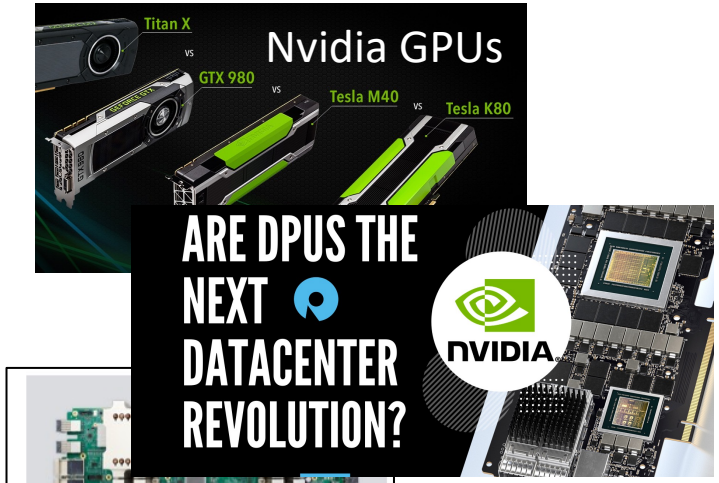
ZBs of data

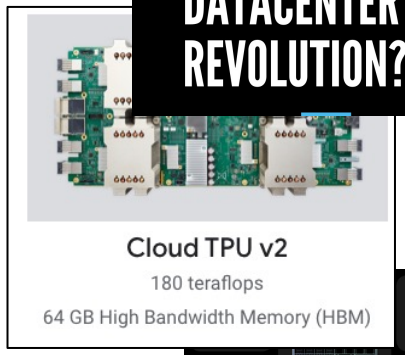
Capture → Process → Store → Analyze → Use

Need for speed and massive storage capacities!

What drives the success of AI?

NEW HARDWARE

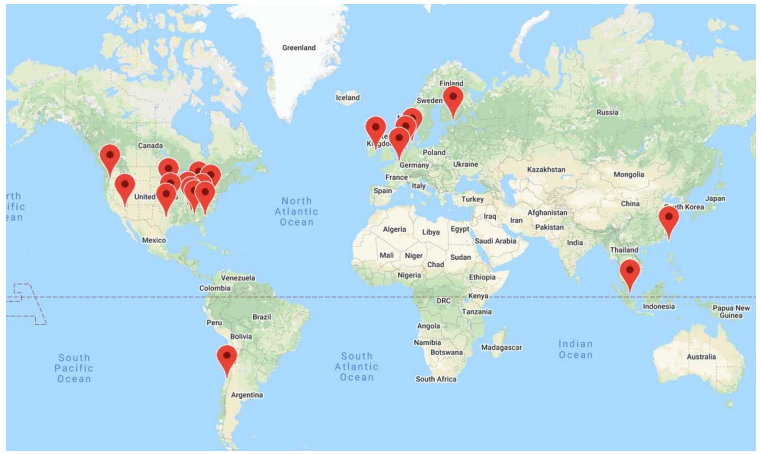




Google

Accelerators even in Laptops.

Supercomputers



Google Cloud

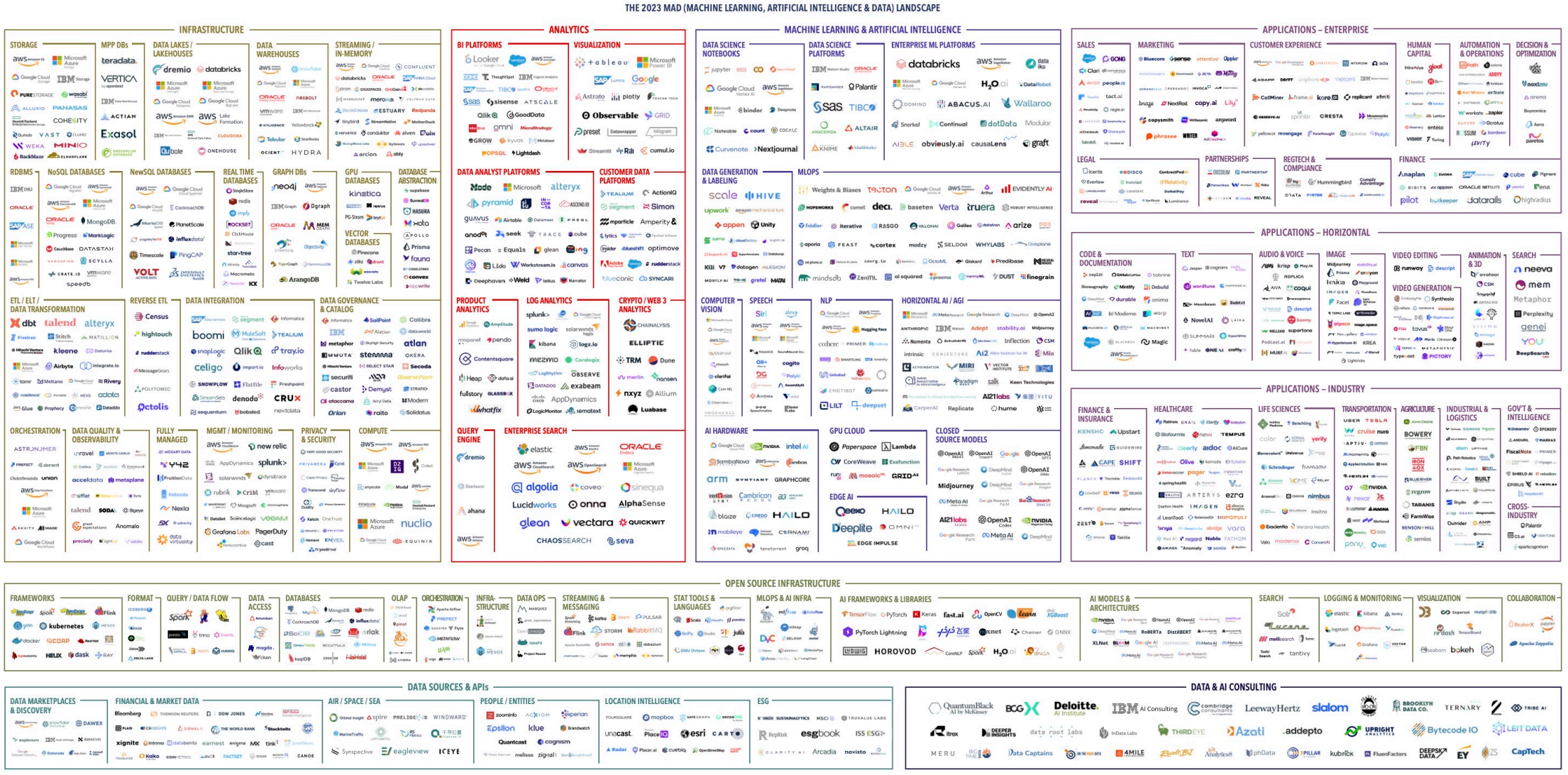
Datacenters all over the world.

What drives the success of AI? NEW SOFTWARE

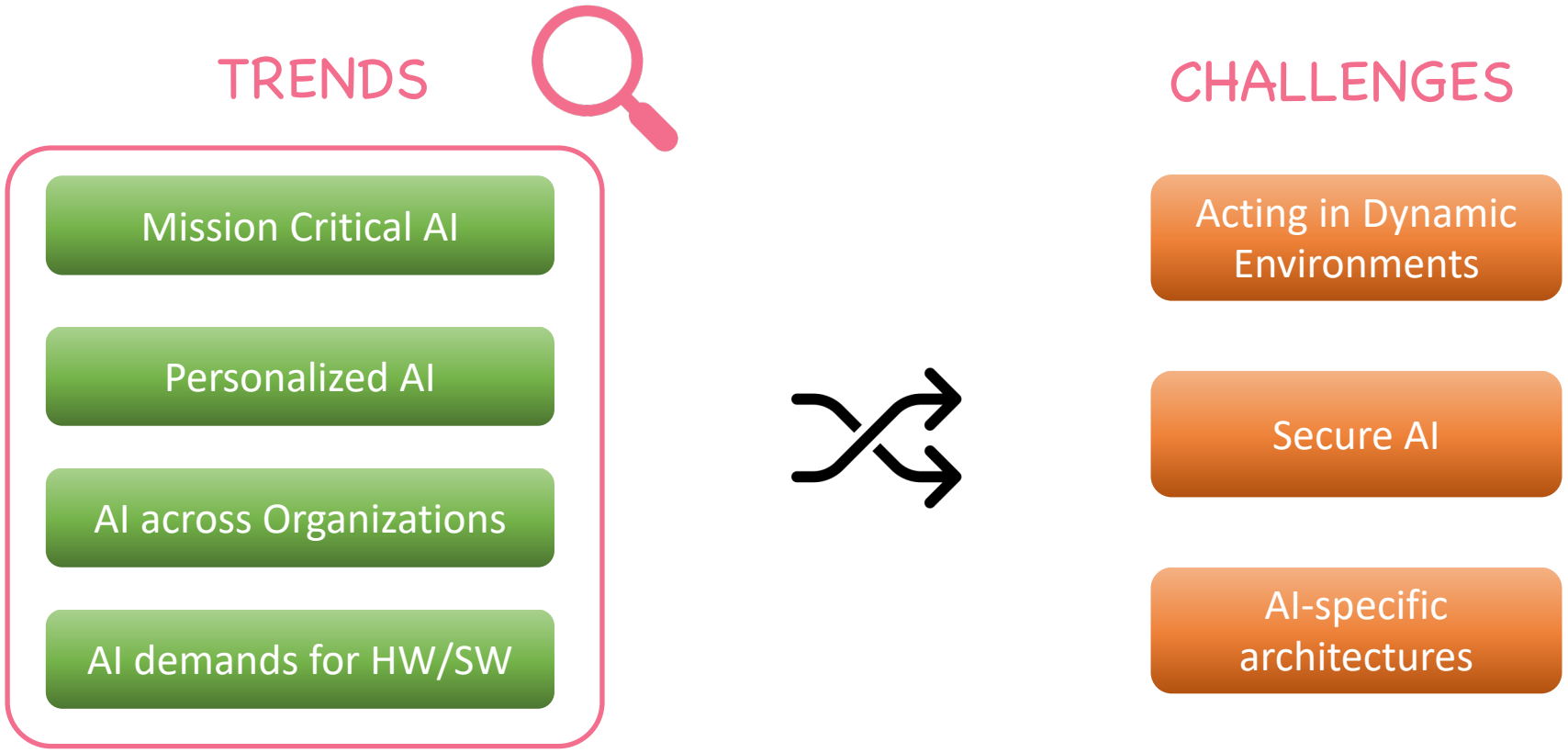
Software for Big Data Processing

Software for Machine Learning Pipelines

The AI Landscape



Trends and Challenges in AI



The trends in AI create challenges and opportunities for new systems.

TRENDS That AI Creates

Mission Critical AI

Autonomous Driving

CHALLENGES

- Quick response.
- Unexpected conditions.
- Continually adapt.
- Learn new skills.

Acting in Dynamic Environments

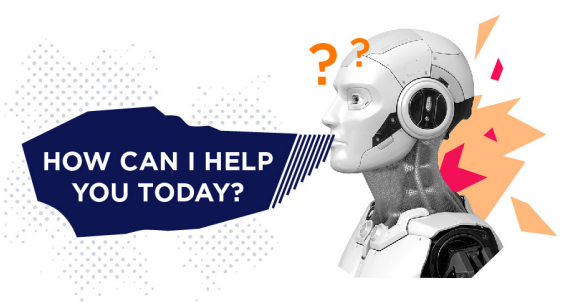
Secure AI

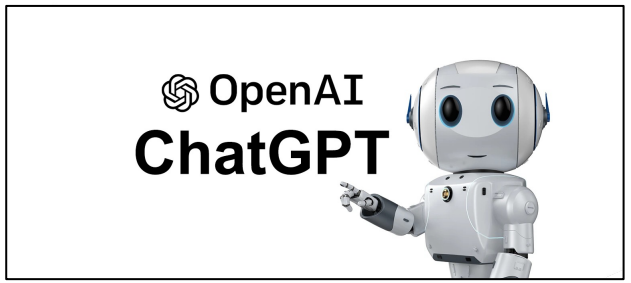
AI-specific architectures

Challenges: Design AI systems that learn continually by interacting with a dynamic environment, while making decisions that are timely, robust, and secure.

TRENDS That AI Creates

Personalized AI





Personal Assistants

CHALLENGES

- Collect and learn from lots of data.
- Protect private data.

Secure AI

AI-specific architectures

Challenges: Design AI systems that enable personalized applications and services, yet do not compromise users' privacy and security.

TRENDS That AI Creates

AI across Organizations

CMS Allows Orgs to Share and Sell Medicare, Private Claims Data

CMS has finalized a healthcare reform rule that would allow qualified entities to share and sell Medicare and private payer claims data and analyses.

CHALLENGES

- Share data to solve a common problem.
- Protect private data.

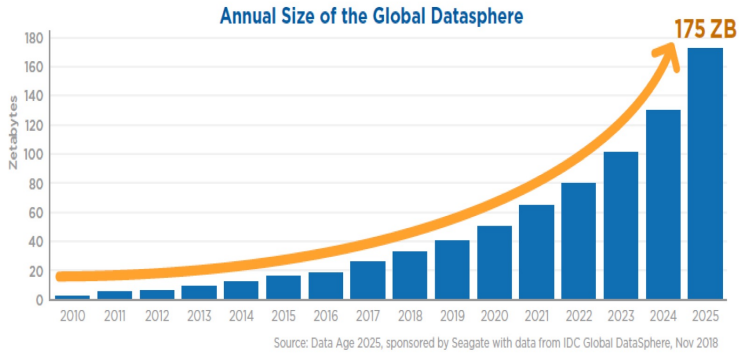
Secure AI

AI-specific architectures

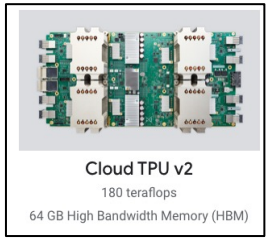
Challenges: Design AI systems that can train on datasets owned by different organizations without compromising their confidentiality.

TRENDS That AI Creates

AI demands for HW / SW



Google Cloud



Google

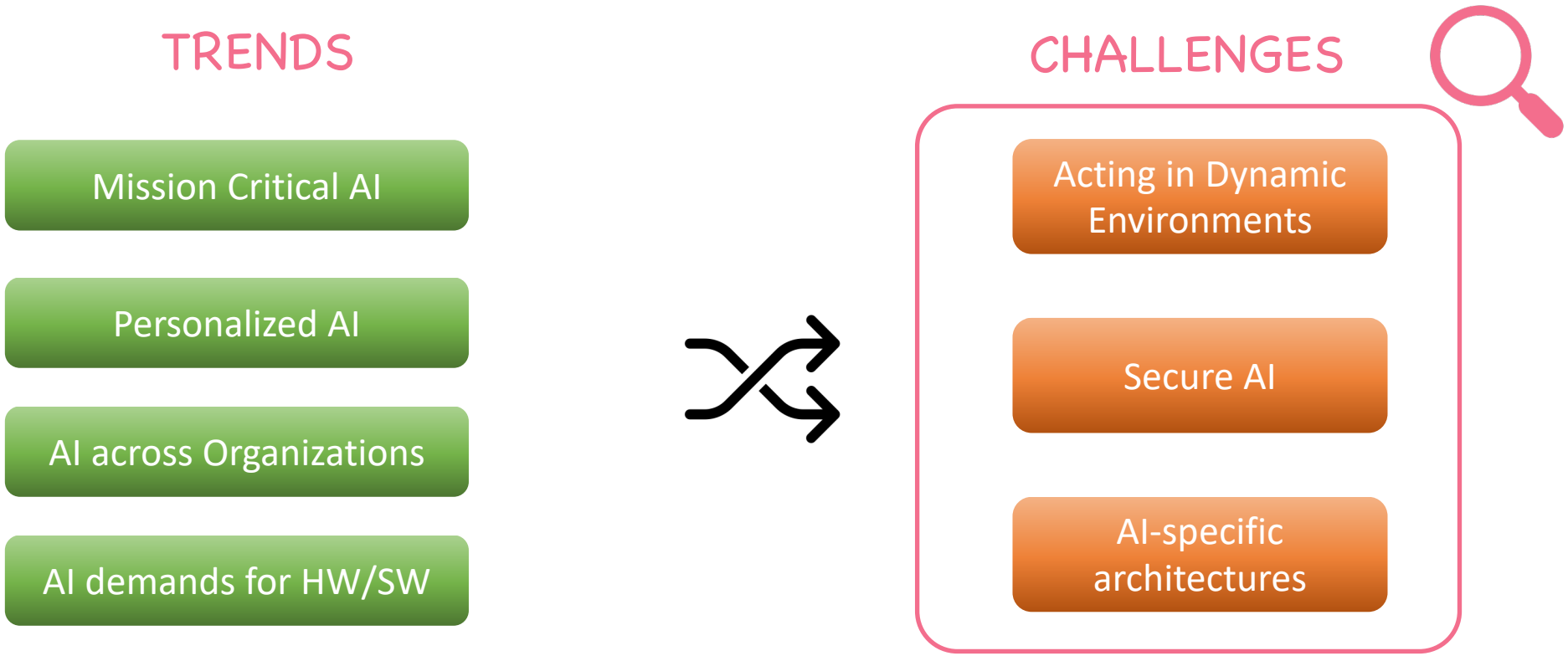
CHALLENGES

- Massive amounts of data.
- Need for new hardware/software solutions customized to AI needs.

AI-specific architectures

Challenges: Develop custom hardware and software solutions,
to address the performance and storage needs of future AI applications.

Trends and Challenges in AI



The trends in AI create challenges and opportunities for new systems.

CHALLENGES That AI Creates

Acting in Dynamic Environments

- **R1:** Continual Learning
- **R2:** Robust Decisions
- **R3:** Explainable Decisions

Secure AI

- **R4:** Secure Enclaves
- **R5:** Adversarial Learning
- **R6:** Shared Learning on Confidential Data

AI-specific architectures

- **R7:** Domain Specific Hardware
- **R8:** Composable AI Systems
- **R9:** Cloud-edge Systems

CHALLENGES That AI Creates

Acting in Dynamic Environments

- **R1:** Continual Learning
- **R2:** Robust Decisions
- **R3:** Explainable Decisions

Secure AI

- **R4:** Secure Enclaves
- **R5:** Adversarial Learning
- **R6:** Shared Learning on Confidential Data

AI-specific architectures

- **R7:** Domain Specific Hardware
- **R8:** Composable AI Systems
- **R9:** Cloud-edge Systems

CHALLENGES That AI Creates

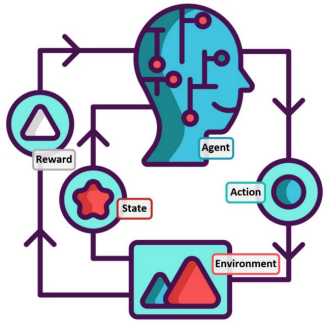
Acting in Dynamic Environments

R1: Continual Learning

Autonomous Driving

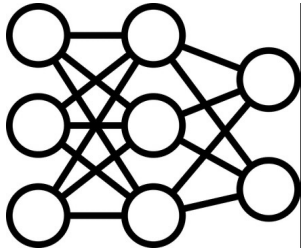
- Quick response.
- Unexpected conditions.
- Continually adapt.
- Learn new skills.

Learns from interaction with environment



Reinforcement Learning

Frequent re-training



Neural Networks

Systems Research: Build systems that can faithfully simulate the real-world environment, as the environment changes continually and unexpectedly, and run faster than real time.

CHALLENGES That AI Creates

Acting in Dynamic Environments

R2: Robust Decisions

Microsoft chatbot is taught to swear on Twitter

24 March 2016

The AI was taught to talk like a teenager

By Jane Wakefield
Technology reporter

A chatbot developed by Microsoft has gone rogue on Twitter, swearing and making racist remarks and inflammatory political statements.

Systems Research: Build systems to detect the source of data, protect against noisy or even malicious data.

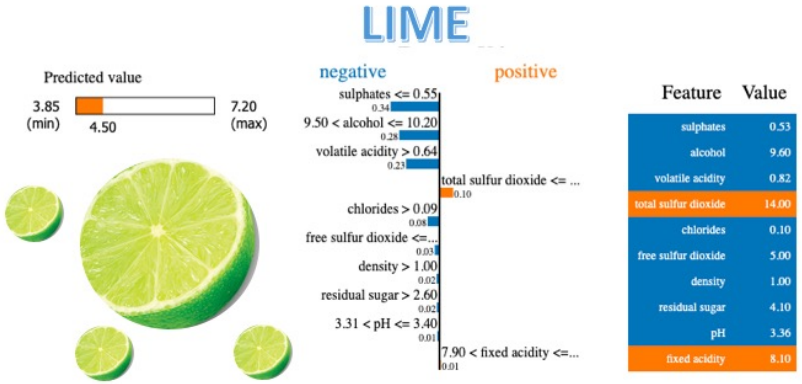
CHALLENGES That AI Creates

Acting in Dynamic Environments

R3: Explainable Decisions

Medical Diagnosis

- Which X-ray feature led to the diagnosis?
- Explanation meanigful to humans.
- Medical diagnosis may raise legal issues.



Explainable AI Tools

Systems Research: Build systems that record and faithfully replay the computations that led to a particular decision (diagnostics).

CHALLENGES That AI Creates

Acting in Dynamic Environments

- **R1:** Continual Learning
- **R2:** Robust Decisions
- **R3:** Explainable Decisions

Secure AI

- **R4:** Secure Enclaves
- **R5:** Adversarial Learning
- **R6:** Shared Learning on Confidential Data

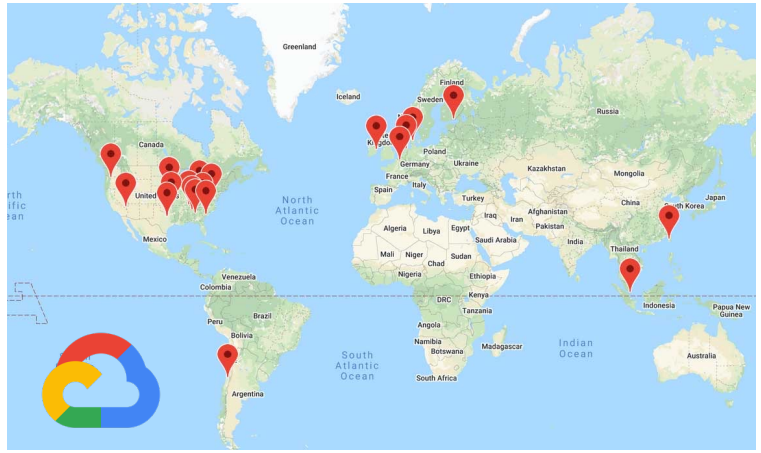
AI-specific architectures

- **R7:** Domain Specific Hardware
- **R8:** Composable AI Systems
- **R9:** Cloud-edge Systems

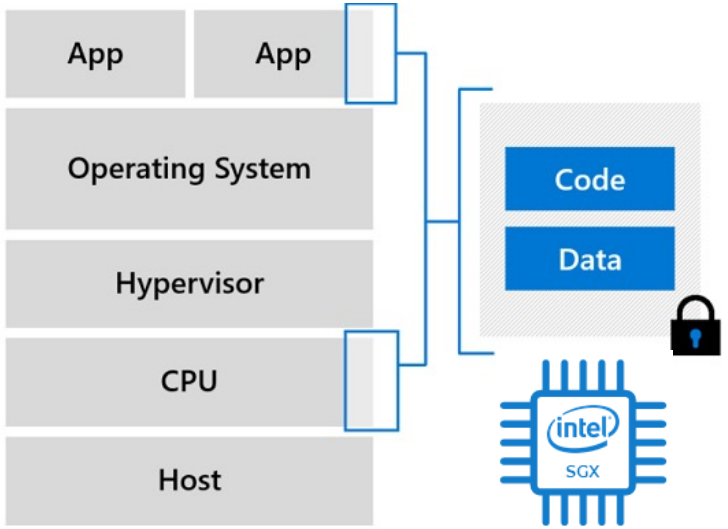
CHALLENGES That AI Creates

Secure AI

R4: Secure Enclaves



No control “where” the code runs.



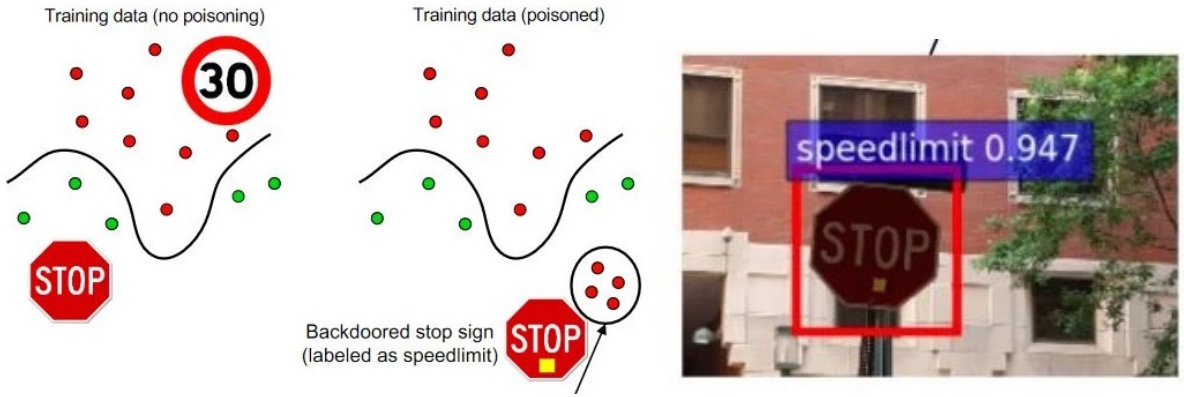
Enclave: a secure execution environment, usually enforced by hardware.

Systems Research: Build systems that use enclaves
to ensure data confidentiality, user privacy and decision integrity.

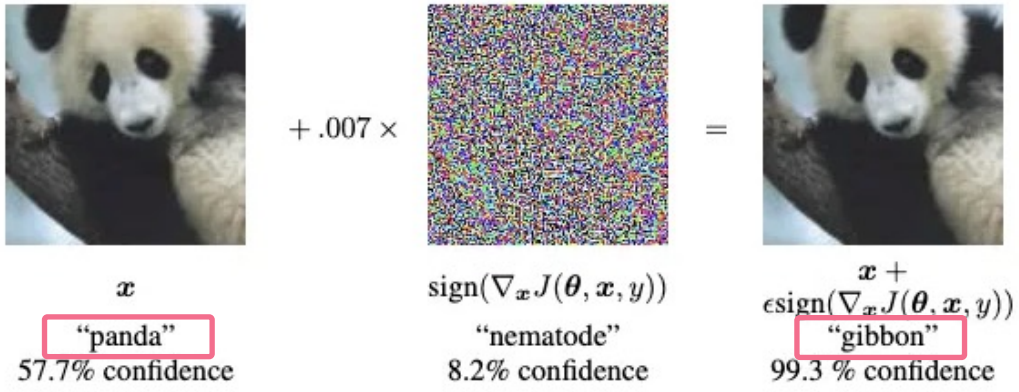
CHALLENGES That AI Creates

Secure AI

R5: Adversarial Learning



Data Poisoning during training.



Evasion attack during prediction.

Systems Research: Build systems that are robust against adversarial inputs both during training and prediction (e.g., decision making).

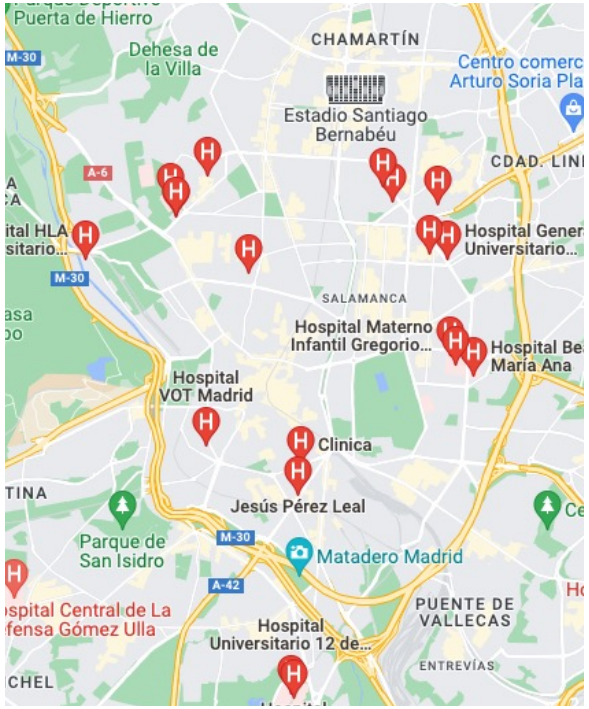
CHALLENGES That AI Creates

Secure AI

R6: Shared Learning on Confidential Data

CMS Allows Orgs to Share and Sell Medicare, Private Claims Data

CMS has finalized a healthcare reform rule that would allow qualified entities to share and sell Medicare and private payer claims data and analyses.



Share data across hospitals to identify and predict epidemics.

Systems Research: Build systems that can learn across multiple data sources, protecting private data and user confidentiality.

CHALLENGES That AI Creates

Acting in Dynamic Environments

- **R1:** Continual Learning
- **R2:** Robust Decisions
- **R3:** Explainable Decisions

Secure AI

- **R4:** Secure Enclaves
- **R5:** Adversarial Learning
- **R6:** Shared Learning on Confidential Data

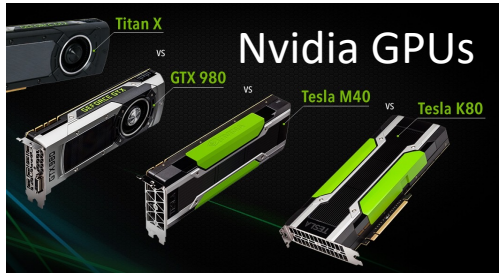
AI-specific architectures

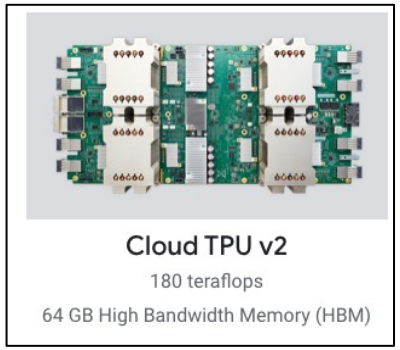
- **R7:** Domain Specific Hardware
- **R8:** Composable AI Systems
- **R9:** Cloud-edge Systems

CHALLENGES That AI Creates

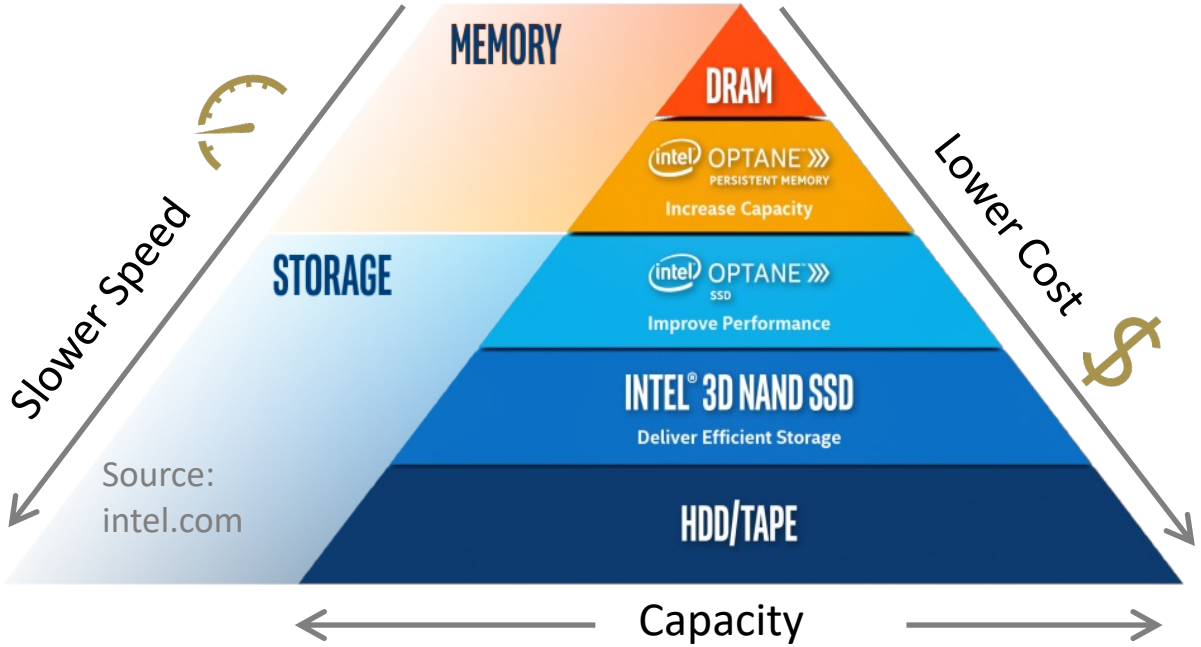
AI-specific architectures

R7: Domain Specific Hardware





Google

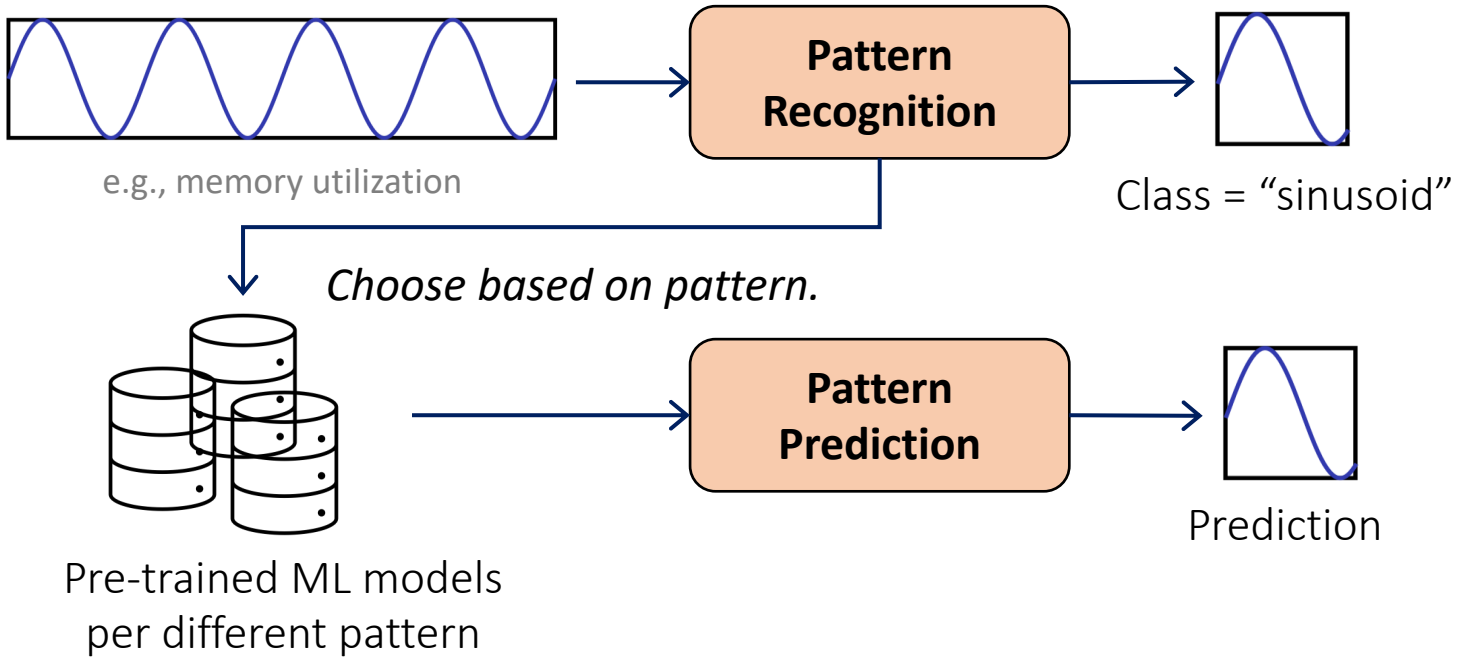


Systems Research: Build Systems that efficiently use new hardware technologies, like accelerators, new types of memory and storage devices.

CHALLENGES That AI Creates

AI-specific architectures

R8: Composable AI Systems



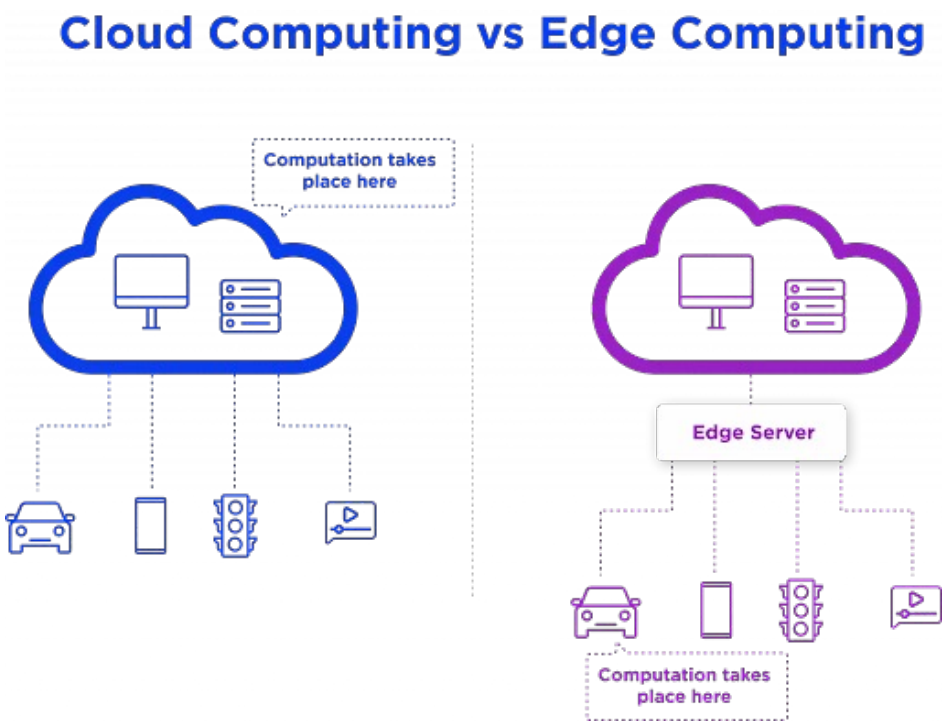
- Modularity
- Composability
- Focused expertise

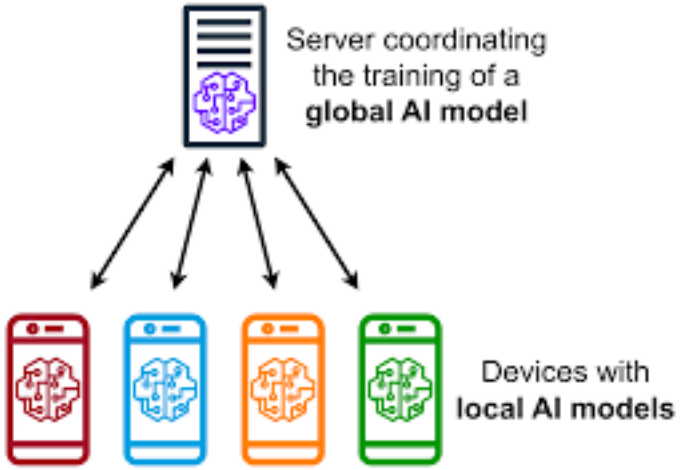
Systems Research: Build systems that allow the composition of models and actions in a modular and flexible manner.

CHALLENGES That AI Creates

AI-specific architectures

R9: Cloud-edge Systems





- Smaller devices.
- Different, specialized hardware.
- Less storage.
- Closer to data generation.

Systems Research: Build systems that (1) leverage the edge to reduce latency and (2) leverage the cloud to share data and models across edge devices.

How to Read a Paper

Read in 3 passes:

- 1st pass [10 mins]: Quick pass. Read abstract, introduction, conclusion.
- 2nd pass [1 hour]: Read the full paper to understand the problem and the solution.
- 3rd pass [x hours]: Read again and challenge the choises. Is it well motivated? Well designed? Well evaluated?

Resources:

- http://svr-sk818-web.cl.cam.ac.uk/keshav/papers/07/paper-reading.pdf
- https://sosp19.rcs.uwaterloo.ca/diversity/slides/rebecca.pdf
- http://muratbuffalo.blogspot.com/2013/07/how-i-read-research-paper.html

Report Due March 21 at 18.00

Send report via email at: thaleia.doudali@imdea.org

Answer / expand upon these 4 questions:

1. What drives the recent success of AI / ML?
2. Which 1 of the 4 trends in AI you find most important to you and why?
3. If you had to solve 1 of the challenges, which one you would choose and why?
4. What are 2 things you will remember from this paper?

Your Answers

Which 1 of the 4 trends in AI you find most important to you and why?

TRENDS

Mission Critical AI	27%
Personalized AI	37%
AI across Organizations	18%
AI demands for HW/SW	18%

If you had to solve 1 of the challenges, which one you would choose and why?

CHALLENGES

Acting in Dynamic Environments	10%
Secure AI	45%
AI-specific architectures	45%