Software Engineering for Artificial Intelligence

Introduction

Attendance via Zoom

☺ Let's try to make this a great experience for all of us:

✓ Please check your setup before the meeting. We start all calls 10 minutes early, where you can do so.

🕐 Please connect before the meeting starts.

👤 Please join using your full name. If you use a nickname or pseudonym, tell the advisors (needed for grading).

🎤 We encourage you to use a microphone and a camera: It improves the overall experience in interactive parts.

🔇 Please mute your microphone when not speaking

Agenda

- Motivation

- This Seminar

- Seminar Structure & Grading

- Schedule

- Registration

Advisors & Contact

Any questions, suggestions, interested in research or collaborations?
Talk to us or drop a mail!

Prof. Mira Mezini

mezini@st.informatik.tu-darmstadt.de

Dr. Krishna Narasimhan

kri.nara@st.informatik.tu-darmstadt.de

Daniel Sokolowski, MSc

sokolowski@cs.tu-darmstadt.de

Most prob faster reply…

Most prob faster reply…

Non-SE4AI things
(Research, thesis, jobs,
general advice)

SE4AI: Topic related
questions

SE4AI: Organizational questions,
enrollment and submissions

Motivation

Imagine: We build together *AcaWhooo!* a „Google Translate" for scientific text.


What are challenges apart from accuracy?

From Data Science to Production

- A data scientist can build our program, but…

  - They are used to fixed datasets and focus heavily on accuracy.

  - They prototype, often using Jupyter notebooks, etc.

  - They are experts in modeling and feature egineering, but stability, size, updateability and other aspects, which are important in production, do mostly not matter.

- A software engineer is focused on production grade software

  - Concerned about **many different kinds of product quality**: performance, security, safety, stability, release time, cost, customer satisfaction, maintainability, reliability, scalability, fault tolerance, ...

- **Both worlds need to be brought together!**

From Data Science to Production

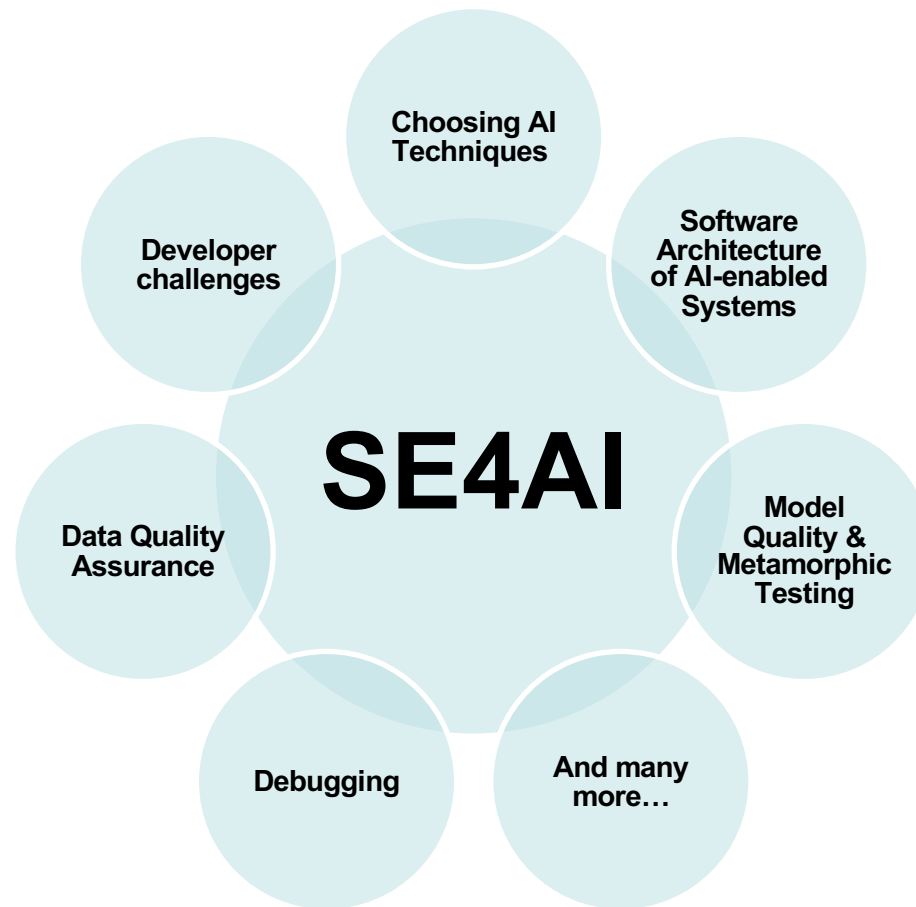
SE

AI
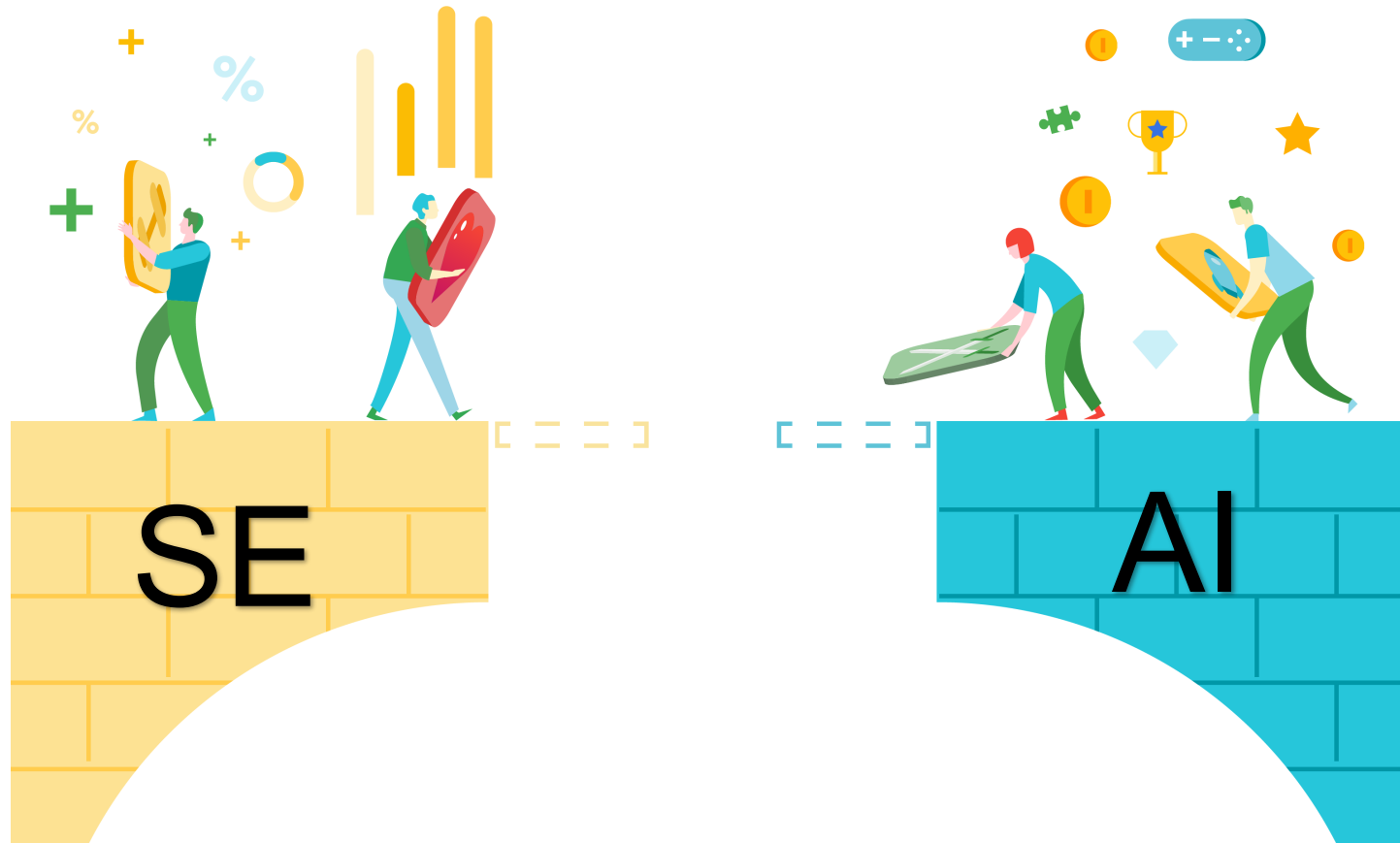
This Seminar Changed!

- Last semester: Overview Seminar

  - Presented various topics in improving SE for AI Systems

  - Gave an overview and some insights, but not very deep

- This semester: **Think Tank Seminar**

  - Interactively critique and analyse one problem in-depth

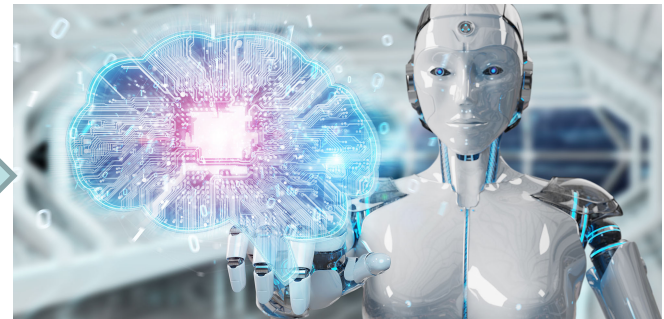 - Develop ideas collectively to improve the state-of-the-art

Topics of SE4AI

Our Focus

DEVELOPER CHALLENGES WHEN BUILDING AI-BASED SOFTWARE

Behind every Intelligent system is a developer struggling with AI libraries

- Exhaustive study to understand developer challenges using ML

- Study of deep-learning bug characteristics, and fix patterns

- Misuse detection for Machine learning Libraries

IOWA STATE UNIVERSITY
Digital Repository

Graduate Theses and Dissertations   Iowa State University Capstones, Theses and Dissertations

2020

Towards understanding the challenges faced by machine learning software developers and enabling automated solutions

Md Johirul Islam
Iowa State University

https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=9156&context=etd

Study of developer challenges using ML – Setup

- Manually analyzed posts in **stack overflow**

- Ranked the posts based on reputation and like/dislike ratio

- Categorized the issues based on various factors:

  - Library (Caffe, Tensorflow, Keras)

  - Stage in the ML Pipeline

  - Lifetime of the issue

- Answered research questions like:

  - Problematic Libraries

  - Difficult stages, and many more…

Study of Deep learning bug characteristics - Setup

- Analyzed bug related posts **stack overflow**

- Included bug-fix commits from **GitHub**

- Studied prevalent type of bugs, root causes and impact among other things

Amimla: Misuse detection for Machine learning Libraries

Combines various sources/techniques to detect misuses motivated by study

Detected misuses with 70% precision and 80% recall

In this Seminar

- We learn about developer challenges facing AI software.

- We learn how to write scientific text collaboratively.

- We will critically analyse a state-of-the-art thesis in this domain.

- We will write a vision paper (4-5 pages) proposing new ideas based on the thesis.
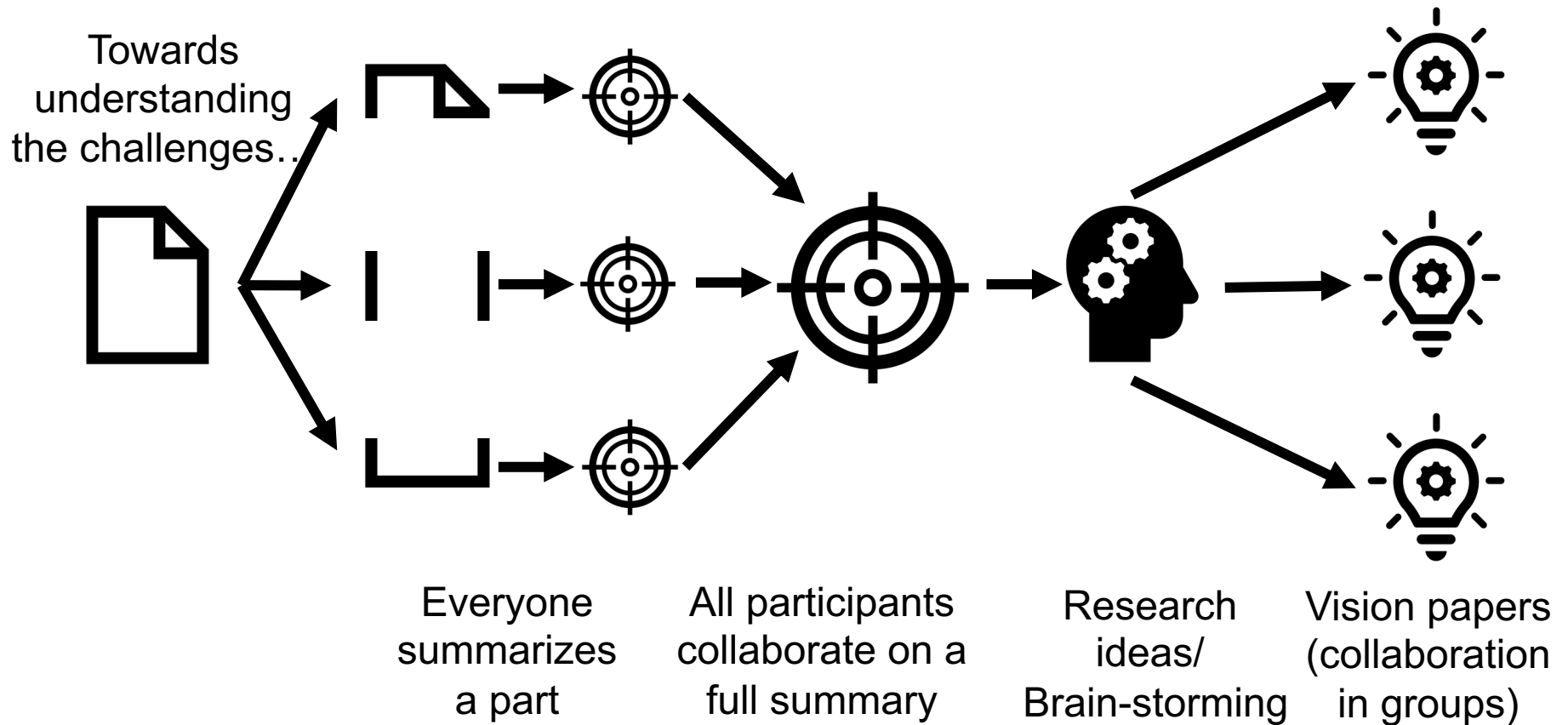
Recommended resources for this Course

- Books

  - Especially: Hulten, Geoff. [Building Intelligent Systems](#) : A Guide to Machine Learning Engineering. Apress, 2018

- Research thesis – Towards understanding the challenges faced by machine learning software developers and enabling automated solutions
  https://lib.dr.iastate.edu/cgi/viewcontent.cgi?article=9156&context=etd

- Blogs

Seminar Overview

Towards understanding the challenges…

Everyone summarizes a part

All participants collaborate on a full summary

Research ideas/ Brain-storming

Vision papers (collaboration in groups)

Schedule

Today	• Kick-off meeting
Nov 5	• Seminar registration
Nov 6	• Registration confirmation
Nov 10	• Chapter Assignment
Dec 6	• Short chapter summary submission
Dec 8	• Chapter summary discussion
Jan 10	• Full summary submission
Jan 12	• Full summary presentation and brainstorm ideas
Jan 19	• Vision paper finalizing
Jan 26, Feb 2	• Workshop
Feb 9	• Last meeting

Grading

- Individual Summary
  - Presentation (5 mins – 2-3 slides) – 10%
  - Submission (1 page + references) – 20%

- Collaborative Summary (2 pages + references) – 20%

- Participation – 10%

- Vision Paper (4-5 pages + references) – 40%
  - 2 pages summary
  - 1-2 pages for the idea
  - 1 page for the impact analysis

Registration

- Send a mail by November 5th to
  sokolowski@cs.tu-darmstadt.de
  - Include why you are interested to participate in this seminar (max 4-5 sentences)
  - If you have related experiences, mention them (courses, projects, …)
- If more than 8 registrations, we will select based on the mail
- We confirm the registration by November 6th via mail
- (Do not forget to register in TUCaN as well)
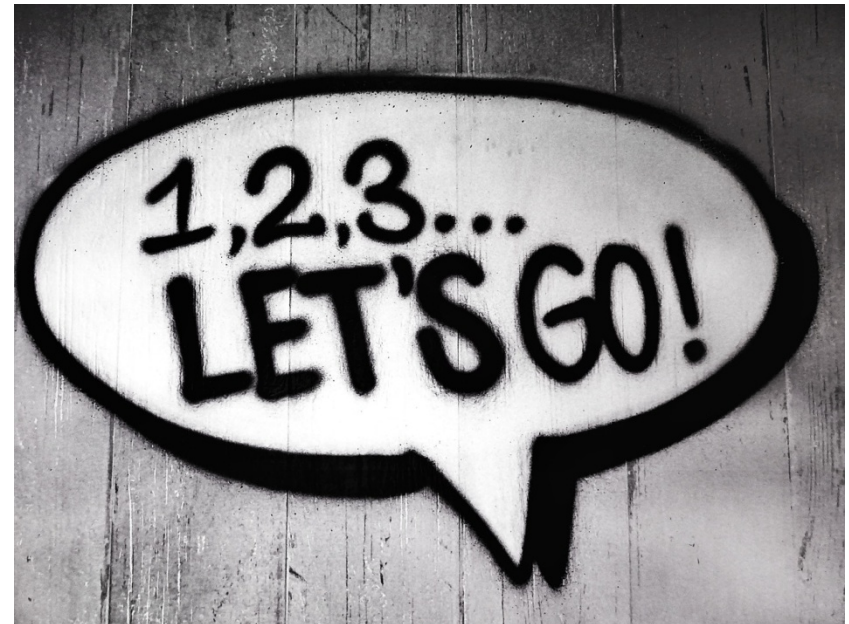
Question & Answers

Acknowledgements & License

- Material Design Icons, by Google under Apache-2.0

- Other images are either by the authors of these slides, attributed where they are used, or licensed under Pixabay or Pexels

- These slides are made available by the authors (Daniel Sokolowski, Krishna Narasimhan) under CC BY 4.0