



ASE 2023: As it Happened!



Shubham Kulkarni, Arya Marda, Karthik Vaidyanathan
SA4S Research Group, Software Engineering Research Center (SERC), IIIT Hyderabad

Title : Towards Self-Adaptive Machine Learning-Enabled Systems Through QoS-Aware Model Switching



Presented our paper on 12 Sep 2023 at ASE 2023 (Core A*) in NIER Track

Our exploration into self-adaptation techniques in Machine Learning-Enabled Systems (MLS) has caught the attention of many, leading to engaging discussions on adaptation decisions and how our approach can be applied to other system goals.

Special Thanks to Microsoft Research India for travel grant support for attending ASE 2023!



Questions Asked During Presentation:

Q1 : Thanks, Can u explain me trigger, trigger for adaptation?
So, the adaptation is adapting, how??

Answer : I explained him, and he understood: Talked about
the process, MAPE-K again

Follow Up : Have u encountered any frequent or unnecessary
model switching, say from model A to B to A and so on...

Answer : Yes, it could possibly occur , mitigating it with avg of
last n results, also latency consideration is required plus to
avoid it we do have twait., also have preloading of models

AdaMLs: Its Out & Popular

Questions Asked During Presentation:



Q2 : Thanks, I really would like like to continue this conversation later with u, question is if u just consider one parameter in terms of adaptation so just switching the model or there are also multiple parameters that are there?

- Alessandro Tundo, University of Milano-Bicocca Italy

Answer : Explained him our logic again. (Have full video of this for reference)



He and his prof are interested in collaborative research ahead with us!

Their work- Identified a need for balancing energy consumption with application objectives in AI-based applications on edge devices, especially for IoT scenarios like smart city monitoring.

2. Energy-Aware Approach:

- Proposed an energy-aware design and deployment strategy for self-adaptive AI-based applications to address the identified problem.

3. Meta-Heuristic Search:

- Utilized a meta-heuristic search procedure to find a set of configurations for self-adaptation with minimal empirical samples required.

4. Configuration Selection:

- Employed Weighted Gray Relational Analysis to finalize configurations and map them to the operation modes of the self-adaptive application.

5. Validation and Results:

- Validated the approach on a pedestrian detection application, achieving up to 81% energy savings with only a 2% to 6% drop in accuracy compared to non-adaptive baseline configurations.

Some More Accolades!



Doing great work, liked the idea and appreciated the PPT and presentation!

Interested and asked about future and asked to look at her lab for PhD or further research!

Prof Dr. Ajitha Ranjan
University of Edinburgh,
UK

- Said need of time, liked the examples and PPT.

Interested and asked me if I can join PhD at her lab



All of them liked our work.
1. Debashee Das – MS, ETH Zurich

2. Prof Jyothi Vemudara, IITH

3. Prof Subhajit Roy, IITK
(Appreciated Tradeoff and asked us to work more on it instead of other KPIs, again asked if I can join PhD with him at IITK)

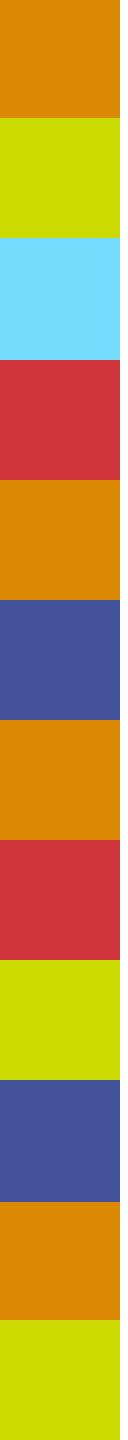
4. Tamal Mondal, MS, IITH





Some more Accolades

- Prof Raghavan, IISc : Appreciated work, liked tradeoff and unsupervised learning. Also surprised to see MS student there, so later appreciated SERC and culture!
- This group is from Peking University, China, they work on amazing ideas all students of prof. Dan Hao, they liked and interested to use our idea!



Met and Discussed our paper with The Danny Tarlow, Keynote Speaker, ASE 2023, Research Scientist, Google Brain - Montreal

- Found our idea interesting! Really liked the experiments!
- His comments:
 1. Service Assurance and fairness should be considered, why if some models will always suffer...
 2. Context based self-adaptation should be considered based on type of request rather than on KPIs
 3. I am afraid that these idea with services like GPT-4 which is switching models and changing service quality – Damm Scary
 4. Where is the stability? Define some things in that direction
 5. Object detection is way different from GenAI, where u want to switch?
 6. If I am paying for some service, I am hesitant for such ideas there
 7. But Why and Why? He also cracked some jokes when I asked will u do this with your current projects at Google, he said I am afraid to do it! 😊

And That Was ASE for our research!

Overall, our presentation was very much understandable, people liked it and appreciated it!

had so many discussions with friends, bus-mates and various profs all around got to know about their research work, culture, countries etc....!! Also had so many discussions with friends, bus-mates and various profs all around got to know about their research work, culture, countries etc....!!

Lets see what others were doing and how was the experience!



Keynote Speakers!



Prof. Julia Lawall
INRIA

She is sharing her experience on the automation of the evolution of software in the Linux Kernel  and Coccinelle !



Danny Tarlow
DeepMind Google,
Canada

Danny Tarlow on DIDACT

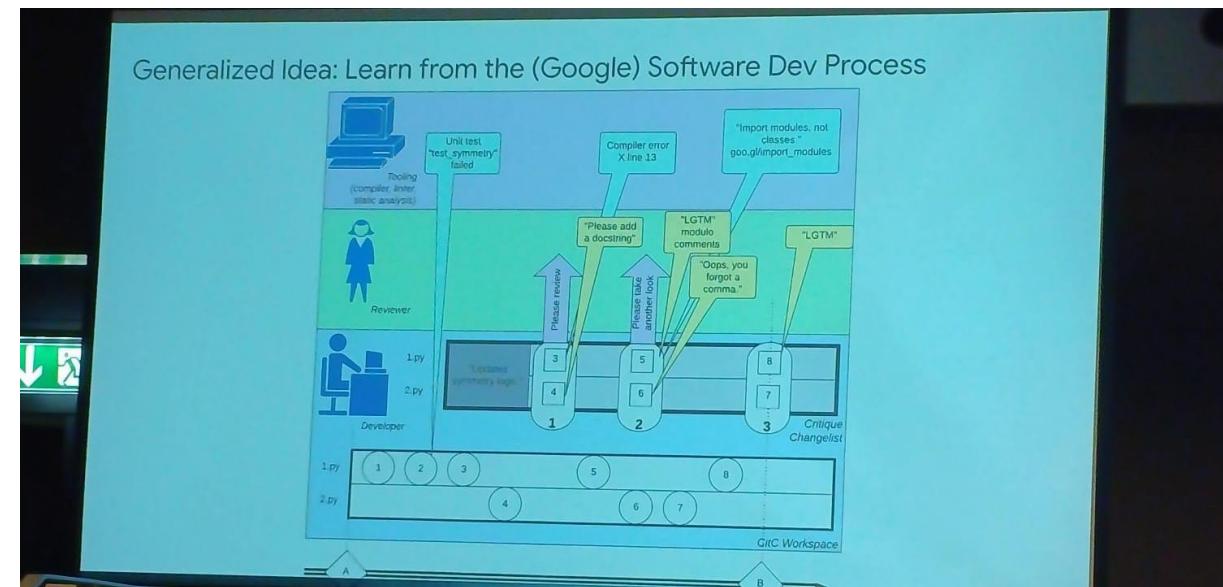
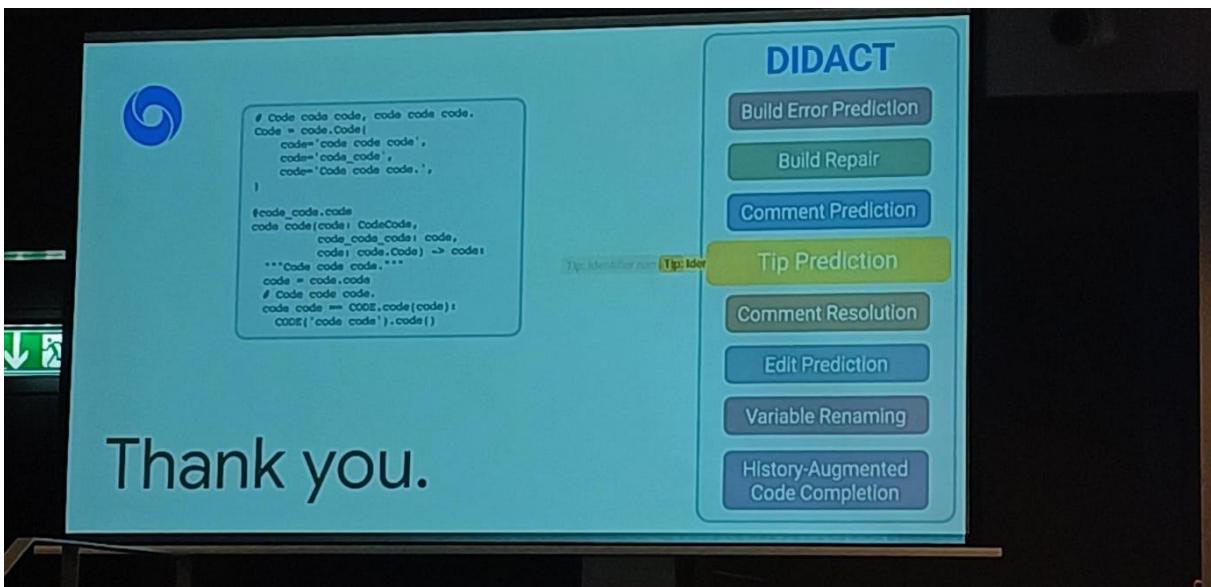
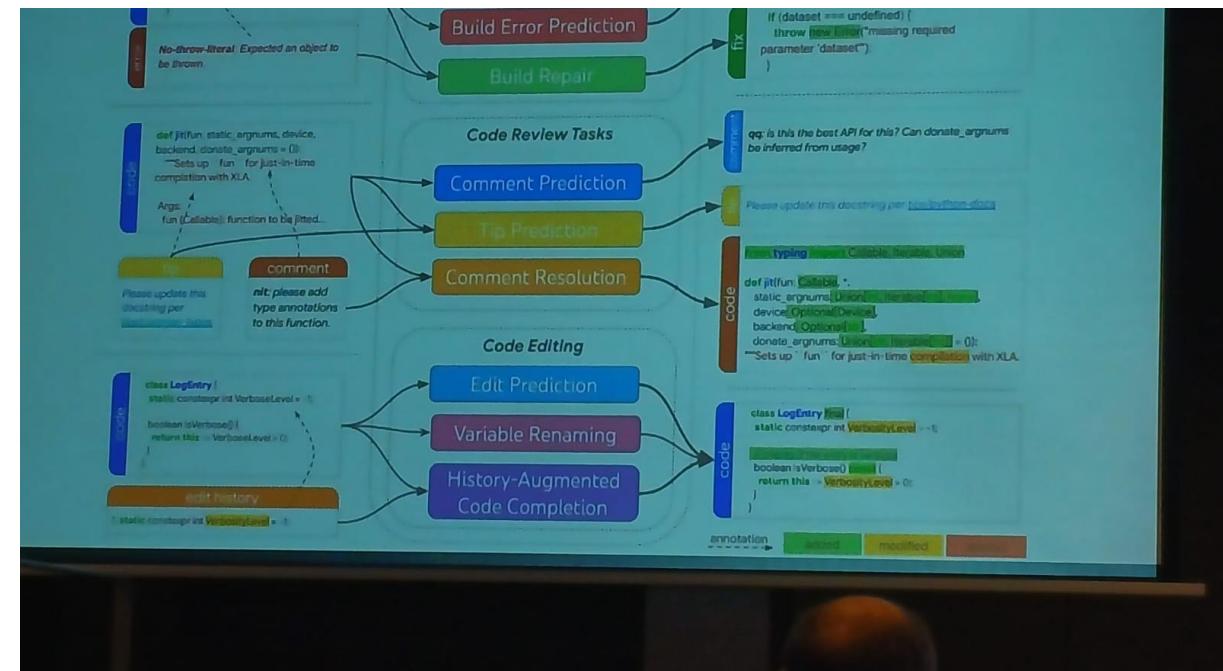


DIDACT (Dynamic Integrated Developer Activity) methodology:

1. Iterative Nature of Software Development: Highlighted the step-by-step improvement process involving various stakeholders and tools in software engineering.
2. Introduction to DIDACT: Introduced DIDACT for training large ML models using the software development process as training data, not just the final code.
3. Training Data Source: Utilized real-time developer interactions and responses during development as a rich training data source.
4. Google's Software Development Instrumentation: Leveraged Google's development environment to enhance the quantity and diversity of developer-activity data.
5. **Promising Outcomes: Showcased the usefulness of DIDACT to professional developers and its potential in imbuing ML models with practical software development skills.

DIDACT: Large sequence models for software development activities

Danny Tarlow
Research Scientist



Caveat

- We are fully aware that any intelligent capabilities displayed by LLMs are emergent behaviour enabled indirectly via mimicking human generated texts :)



Emergent behavior is a concept commonly used in various fields, including physics, biology, sociology, and computer science, to describe complex phenomena that arise from the interactions of simpler components or agents within a system. In essence, emergent behavior refers to patterns, properties, or behaviors that emerge at a higher level of organization or complexity but are not explicitly programmed or designed into the individual components of the system.

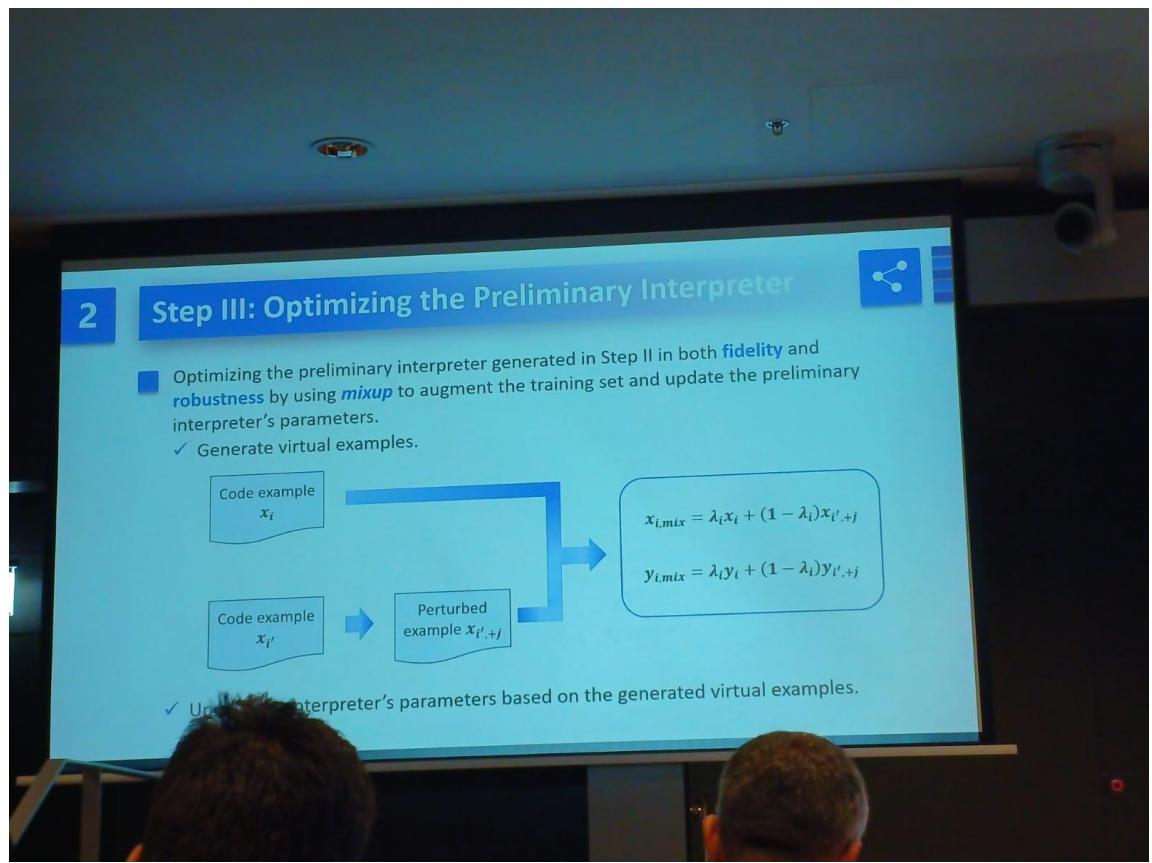
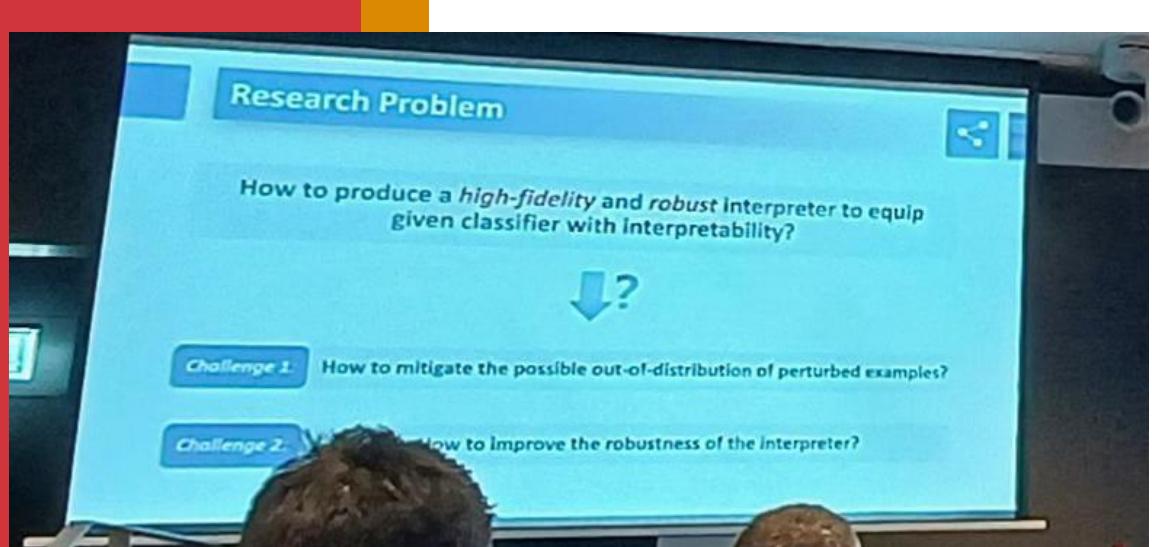
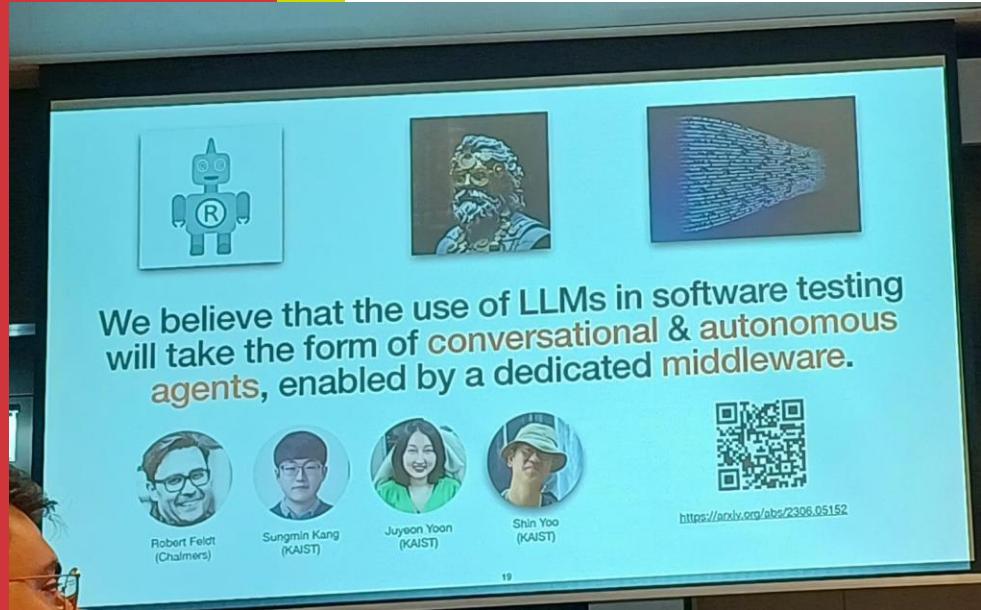
CASTLE

Finding failure-inducing test cases is challenging

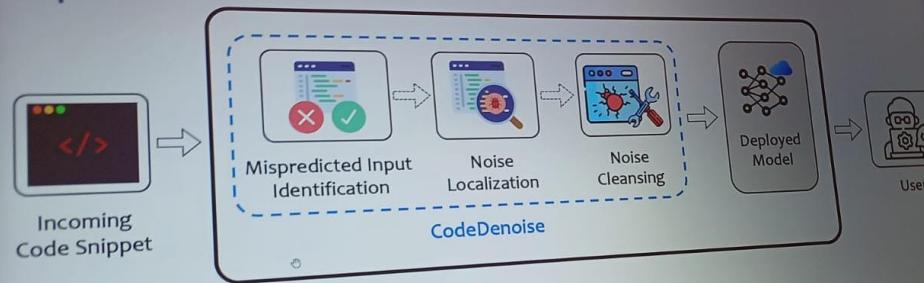
- Challenge 1: search space for failure-inducing test inputs is vast
- Challenge 2: constructing a test oracle is often an undecidable problem



Tzu-Chi Li | tcli@connect.ust.hk



Overview of CodeDenoise



The usage of CodeDenoise in practice:

- We treat CodeDenoise as a **post-processing module** and integrate it with the deployed deep **code model** as a system for making predictions in practice.
- In particular, CodeDenoise **just accesses the deep code model but not modifies it**, and thus it can accomplish the goal of on-the-fly improving performance of (deployed) models.

Input Denoising for Deep Code Models

```

1 def sort(x_list, y_length):
2     a1_selection = 0
3     flag = True
4     while flag:
5         flag = False
6         for i in range(1, y_length):
7             j = y_length - i
8             if x_list[j] < x_list[j-1]:
9                 x_list[j], x_list[j-1] = \
10                 x_list[j-1], x_list[j]
11             flag = True
12     a1_selection += 1
13 return x_list, a1_selection
  
```

Ground-truth Label: Bubble Sort
Prediction Result: Selection Sort
Noisy Identifier: a1_selection

(1) Noisy Code

Noisy identifiers: the identifier makes the largest contribution to the misprediction.

This motivates the potential of on-the-fly improving performance of (deployed) deep code models through input denoising.

```

1 def sort(x_list, y_length):
2     count = 0
3     flag = True
4     while flag:
5         flag = False
6         for i in range(1, y_length):
7             j = y_length - i
8             if x_list[j] < x_list[j-1]:
9                 x_list[j], x_list[j-1] = \
10                 x_list[j-1], x_list[j]
11             flag = True
12         count += 1
13 return x_list, count
  
```

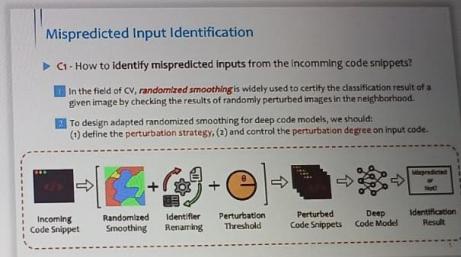
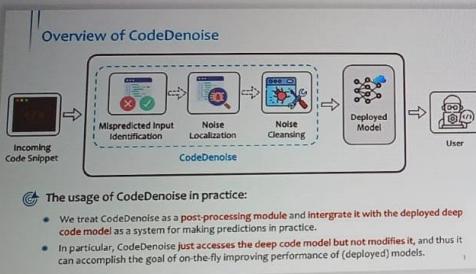
Ground-truth Label: Bubble Sort
Prediction Result: Bubble Sort
Denoised Identifier: count

(2) Denoised Code

Challenges

- How to identify mispredicted inputs from the incoming code snippets?
- How to localize noise (identifier-level) resulting in misprediction from a given code snippet?
- How to cleanse noise to make the code snippet be predicted correctly?

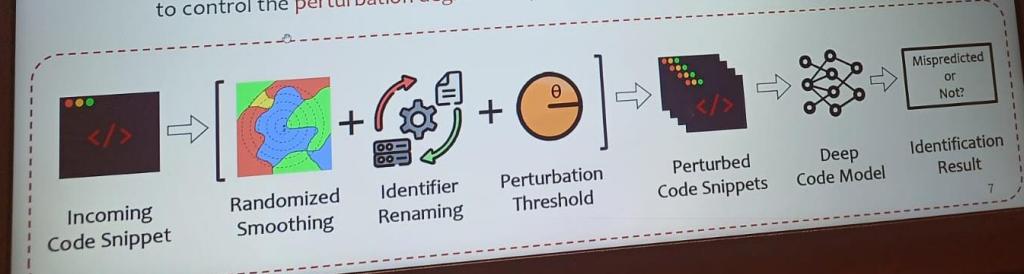
Summary & Thanks !

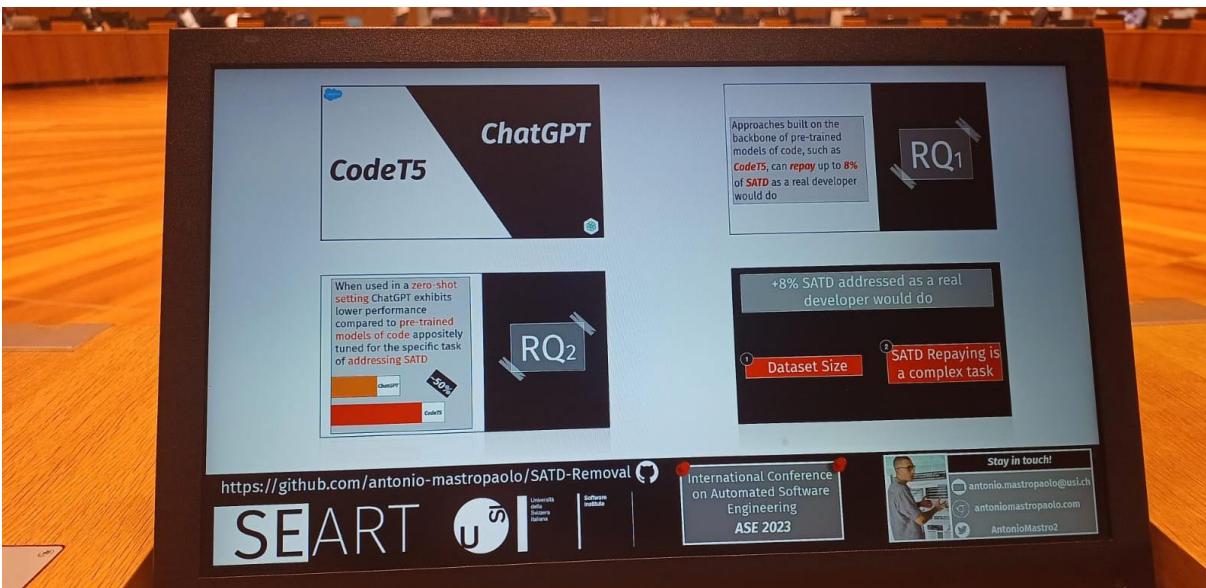
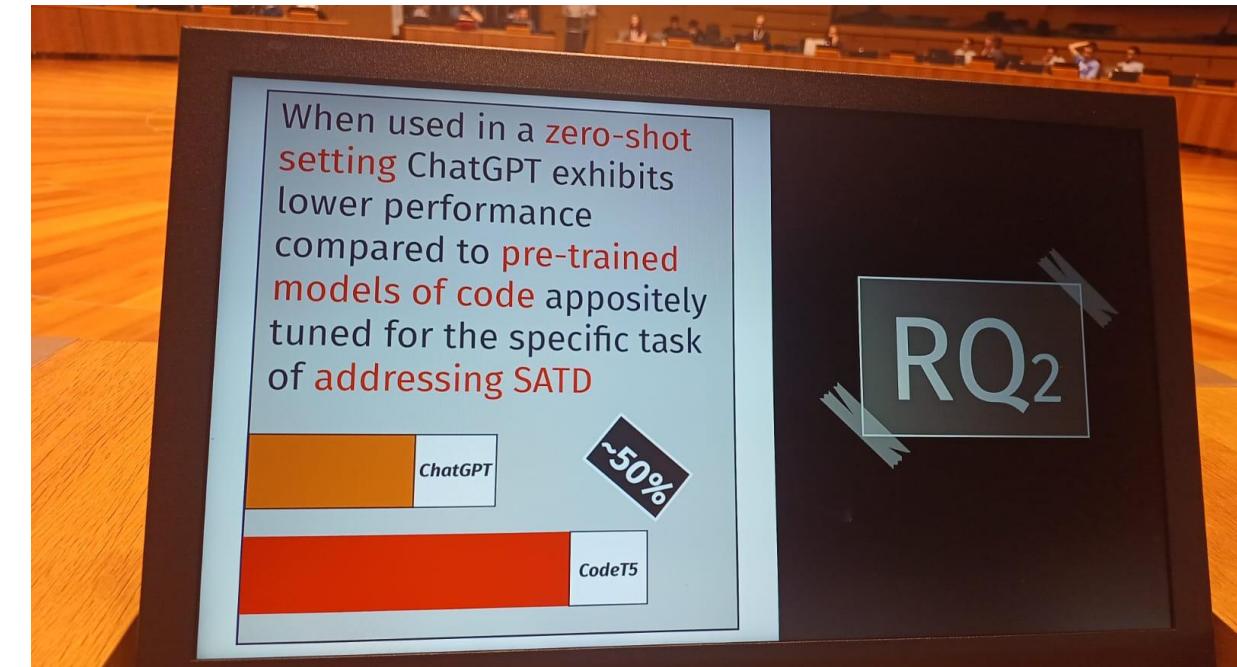
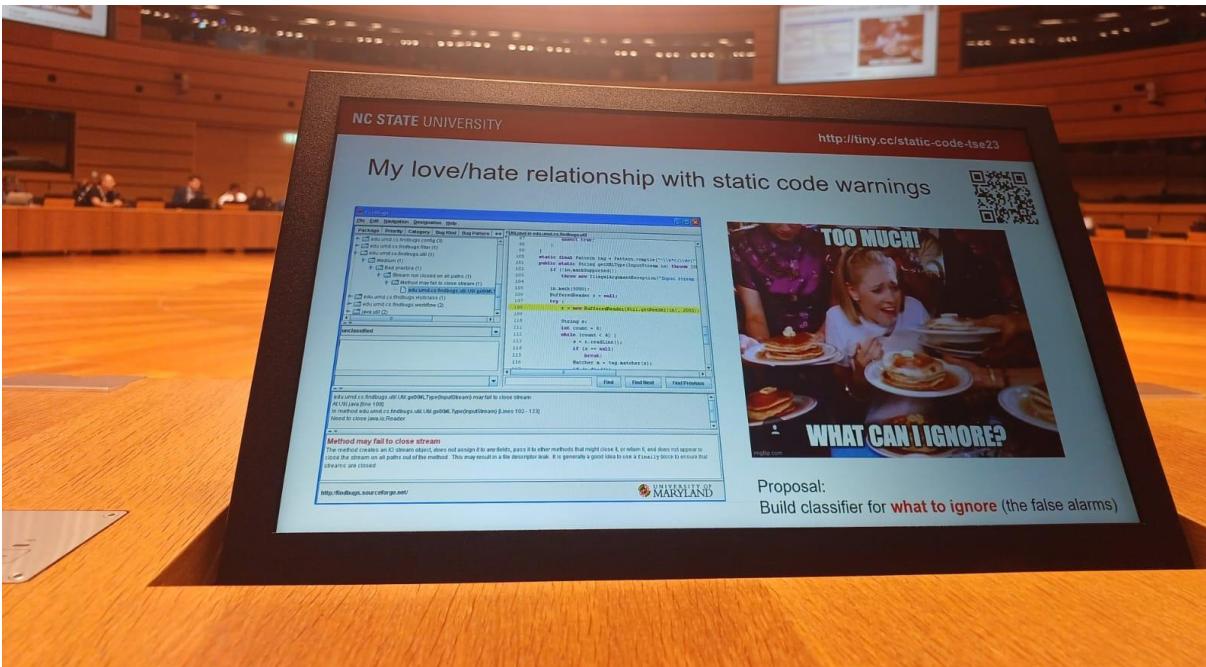


Mispredicted Input Identification

C1 - How to identify mispredicted inputs from the incoming code snippets?

- In the field of CV, **randomized smoothing** is widely used to certify the classification result of a given image by checking the results of randomly perturbed images in the neighborhood.
- To design adapted randomized smoothing for deep code models, we should: (1) define the number of perturbed code snippets for the **perturbation strategy**, (2) and define the threshold to control the **perturbation degree** on input code.

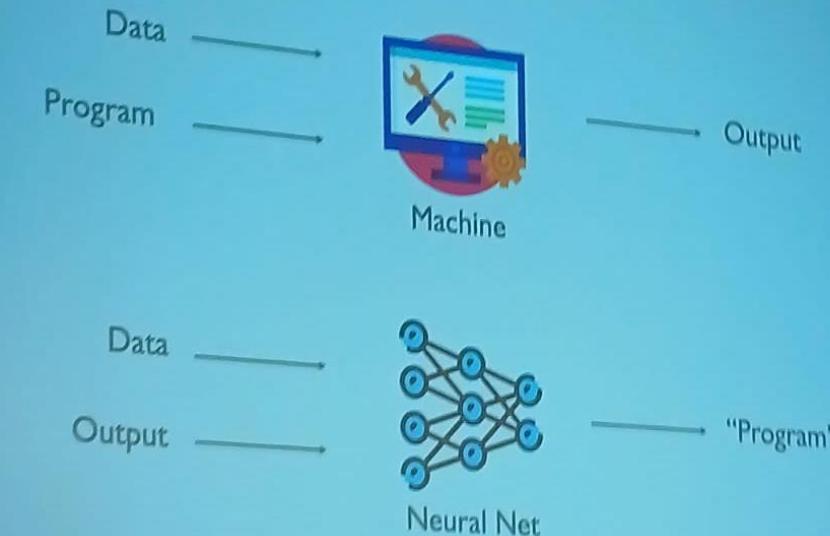




YESTERDAY'S DEVS VS. TOMORROW's DEVS

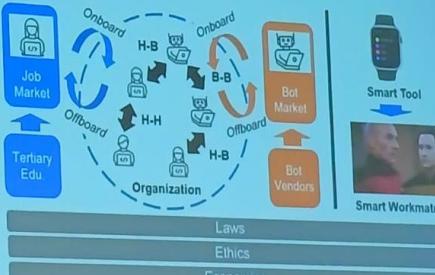


SOFTWARE 1.0 V.S SOFTWARE 2.0



Vision: Software Engineering 2.0 (SE 2.0)

Symbiotic workforce of autonomous, responsible, intelligent
bots and software engineers



Trustworthy and Synergistic AI4SE:
Vision and the Road Ahead

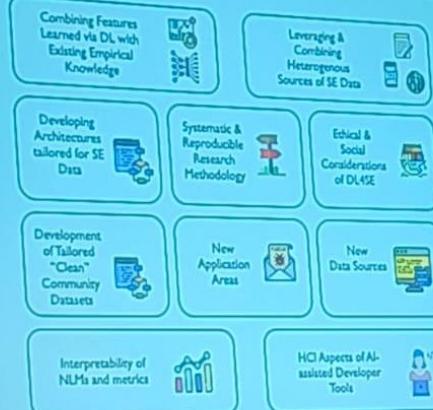
David Lo

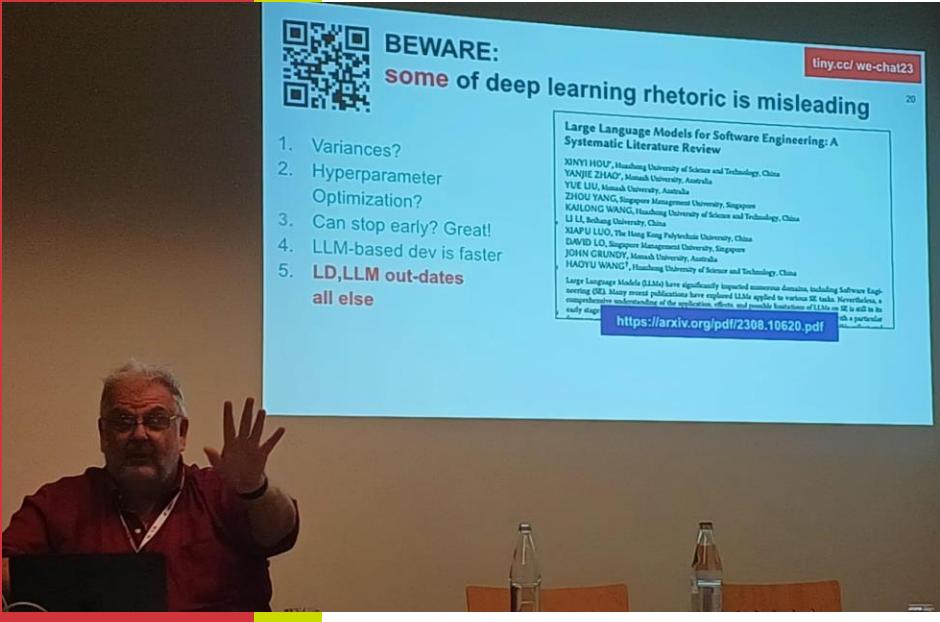
ICSE'23 Future of SE Talk

SCHOOL OF
COMPUTING AND
INFORMATION SYSTEMS

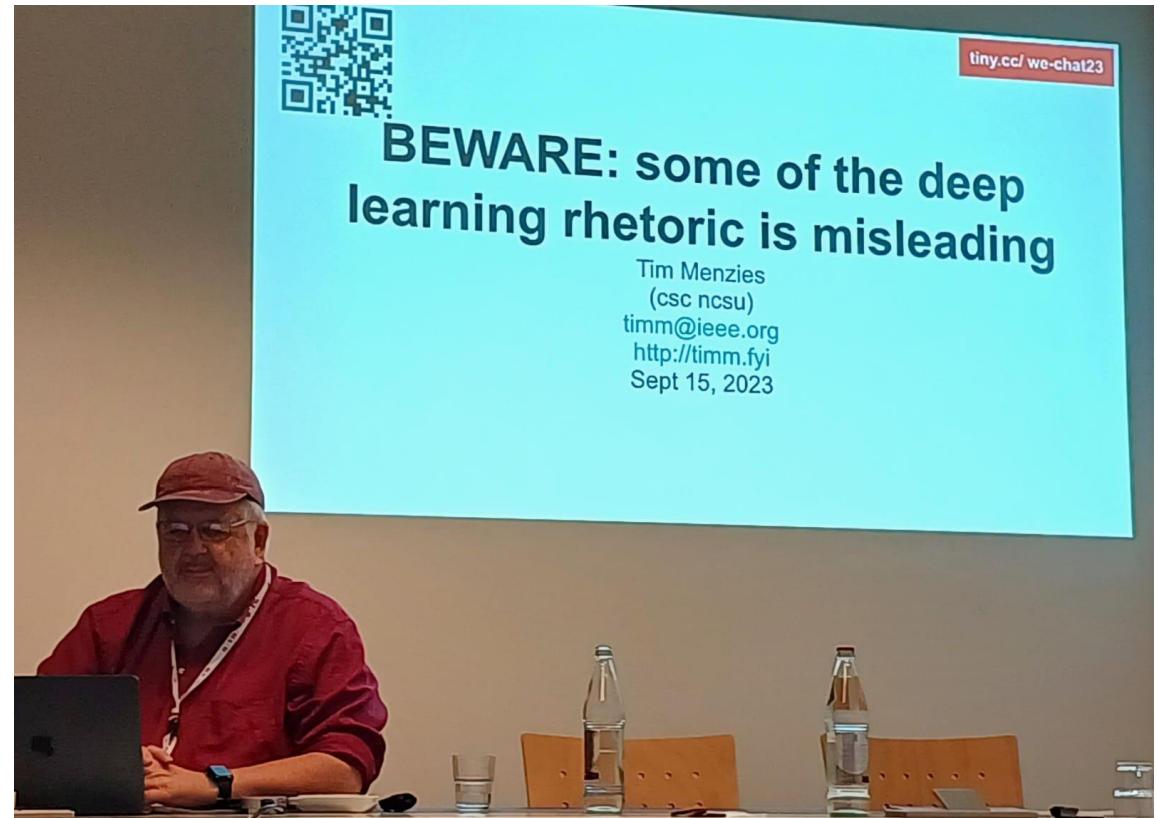
Conference presentation

THE RESEARCH DIRECTIONS IN DL4SE





Decades of AI, my own set of "lessons learned"						
Lessons	ICSE	TSE	FSE	EMSE	IEEE sware	ASE conf
SE data compresses , a lot, without lost of signal (e.g. active learning, random projections, instance+feature+range selection)	2021	2007,2021, 2020,2023			2007	
Modeling vastly easier in compressed space (e.g. need fewer labels, transfer learning, active learning, semi-supervised learning)		2013,2021		2018, 2020		2007, 2021
Optimization easier in compressed space $O(N \log N)$		2015, 2019,2020		2020		2013
If we only share compressed data, then most data private	2012,2015	2013				
If we generate from compressed data/models, can improve (e.g.) explanation, fairness, etc			2021 2020		2023	



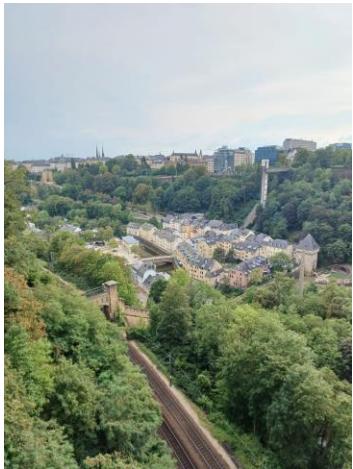
Let's move forward
to see some pics
from the trip :
People and Places!

End of my ~~tale~~ tail



• Questions? Comments?

Luxembourg City Walking Tour From ASE



Cruise & Gala Dinner from ASE 2023:

Rested and relaxed as we were cruising along the river, past the ships and barges and the wine-producing villages on both banks of the Moselle, in the heart of Europe.

Luxembourg – Germany – France and Back to Luxembourg...

Memorable and Pleasant – People and Places!!





And went to Paris!



Conference presentation



Paris Was
memorable,
motivates to
publish more to
travel again! 😊



Airport Bus



Car



Bullet Train



Cycle



Tram



Flight

Cruise



Bus



Conference presentation

Europe
was
awesome:
Smile.
Explore
and Travel





Immense gratitude to Prof. Karthik for your invaluable guidance and mentorship that steered this project towards success. A big thank you to Arya, whose coding expertise and collaborative efforts were instrumental in bringing our research to fruition.

- From Shubham Kulkarni, ASE 2023, European Convention Center, Luxembourg.



Conference presentation

Thank You!!

**Shubham Kulkarni
MS CSE
SERC, IIIT
Hyderabad India**



Agenda

Meet the presenter

Background

Breakout questions

Hypothesis

Making great products

Comparing Contoso to the competition

History

Key takeaways

Tips for businesses

Call to action



Background

As a high school student, Mirjam's interest in American Sign Language sparked her desire to study communication, which led to her focus on technology and how it enhances communication.

“If you can film an idea in your mind, follow that film idea shot for shot, scene for scene, that idea is worth making.”

Craig Mapp



Hypothesis

Virtual communication will be more common in the next 5 years.

Contoso's technology helps people and businesses communicate efficiently through a virtual world.

Call to action



Contoso is the best technology platform on the market



Download Contoso to try it out and provide feedback



Making great products



Take risks to lead in product innovation



Listen to customers and learn what they need



Test with customers until you get the product right

Comparing Contoso to the competition

Contoso empowers businesses to communicate efficiently with technology that's simple and easy to use.

The competition also offers similar technology, but their tools aren't user friendly.



History

Launch Contoso

Feb 20XX

Procure investments

Oct 20XX

Start development

Nov 20XX

JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC JAN FEB MAR APR MAY JUN JUL AUG SEP OCT NOV DEC

20XX

Request funding

May 20XX

Hire designers

Feb 20XX

Open offices

Jul 20XX

Key takeaways



Virtual communication
is the future



Contoso is a leader in
virtual communication



Contoso helps
businesses communicate
effectively



Contoso is free to try
and easy to use



Tips for businesses



Take risks to lead in product innovation



Listen to customers and learn what they need



Test with customers until you get the product right

Thank you

Mirjam Nilsson

206-555-0146

mirjam@contoso.com

www.contoso.com

