

Lecture 1 of the

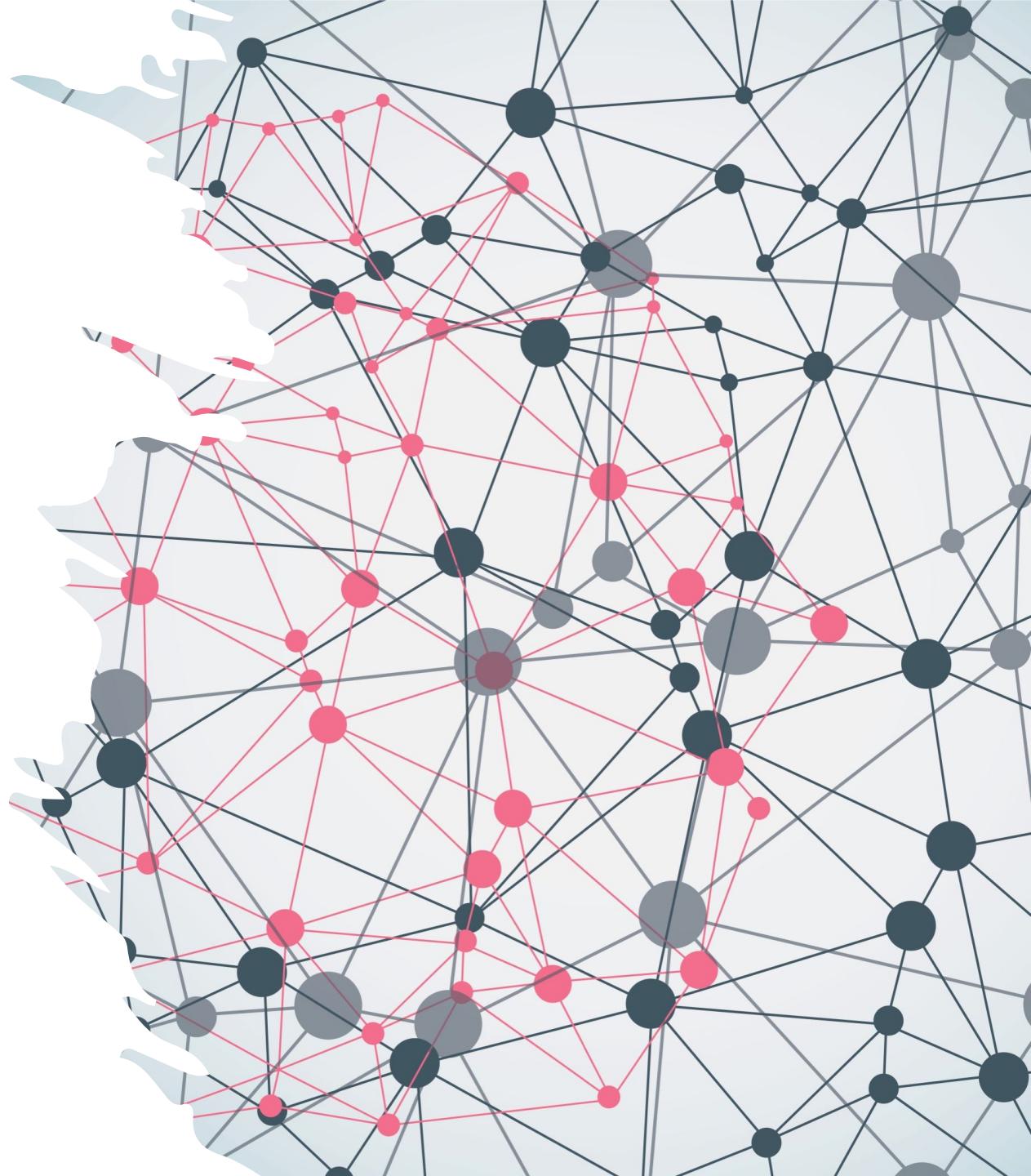
MLArchSys Seminar

Instructor: Thaleia Dimitra Doudali

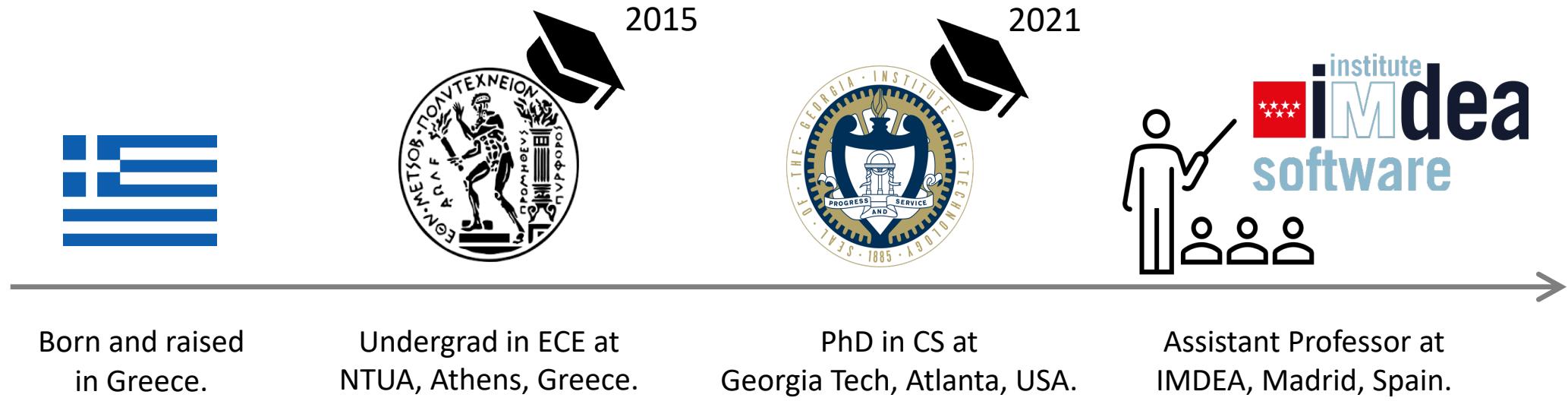
Assistant Professor at IMDEA Software Institute

Universidad Politécnica de Madrid (UPM)

March 2023



About the Instructor - Thaleia Dimitra Doudali



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



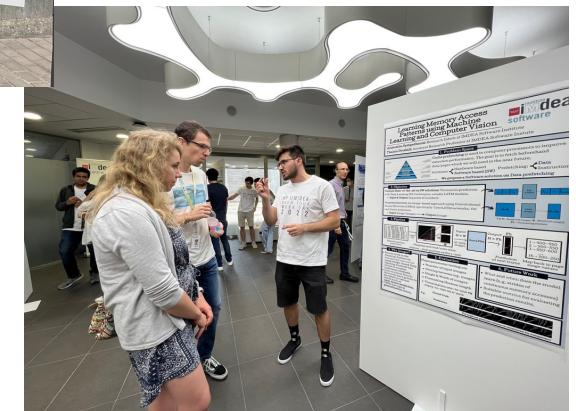
Poster Competition – June 2022



Easter Egg Hunt – April 2022

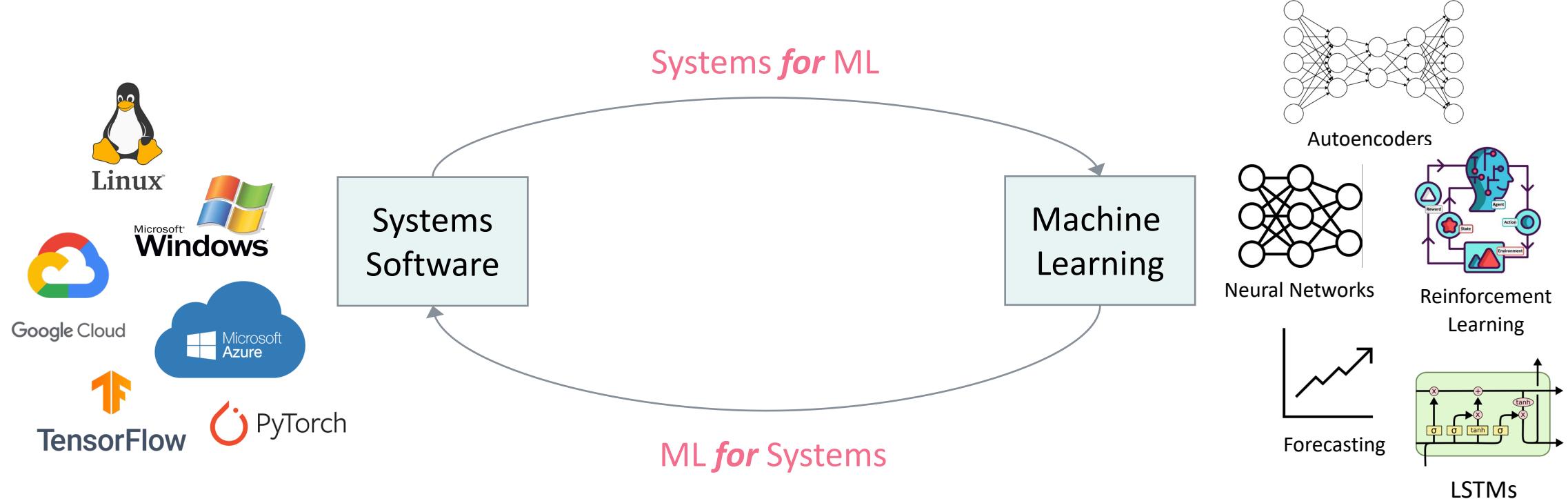


Soft Skills Workshop – May 2022



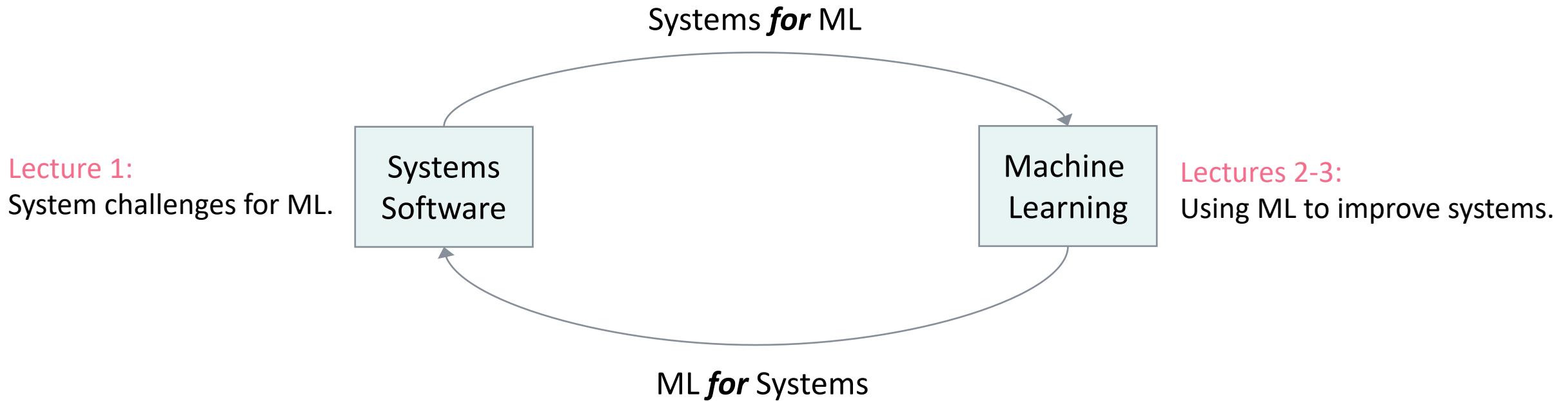
About My Research

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



About This Seminar Series

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



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

Today's Lecture

A Berkeley View of Systems Challenges for AI

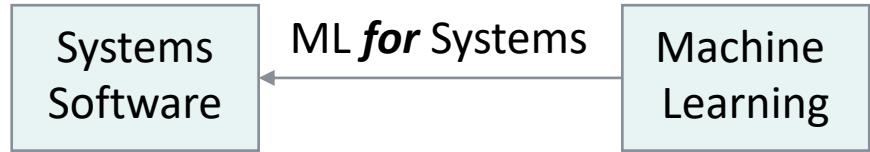
Today's paper:

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?

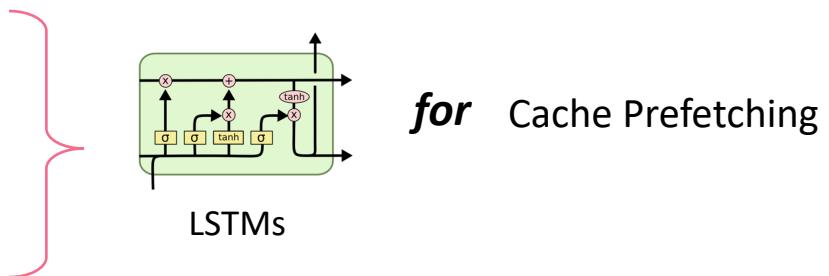
Next Lectures



Lecture 2

Learning Memory Access Patterns

Milad Hashemi¹ Kevin Swersky¹ Jamie A. Smith¹ Grant Ayers^{2*} Heiner Litz^{3*} Jichuan Chang¹
Christos Kozyrakis² Parthasarathy Ranganathan¹



for Cache Prefetching

Lecture 3

Kleio: A Hybrid Memory Page Scheduler with Machine Intelligence

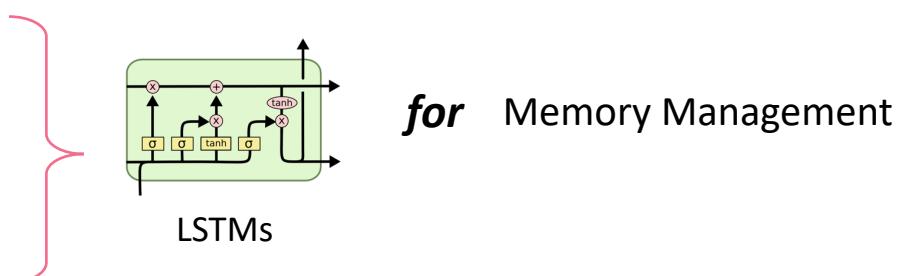
Thaleia Dimitra Doudali
Georgia Institute of Technology
thdoudali@gatech.edu

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

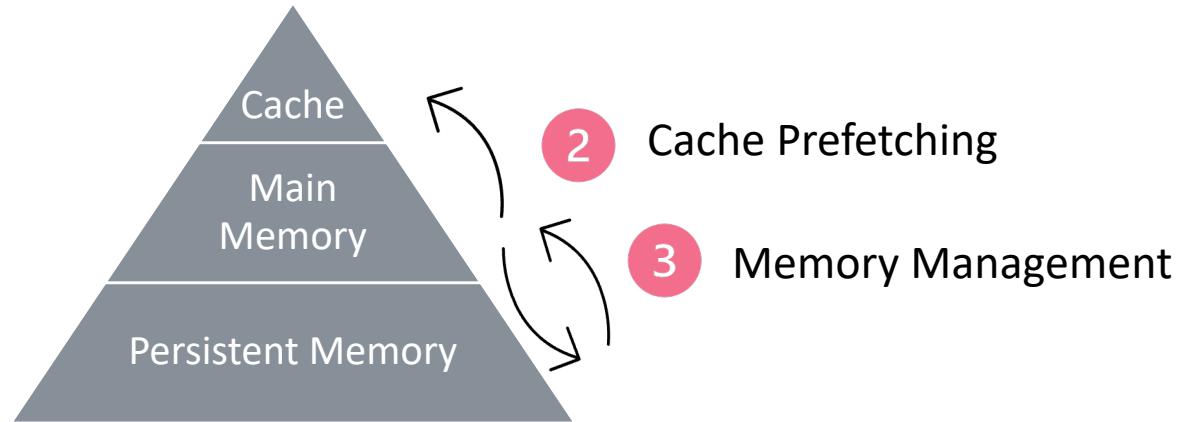


for Memory Management

Next Lectures



Data Transfers



Structure of lectures: 2 3

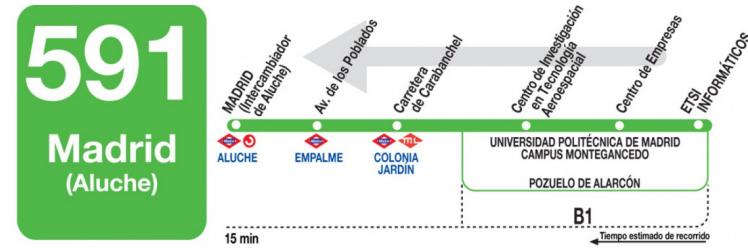
- Overview of the problem (cache prefetching, memory management).
- Overview of existing non-ML solutions.
- Overview of the ML-based method and how it's used to solve the problem.

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.



Horarios de salida de la ETSI Informáticos					
140920					
Lunes a	(Vigente de 1 de septiembre a 30 de junio) - Días lectivos				
De 10:05 a 13:35	cada 30	minutos			
A 13:55	14:10				
De 14:35 a 18:05	cada 30	minutos			
De 18:45 a 21:00	cada 45	minutos			

Contact

- Via email: thaleia.doudali@imdea.org



Website

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

Teaching

Spring 2023

MLArchSys Seminar Series.

At the MUSS and EMSE Master Programs of the School of Computer Science at Universidad Politécnica de Madrid. [MUSS Link](#) [EMSE Link](#)

Seminar 1: Introduction to Machine Learning for Computer Architecture and Systems. [Slides](#) [Paper](#)

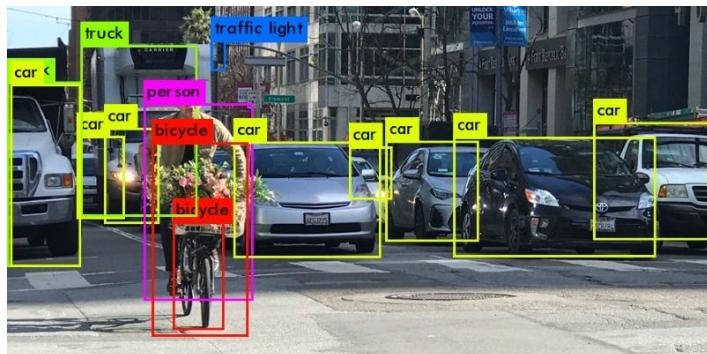
Seminar 2: Machine Learning for Cache Prefetching. [Slides](#) [Paper](#)

Seminar 3: Machine Learning for Hybrid Memory Management. [Slides](#) [Paper](#)

Artificial Intelligence is Everywhere



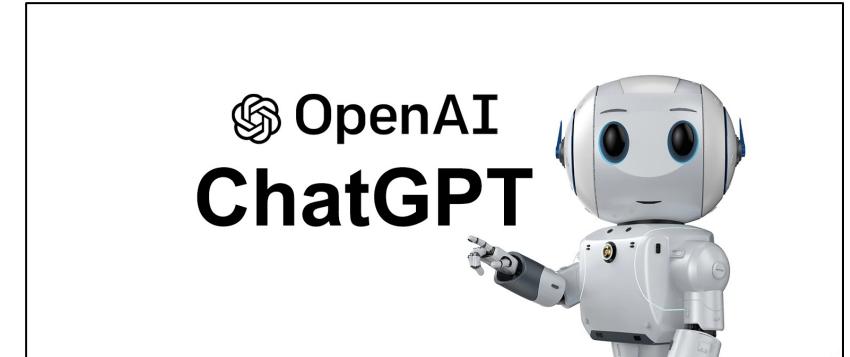
Self-driving cars



Object Recognition



Medical Diagnosis



Chatbot - Personal Assistants



Nurse Education in Practice
Volume 66, January 2023, 103537



Editorial

Open artificial intelligence platforms in nursing education: Tools for academic progress or abuse?

Siobhan O'Connor^a ChatGPT^b

Show more ▾

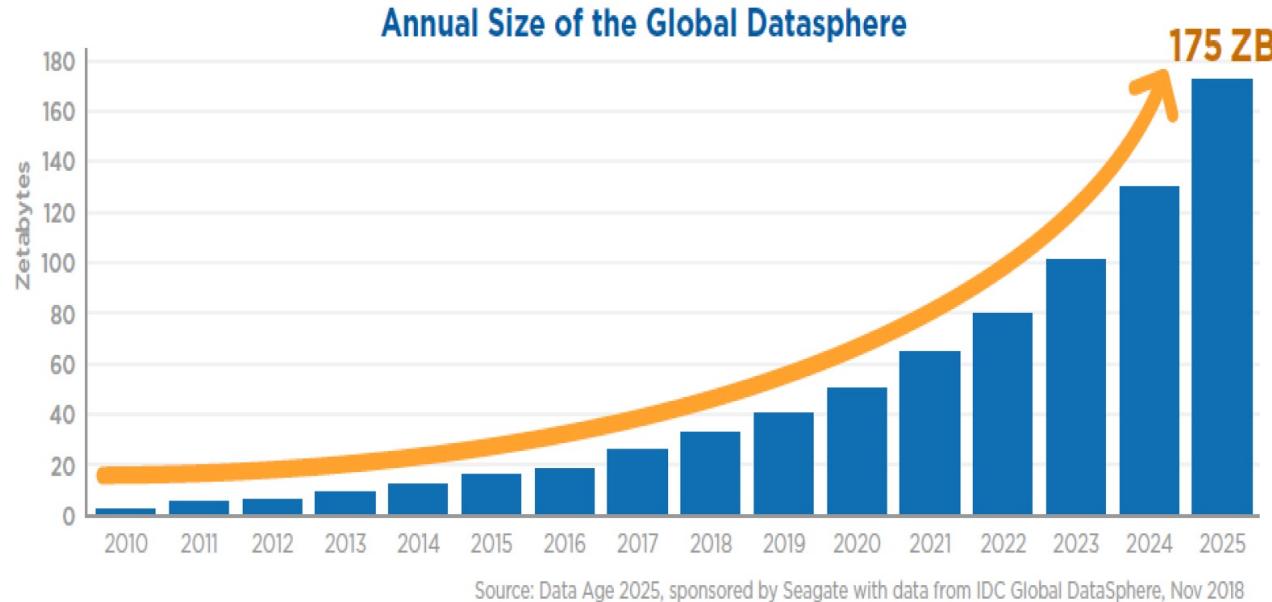
+ Add to Mendeley Share Cite

<https://doi.org/10.1016/j.nepr.2022.103537>

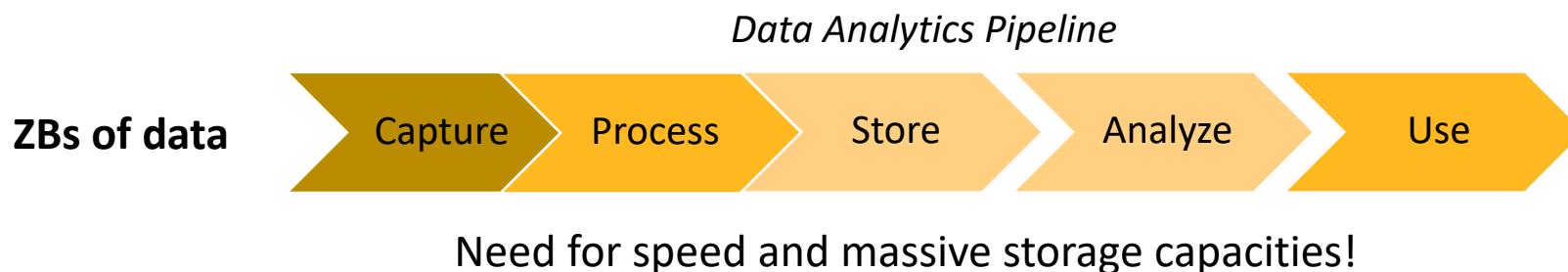
Get rights and content

What drives the success of AI?

BIG DATA

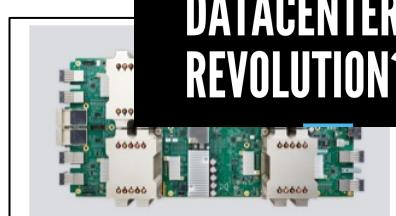
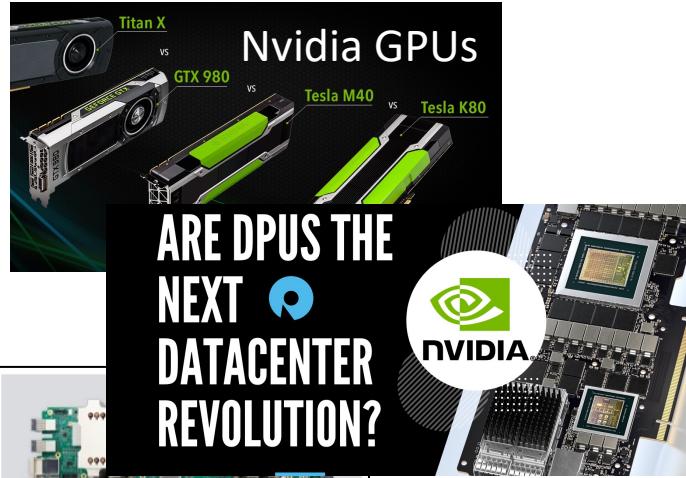


Megabyte	1,000,000 bytes
Gigabyte	1,000,000,000 bytes
Terabyte	1,000,000,000,000 bytes
Petabyte	1,000,000,000,000,000 bytes
Exabyte	1,000,000,000,000,000,000 bytes
Zettabyte	1,000,000,000,000,000,000,000 bytes
Yottabyte	1,000,000,000,000,000,000,000,000 bytes



What drives the success of AI?

NEW HARDWARE



Cloud TPU v2
180 teraflops
64 GB High Bandwidth Memory (HBM)



Accelerators even in Laptops.



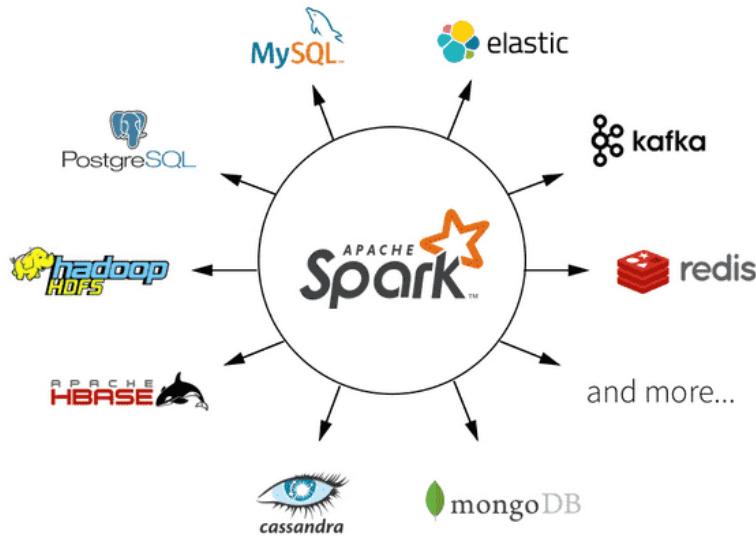
Supercomputers



Datacenters all over the world.

What drives the success of AI?

NEW SOFTWARE

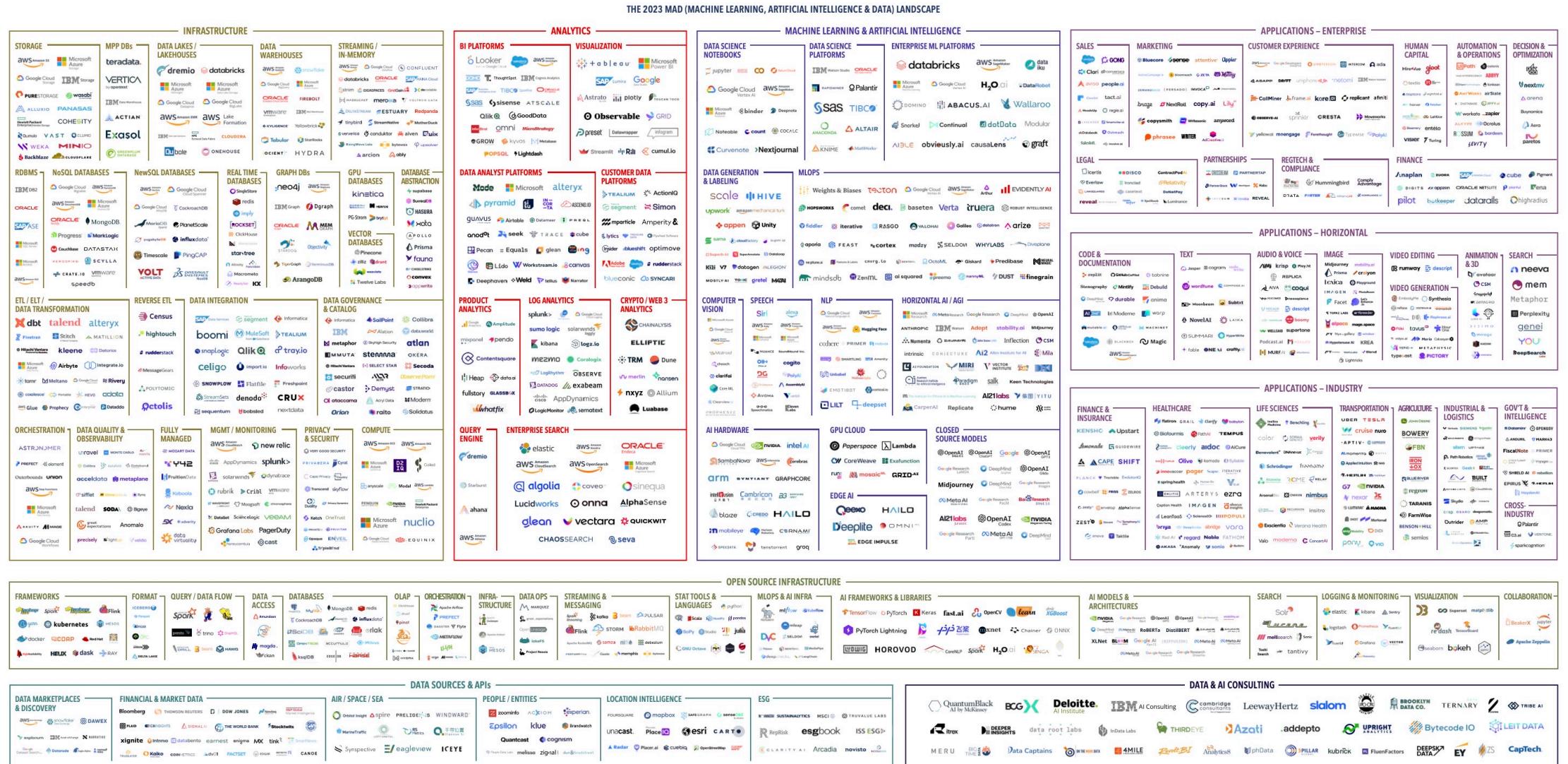


Software for Big Data Processing

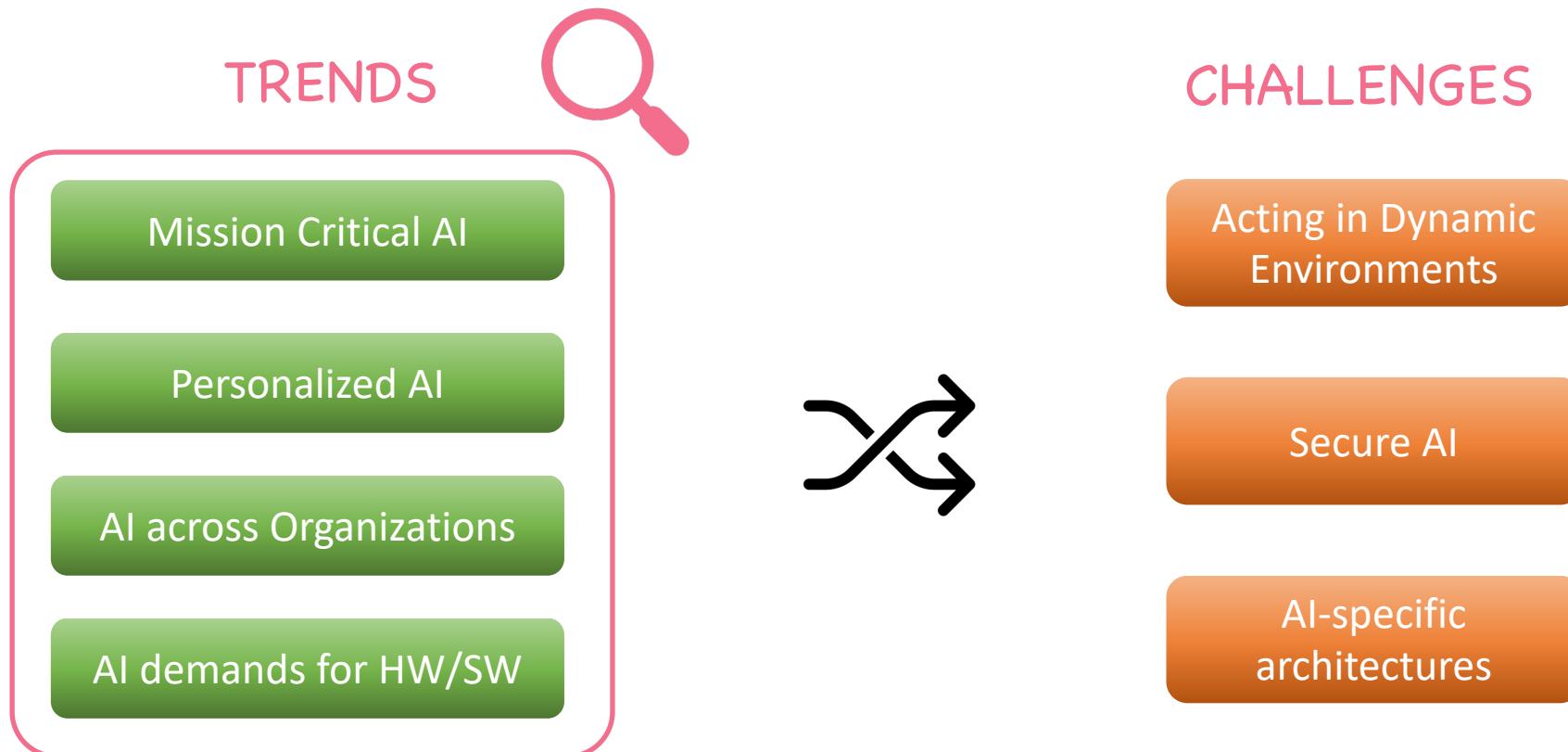


Software for Machine Learning Pipelines

The AI Landscape

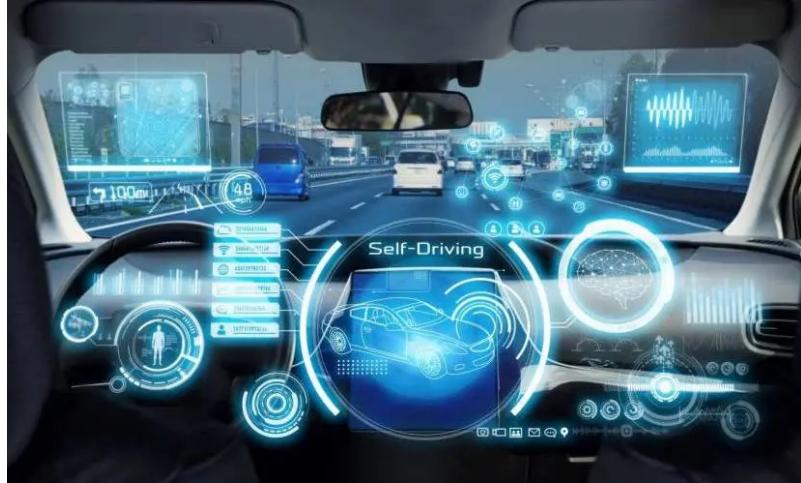


Trends and Challenges in AI



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

CHALLENGES



Autonomous Driving

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

Acting in Dynamic Environments

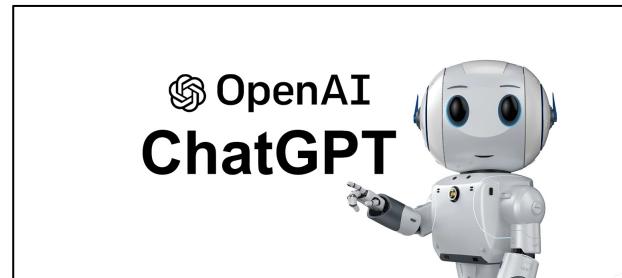
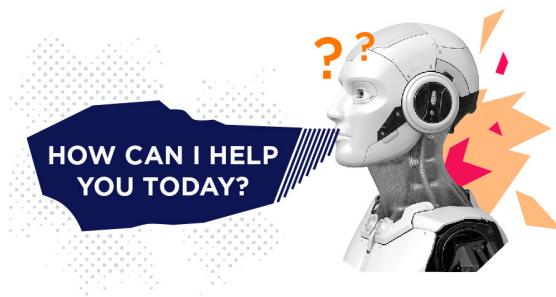
Secure AI

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.

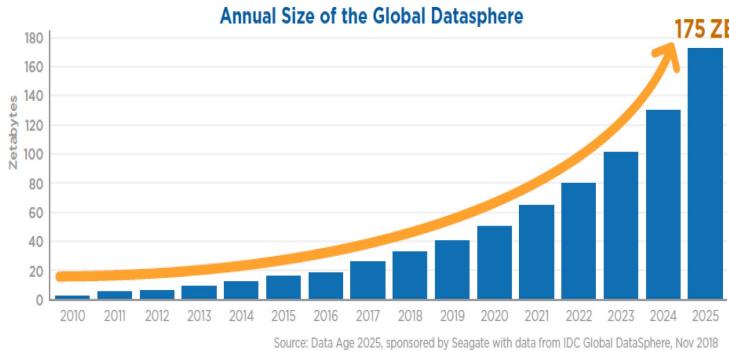
Secure AI

AI-specific architectures

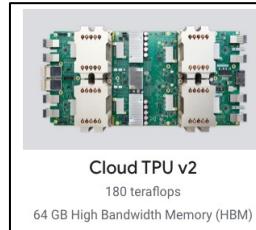
Challenges: Design AI systems that can train on datasets owned by different organizations without compromising their confidentiality.

TRENDS That AI Creates

AI demands for HW / SW



Google Cloud



Google

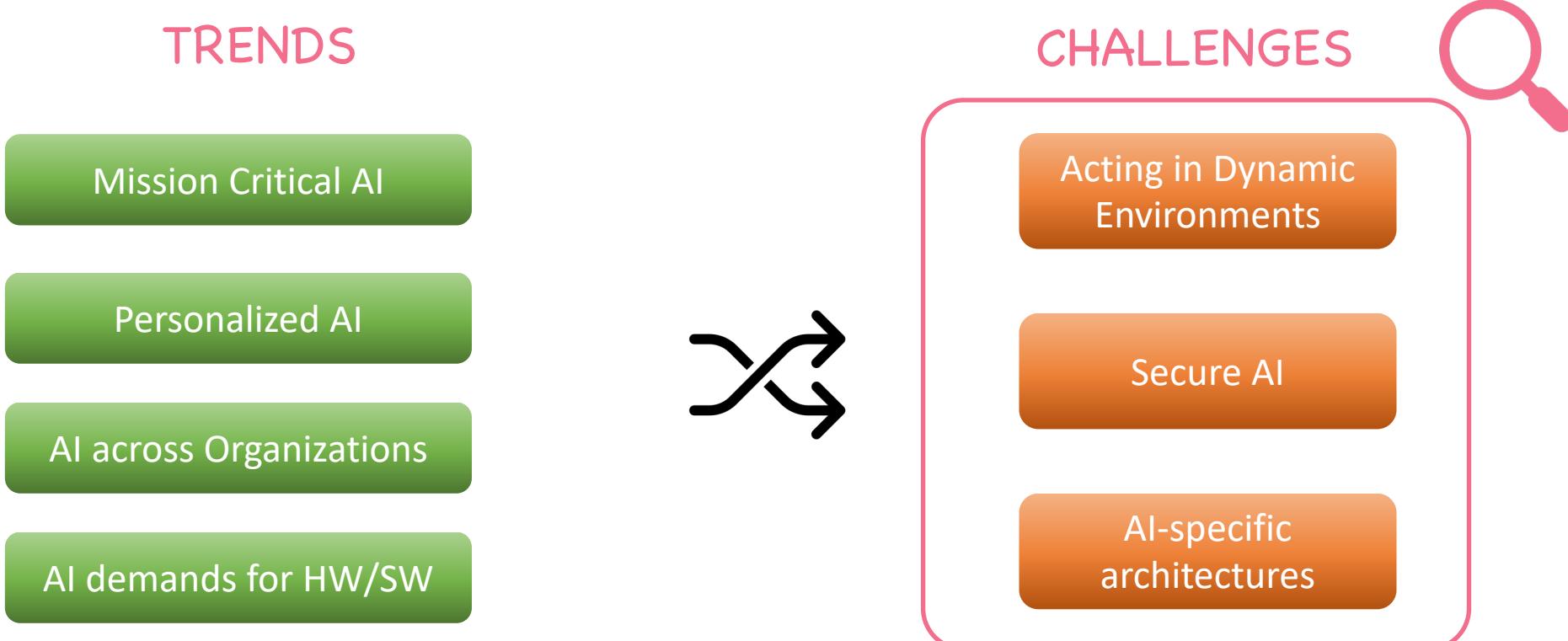
CHALLENGES

- 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

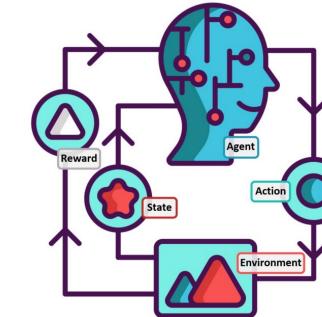
R1: Continual Learning



Autonomous Driving

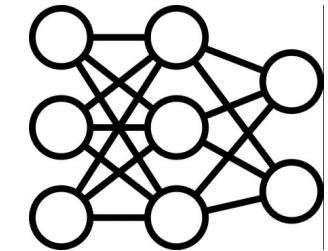
- Quick response.
- Unexpected conditions.
- Continually adapt.
- Learn new skills.

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



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

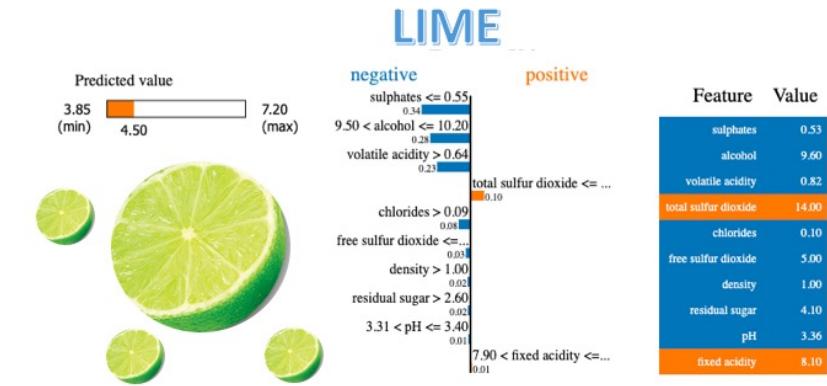
Acting in Dynamic Environments

R3: Explainable Decisions



Medical Diagnosis

- Which X-ray feature led to the diagnosis?
- Explanation meaningful 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

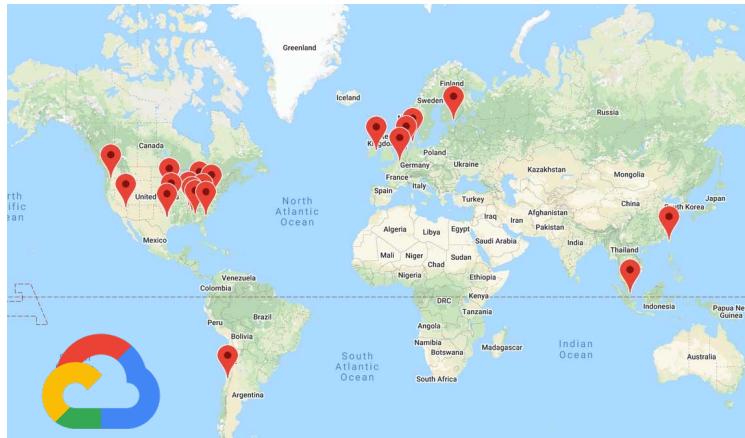
AI-specific architectures

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

CHALLENGES That AI Creates

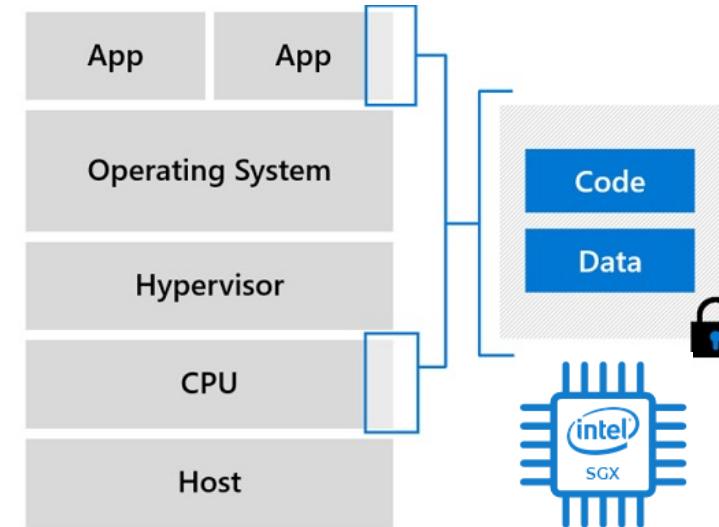
Secure AI

R4: Secure Enclaves



Google Cloud

No control “where” the code runs.



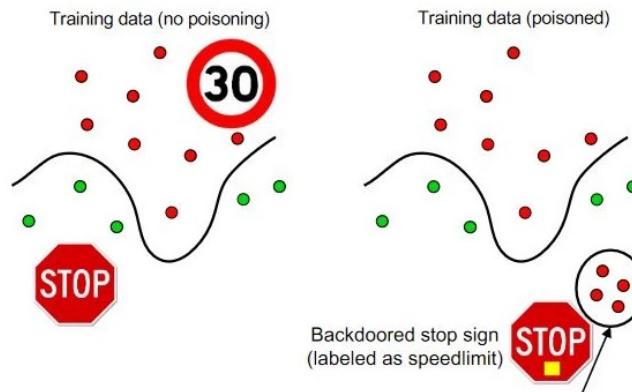
Enclave: a secure execution environment, usually enforced by hardware.

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

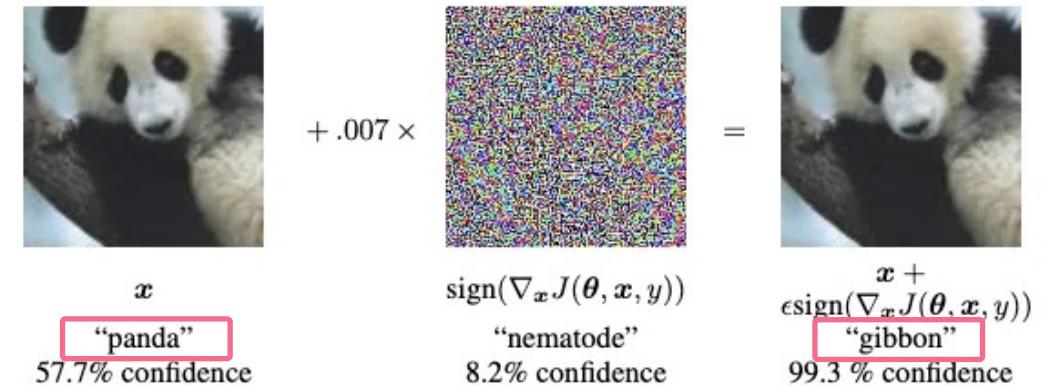
CHALLENGES That AI Creates

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).

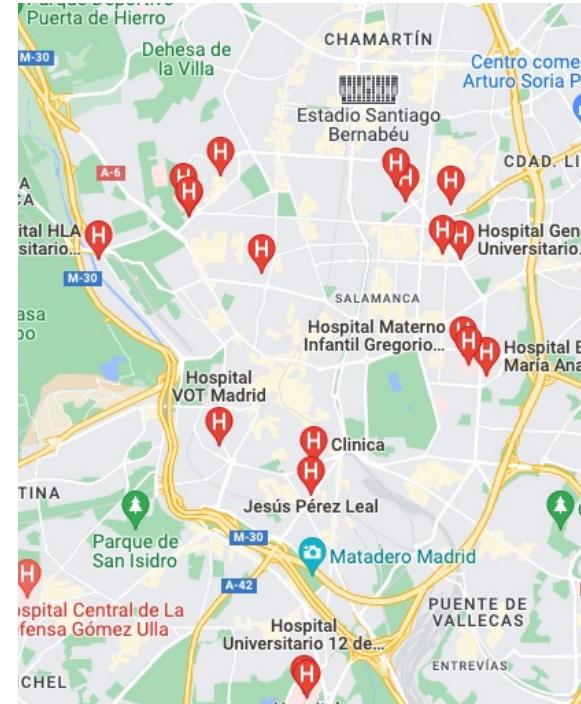
CHALLENGES That AI Creates

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

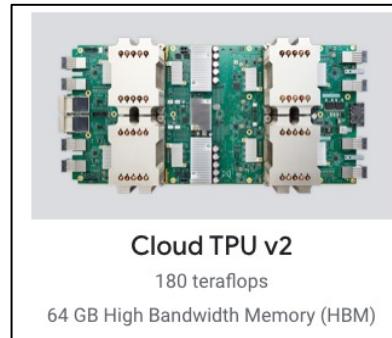
AI-specific architectures

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

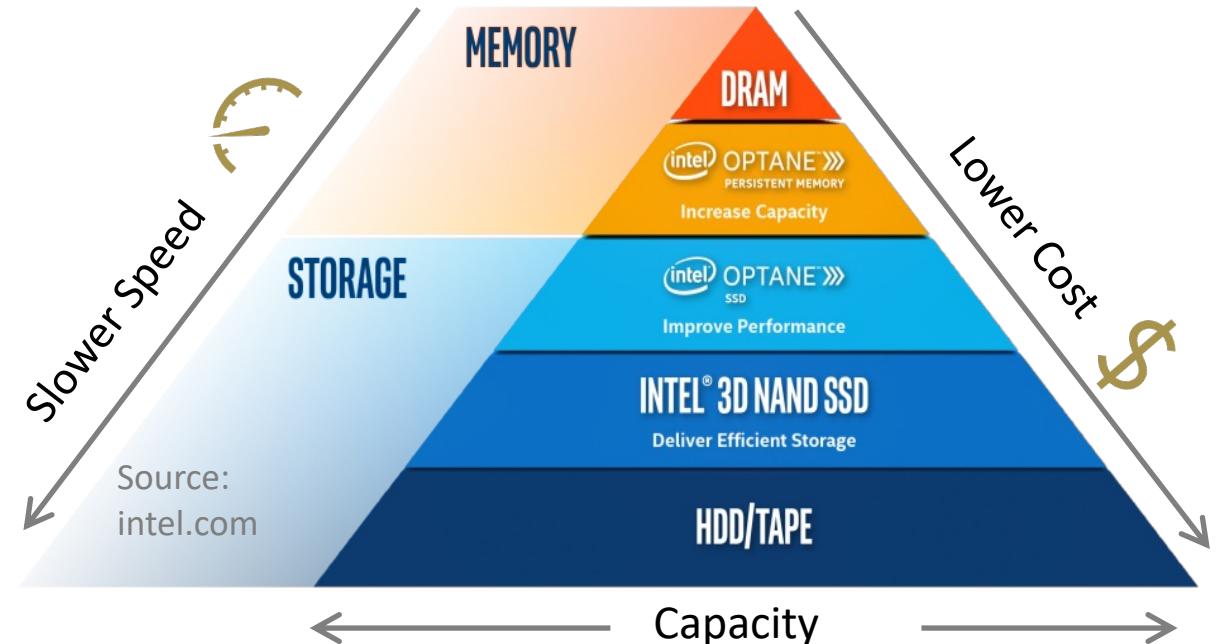
CHALLENGES That AI Creates

AI-specific
architectures

R7: Domain Specific Hardware



Google



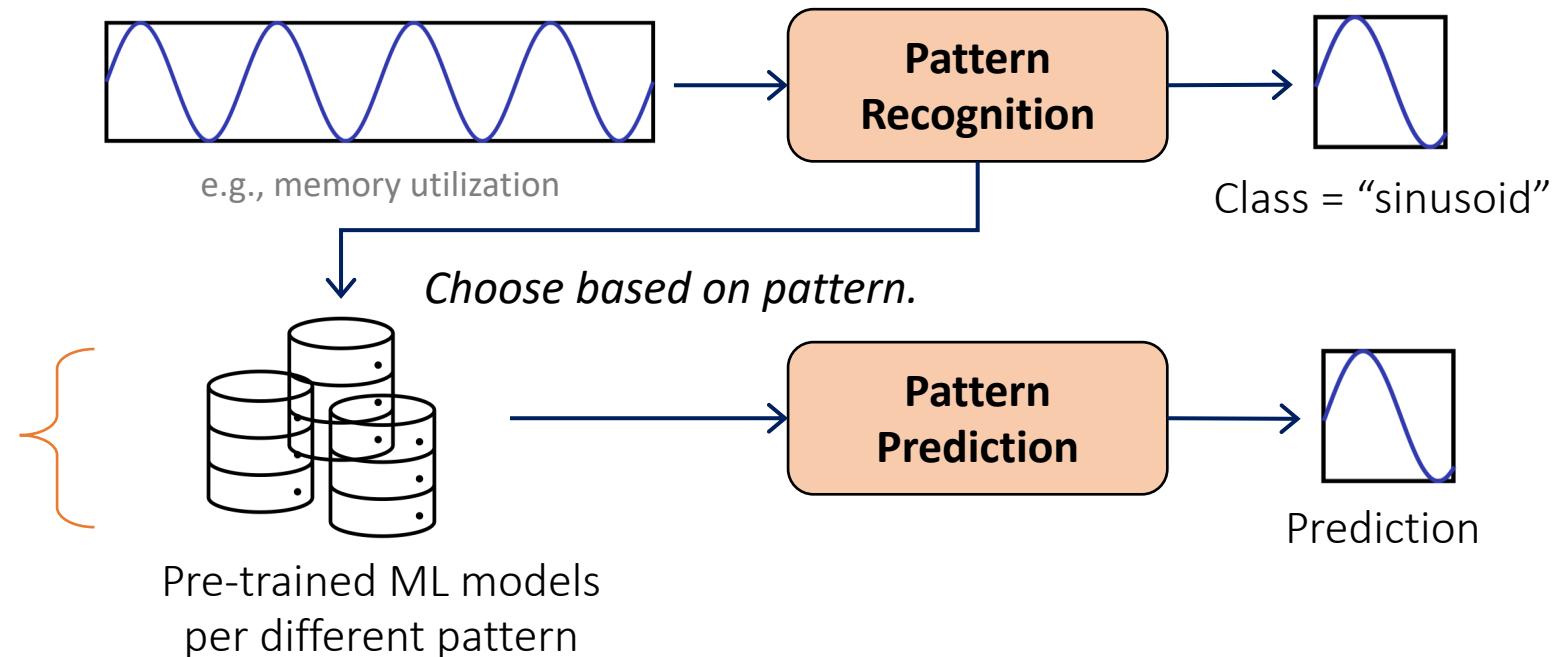
Systems Research: Build Systems that efficiently use new hardware technologies, like accelerators, new types of memory and storage devices.

CHALLENGES That AI Creates

AI-specific
architectures

R8: Composable AI Systems

- Modularity
- Composability
- Focused expertise



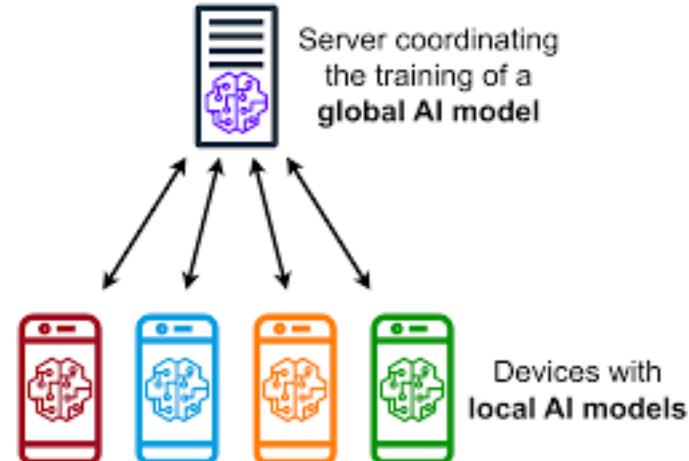
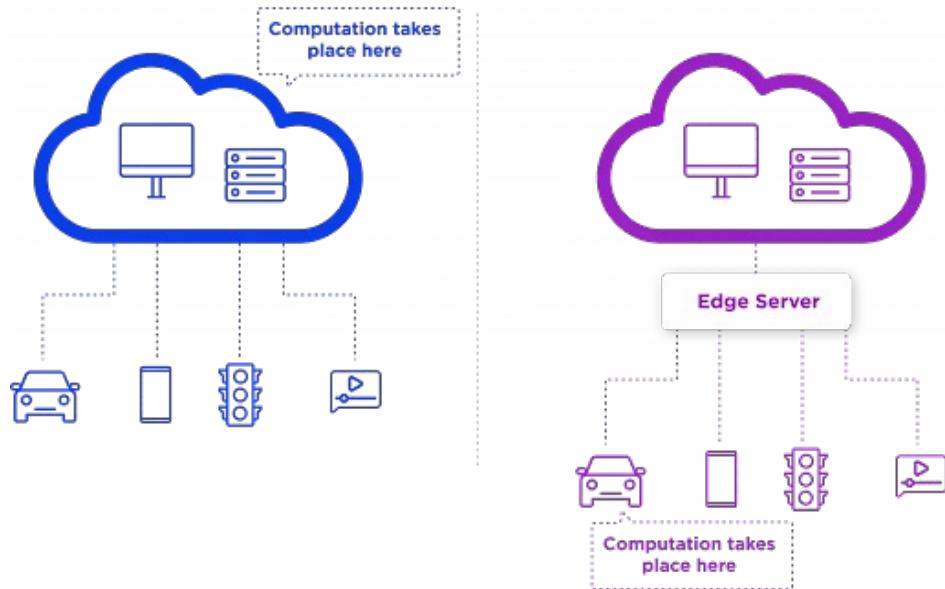
Systems Research: Build systems that allow the composition of models and actions in a modular and flexible manner.

CHALLENGES That AI Creates

AI-specific
architectures

R9: Cloud-edge Systems

Cloud Computing vs Edge Computing



- 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.

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?

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 choices. 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>