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.

Undergrad in ECE at NTUA, Athens, Greece.

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
About My Research

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

- Systems for ML
- Machine Learning
- ML for Systems
- Systems Software
- Machine Learning

- 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.
A Berkeley View of Systems Challenges for AI


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

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

ML for Systems

Machine Learning

for Cache Prefetching

LSTMs

for Memory Management

LSTMs
Structure of lectures:
1. Overview of the problem (cache prefetching, memory management).
2. Overview of existing non-ML solutions.
3. 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/

Teaching

Spring 2023
MLArchSys Seminar Series.
At the MUSS and EMSE Master Programs of the School of Computer Science at Universidad Politécnica de Madrid. MUS Link EMSE Link

Seminar 2: Maching Learning for Cache Prefetching. Slides Paper
Seminar 3: Maching Learning for Hybrid Memory Management. Slides Paper
Artificial Intelligence is Everywhere

Self-driving cars

Object Recognition

Medical Diagnosis

Chatbot - Personal Assistants
What drives the success of AI?

**BIG DATA**

**Data Analytics Pipeline**

- **Capture**
- **Process**
- **Store**
- **Analyze**
- **Use**

Need for speed and massive storage capacities!
What drives the success of AI?

**NEW HARDWARE**

- Nvidia GPUs
- Supercomputers
- Datacenters all over the world.
- Accelerators even in Laptops.
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

CHALLENGES

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

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

CHALLENGES

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

Personal Assistants

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

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.

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

CHALLENGES

• Massive amounts of data.

• Need for new hardware/software solutions customized to AI needs.

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

**R1: Continual Learning**

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

**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

R2: Robust Decisions

**Microsoft chatbot is taught to swear on Twitter**

By Jane Wakefield

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

R3: Explainable Decisions

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

Medical Diagnosis

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

**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

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

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**

**R7: Domain Specific Hardware**

**Nvidia GPUs**

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

**Source:** intel.com

**Google**

---

**Systems Research:** Build Systems that efficiently use new hardware technologies, like accelerators, new types of memory and storage devices.
R8: Composable AI Systems

- Modularity
- Composability
- Focused expertise

Pre-trained ML models per different pattern

\textbf{Pattern Recognition}

\textbf{Choose based on pattern.}

\textbf{Pattern Prediction}

e.g., memory utilization

Class = “sinusoid”

\textbf{Systems Research:} Build systems that allow the composition of models and actions in a modular and flexible manner.
CHALLENGES That AI Creates

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.
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://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?

- **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?

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