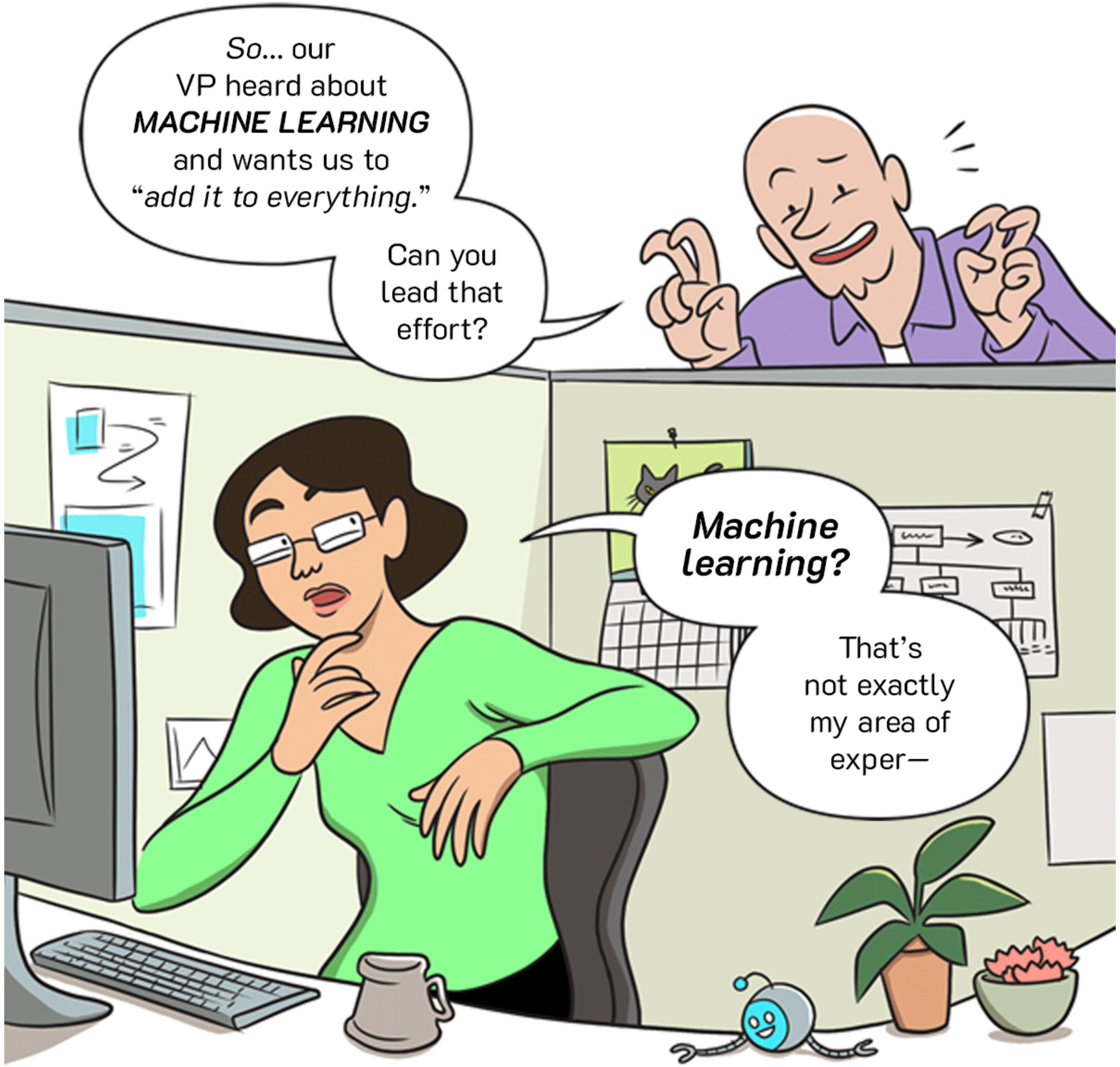

So... our VP heard about MACHINE LEARNING and wants us to "add it to everything."
Can you lead that effort?
Machine learning?
That's not exactly my area of exper—

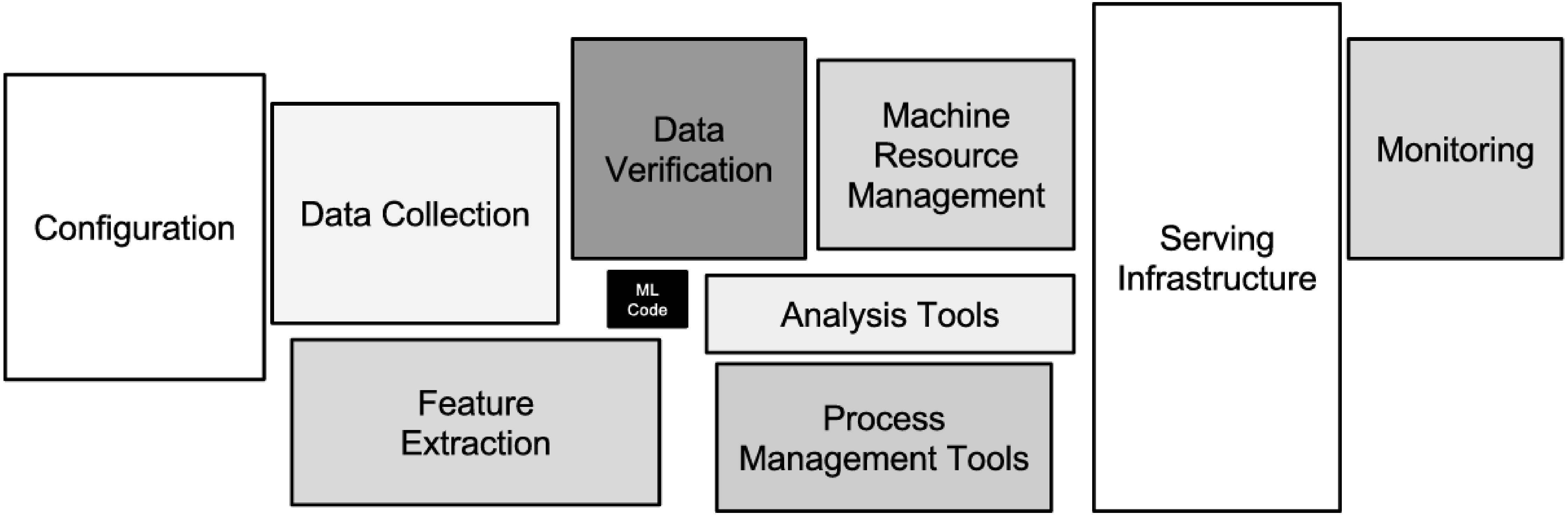
Curtsy: https://cloud.google.com/products/ai/ml-comic-1

"To make great products: Do machine learning like the great engineer you are, not like the great machine learning expert you aren't."

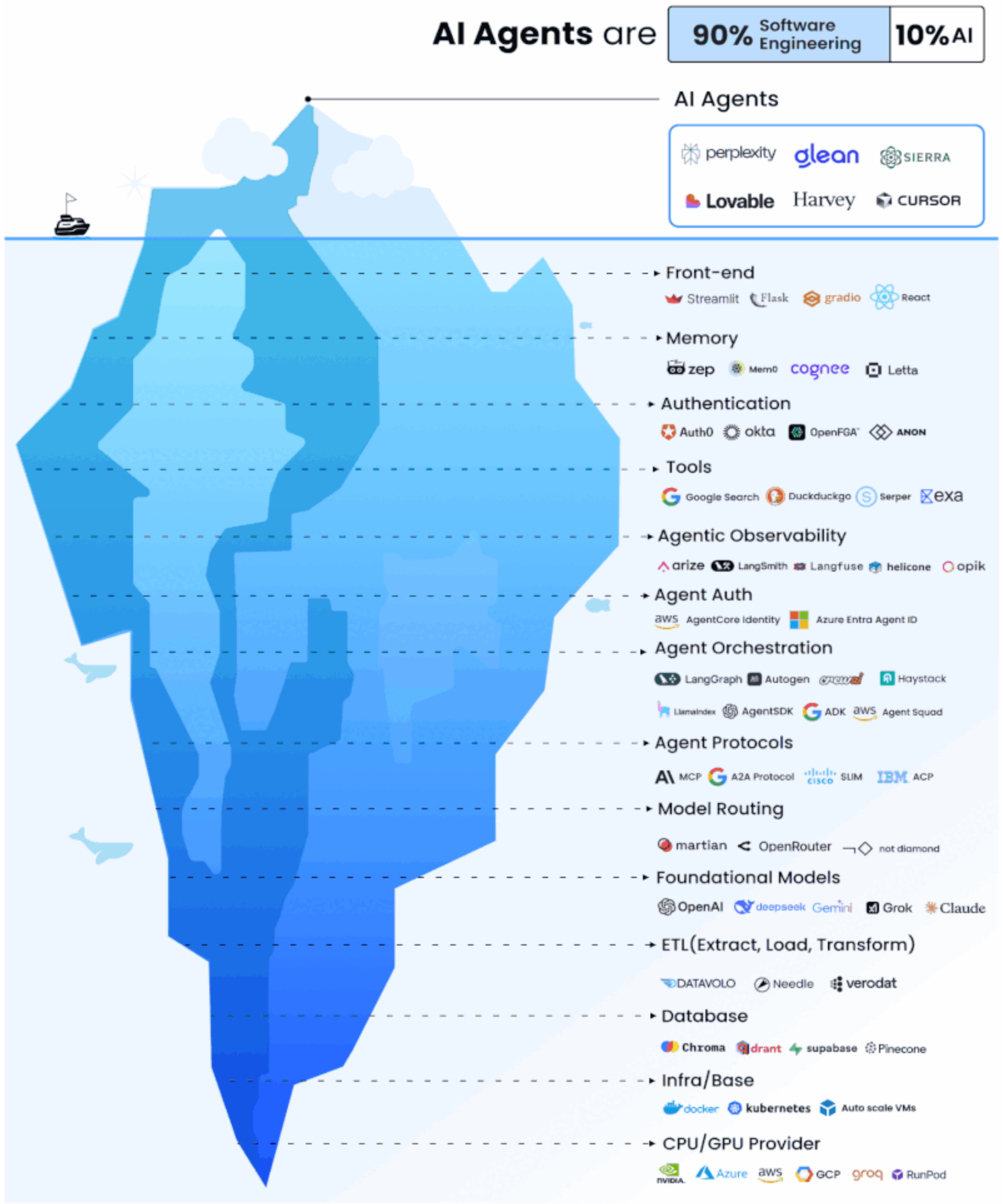
-Google AI Lab

Developing and deploying ML systems are relatively fast and cheap, but maintaining them over time is difficult and expensive.

https://www.youtube.com/watch?v=VfcY0edoSLU



Sculley, David, et al. "Hidden technical debt in machine learning systems." *Advances in neural information processing systems* 28 (2015).



AI Agents are 90% Software Engineering 10% AI

AI Agents

perplexity glean SIERRA
Lovable Harvey CURSOR

Front-end
Streamlit Flask gradio React

Memory
zep Mem0 cognee Letta

Authentication
Auth0 okta OpenFGA ANON

Tools
Google Search Duckduckgo Serper exa

Agentic Observability
arize LangSmith Langfuse helicone opik

Agent Auth
aws AgentCore Identity Azure Entra Agent ID

Agent Orchestration
LangGraph Autogen crewai Haystack
LlamaIndex AgentSDK ADK aws Agent Squad

Agent Protocols
MCP A2A Protocol cisco SLIM IBM ACP

Model Routing
martian OpenRouter not diamond

Foundational Models
OpenAI deepseek Gemini Grok Claude

ETL(Extract, Load, Transform)
DATAVOLO Needle verodat

Database
Chroma Qdrant supabase Pinecone

Infra/Base
docker kubernetes Auto scale VMs

CPU/GPU Provider
NVIDIA Azure aws GCP groq RunPod

Source: LinkedIn post

ENSF 619
Software Engineering for ML-based Systems

Overview and Admin Notes

Fall 2025

Gouri Ginde

Welcome

What you should know (Prerequisites) : Software Engineering basics (SDLC, Waterfall, Spiral, Scrum, Agile methodologies) beginner level, Knowledge of ML

Agenda for today

- Introduction to ENSF 619
- Course outline
- Assignment 1
- QnA

Gouri

Asst. Prof. Gouri Ginde
Dept. of Electrical and Software Engineering
SHAKTI (**S**oftware **H**ub for **A**nalyti**K**s, **T**echnology and **I**nnovation) Lab

- Virtual/in-person office hours: by appointment made by email
- Email: gouri.deshpande@ucalgary.ca
- Office: ICT 254
- Contact number (Office): 403.220.7431
- Website: https://profiles.ucalgary.ca/gouri-ginde-deshpande

NOTE: Include the following in the subject of your email [ENSF 619]
E.g. *[ENSF 619] Need clarification*
[ENSF 619] Meeting request

Gouri outside of classroom

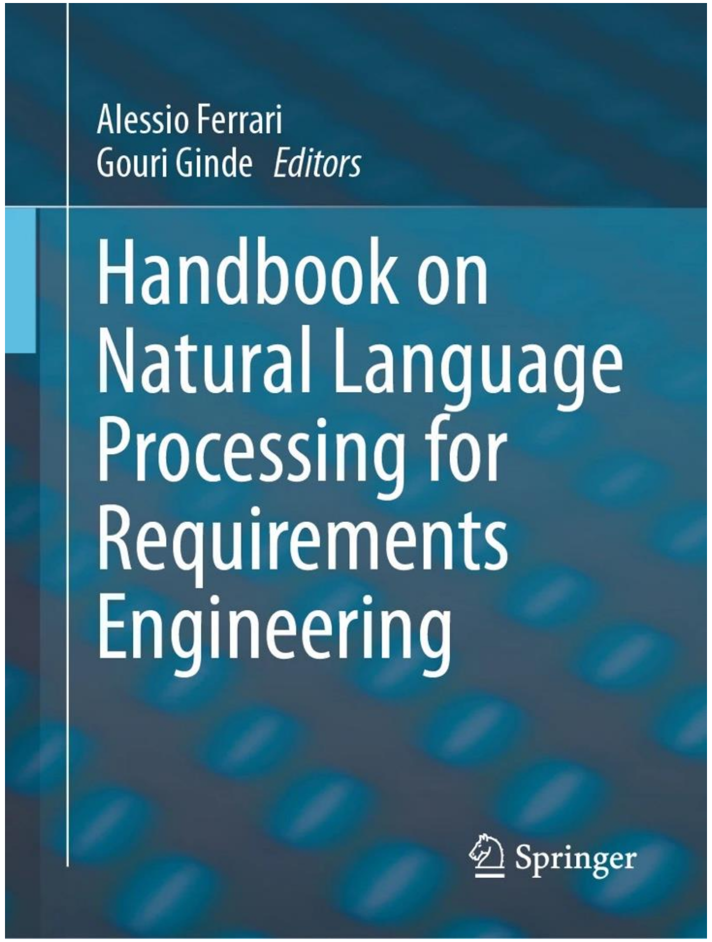
I’m a researcher in the field of Software Engineering (SE)

I specialize in

- Requirements Engineering (RE): Dependency extraction, Elicitation, Open Source Software data mining, Data Analytics
- SE for Healthcare
- Applied NLP and ML, RE for ML and Responsible AI for RE
- Scientometrics

I also dabble in Electronic Publishing: Experimentation, Design, Writing, etc.

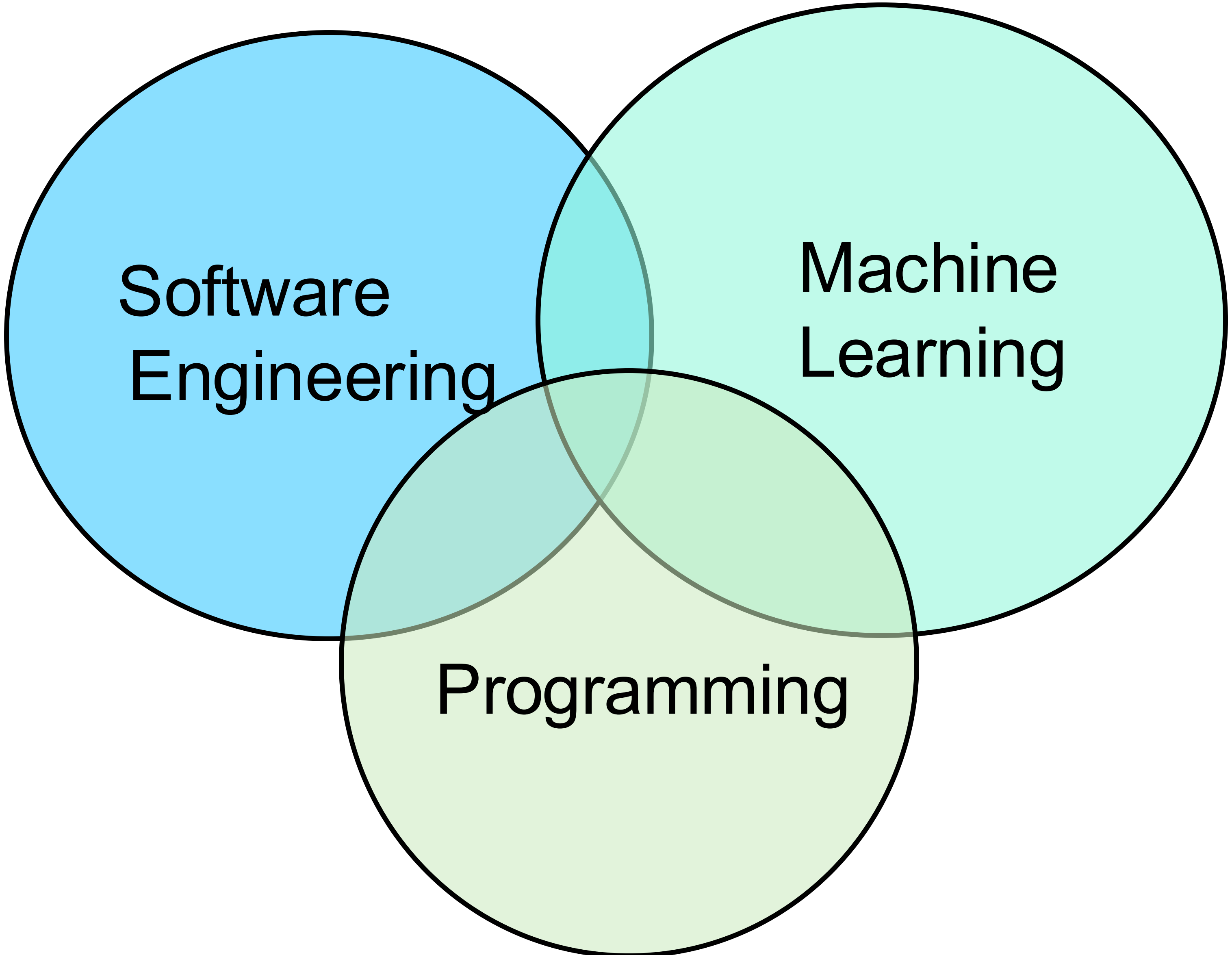
Gouri outside of university


Alessio Ferrari
Gouri Ginde Editors
Handbook on Natural Language Processing for Requirements Engineering
Springer

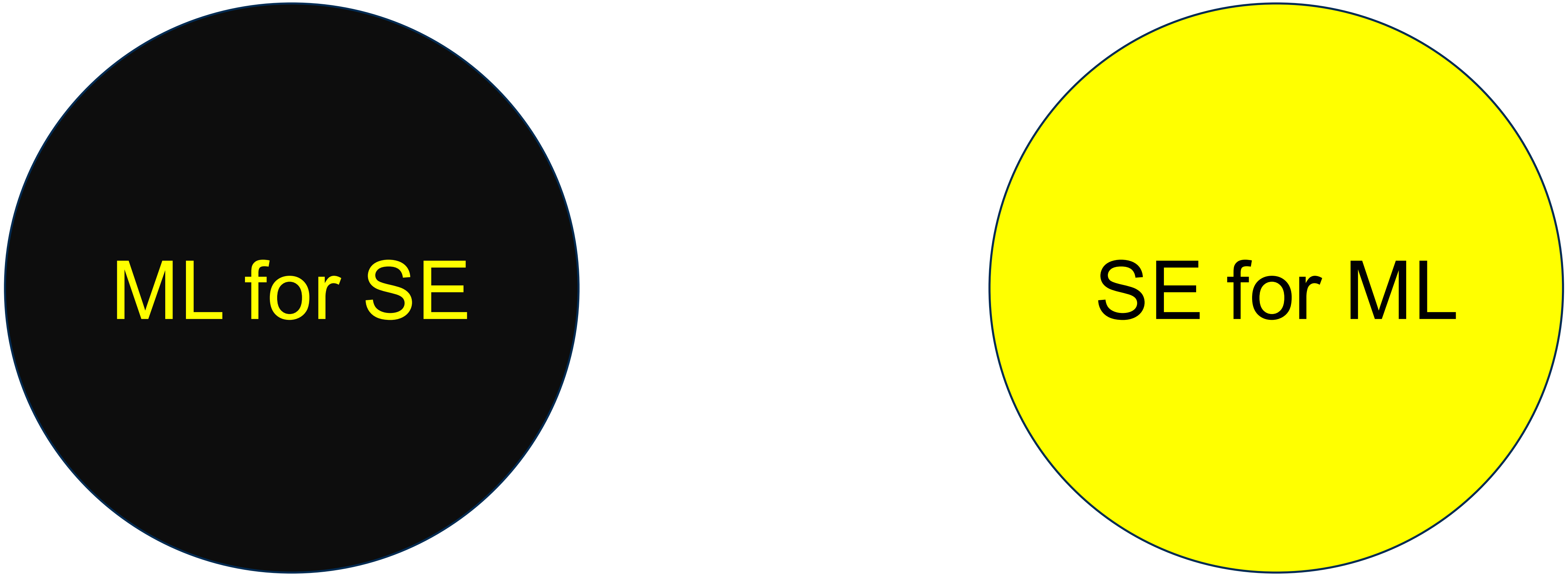
- I do gardening, painting, sculpting, swimming, yoga, learning winter sports such as downhill skiing
- I love programming and debugging!
- I worked in IT for over a decade as a Software and Sr. Software engineer at Hewlett Packard, India, before doing my PhD here at the Dept. of Computer Science.
- I have two kids.

Recently, I co-edited a Book titled “Handbook on NLP for Requirements Engineering.”

ENSF 619.06: Software Engineering for AI/ML-based Systems



What is the difference?

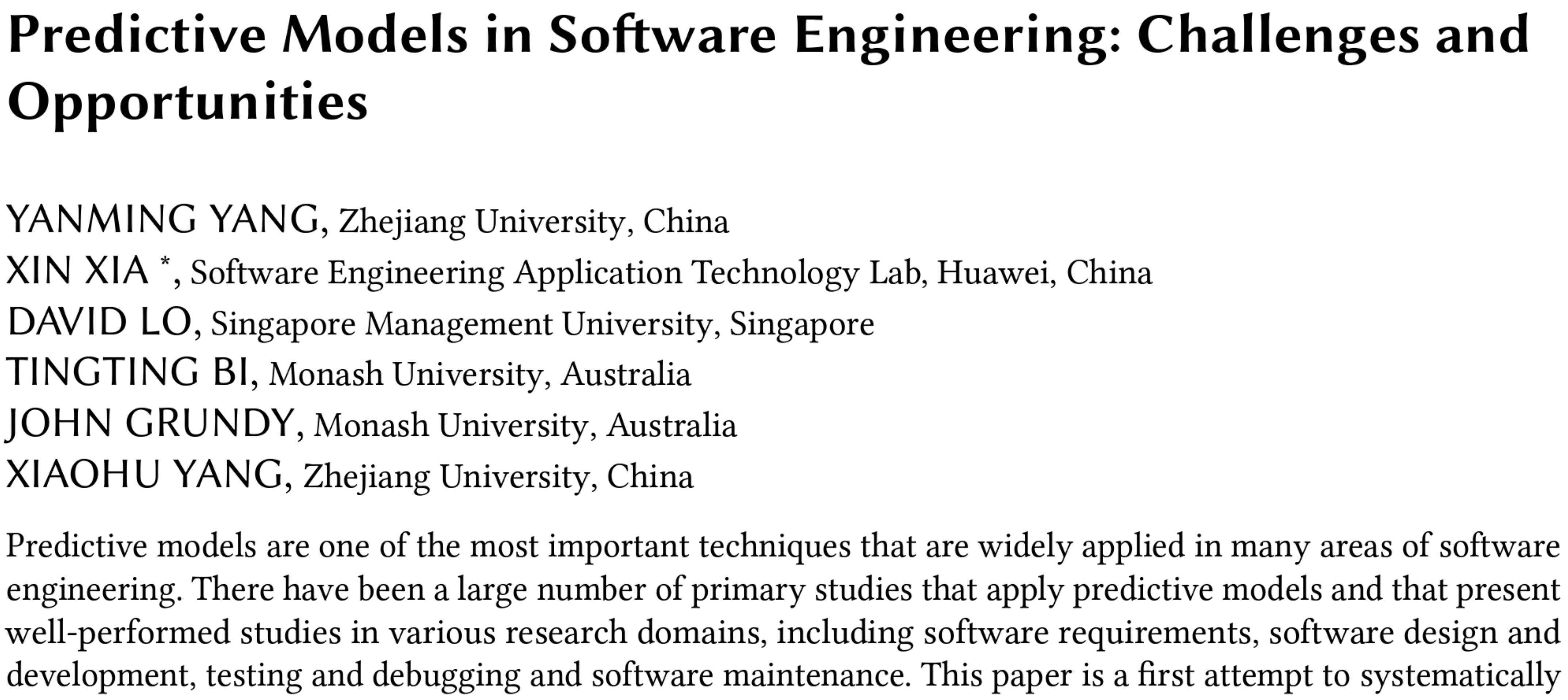


ML for SE

SE for ML

ML for SE

ML for SE



Predictive Models in Software Engineering: Challenges and Opportunities

YANMING YANG, Zhejiang University, China
XIN XIA *, Software Engineering Application Technology Lab, Huawei, China
DAVID LO, Singapore Management University, Singapore
TINGTING BI, Monash University, Australia
JOHN GRUNDY, Monash University, Australia
XIAOHU YANG, Zhejiang University, China

Predictive models are one of the most important techniques that are widely applied in many areas of software engineering. There have been a large number of primary studies that apply predictive models and that present well-performed studies in various research domains, including software requirements, software design and development, testing and debugging and software maintenance. This paper is a first attempt to systematically

Textual data in Software Project

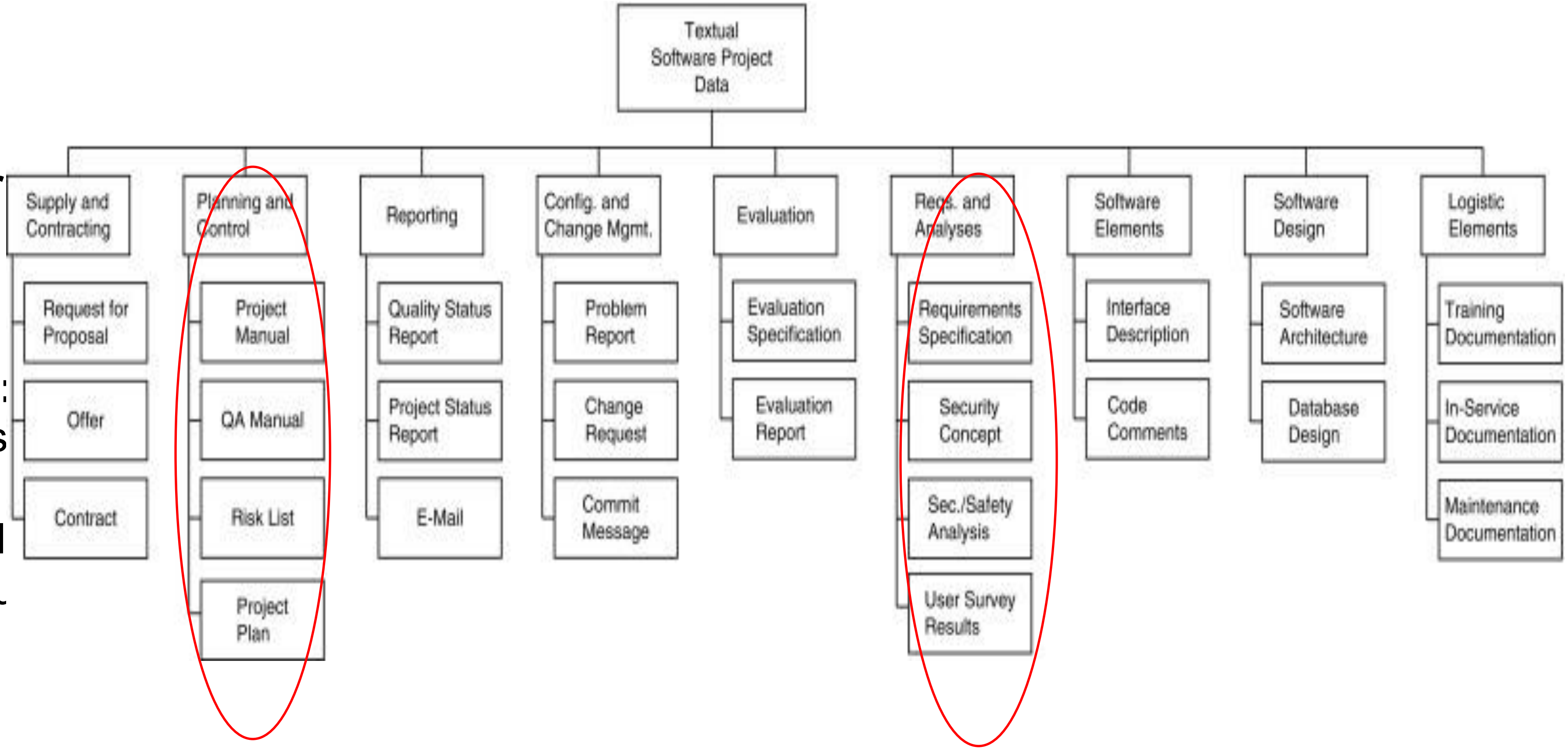
Sour

Also: and commit

mess

Final surveys and

produ



Bird, Christian, Tim Menzies, and Thomas Zimmermann, eds. The art and science of analyzing software data. Elsevier, 2015.

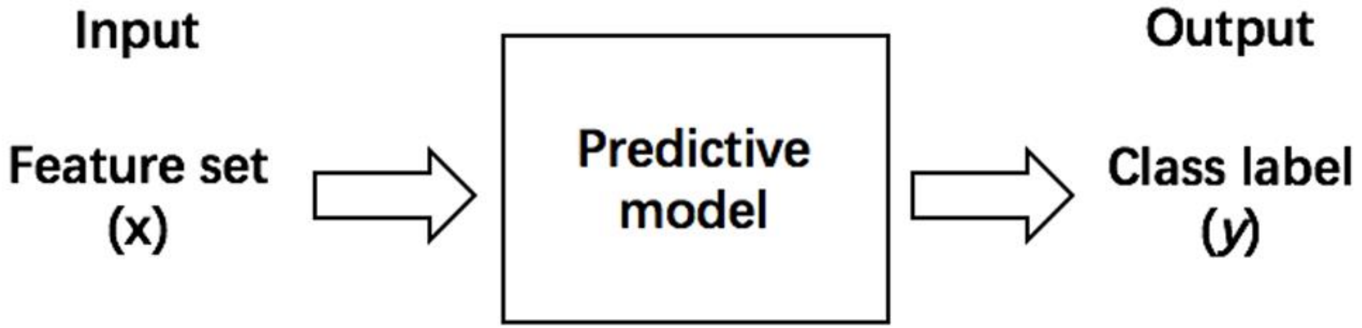


Fig. 1. Predictive model as the task of mapping an input feature set x into its class label y.

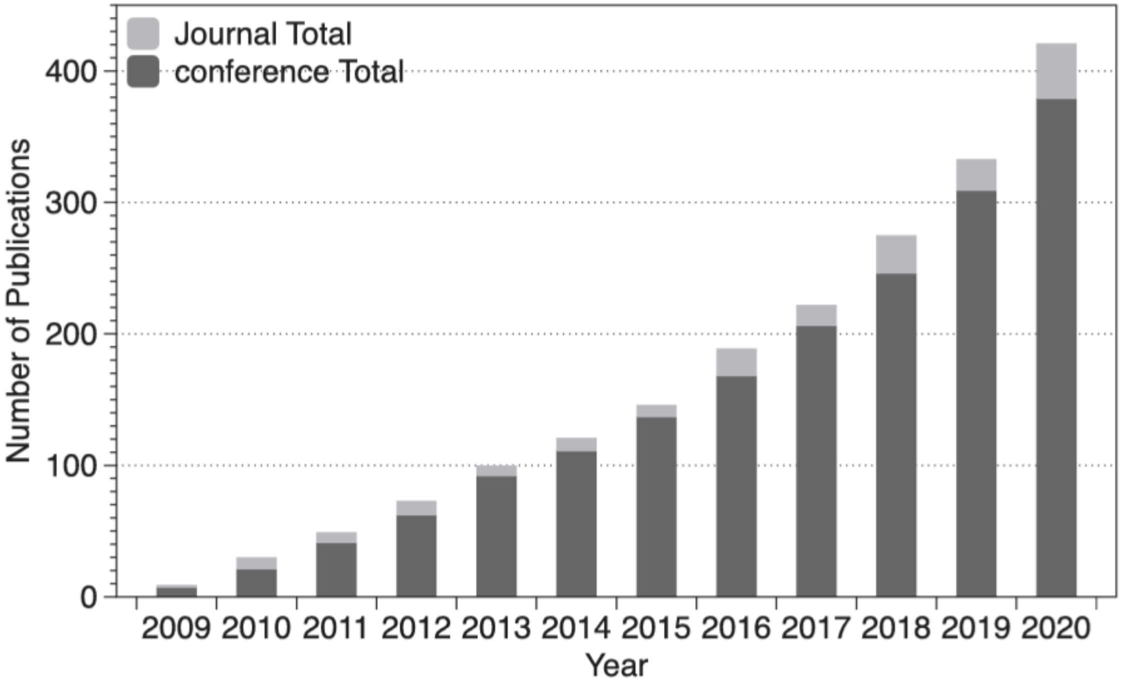


Fig. 4. Venue distribution for primary papers.

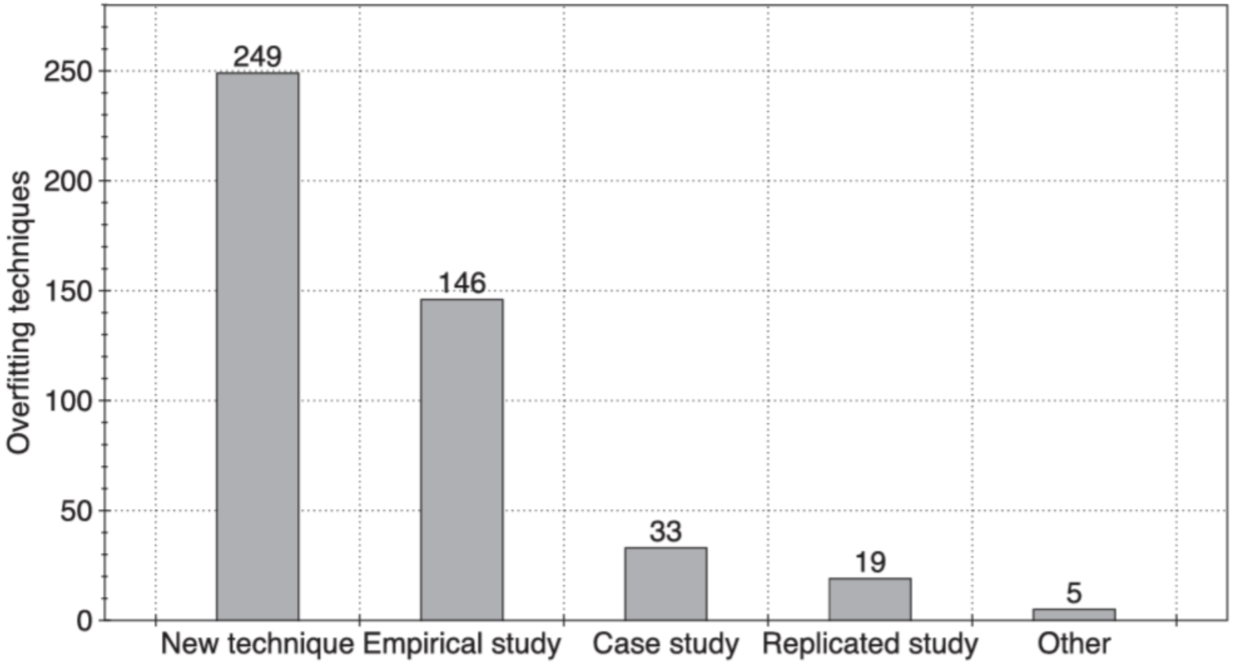


Fig. 5. Type of main contribution.

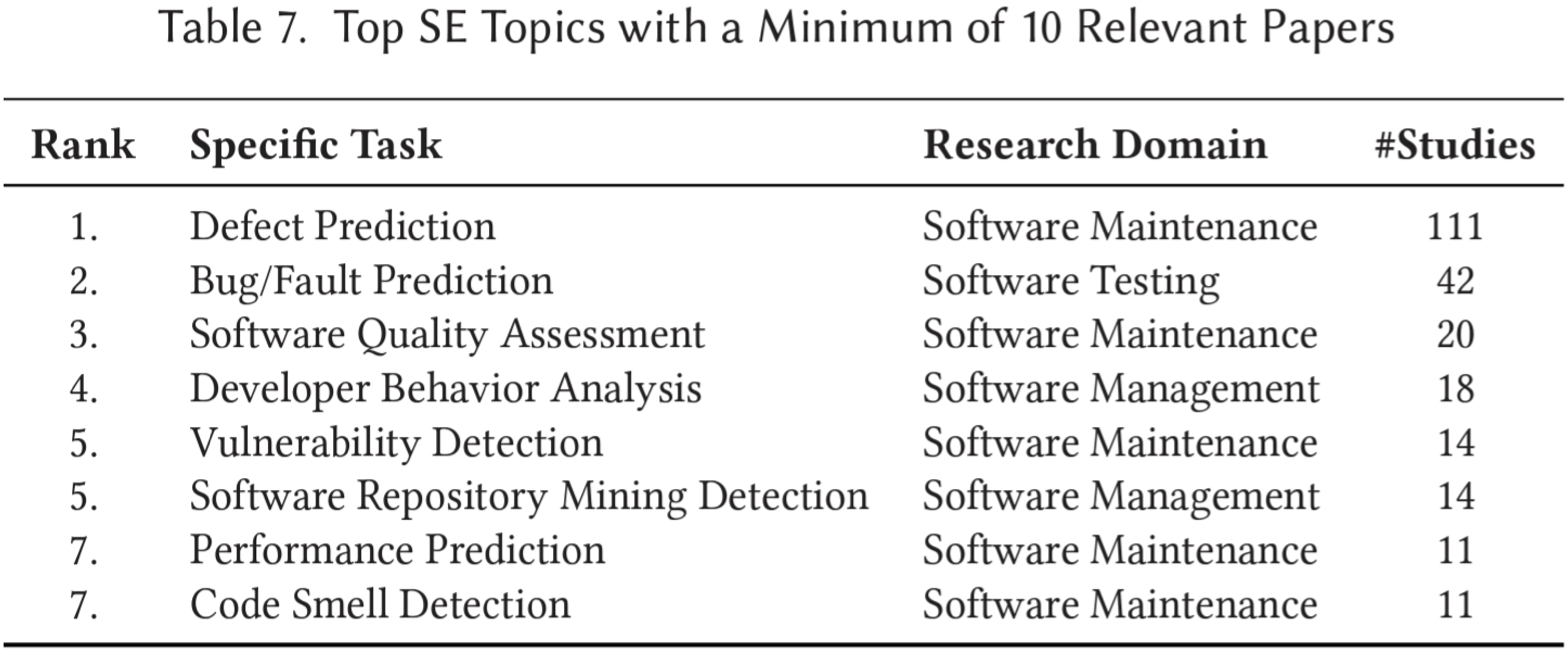
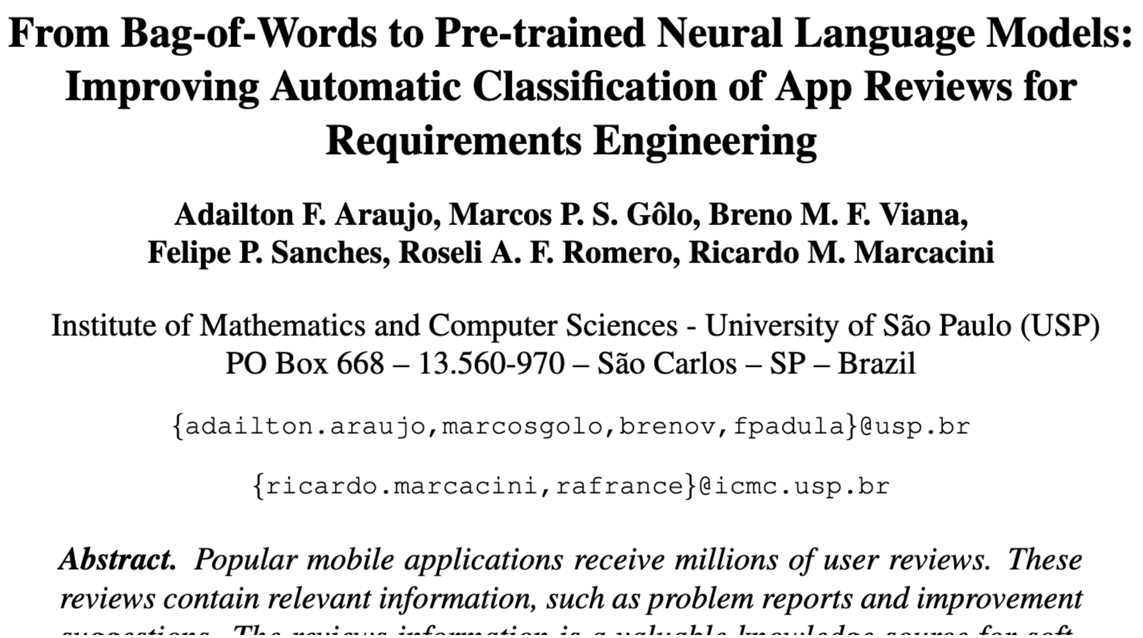


Table 7. Top SE Topics with a Minimum of 10 Relevant Papers

Rank	Specific Task	Research Domain	#Studies
1.	Defect Prediction	Software Maintenance	111
2.	Bug/Fault Prediction	Software Testing	42
3.	Software Quality Assessment	Software Maintenance	20
4.	Developer Behavior Analysis	Software Management	18
5.	Vulnerability Detection	Software Maintenance	14
5.	Software Repository Mining Detection	Software Management	14
7.	Performance Prediction	Software Maintenance	11
7.	Code Smell Detection	Software Maintenance	11

Applied NLP (and ML) in RE: Examples



From Bag-of-Words to Pre-trained Neural Language Models: Improving Automatic Classification of App Reviews for Requirements Engineering

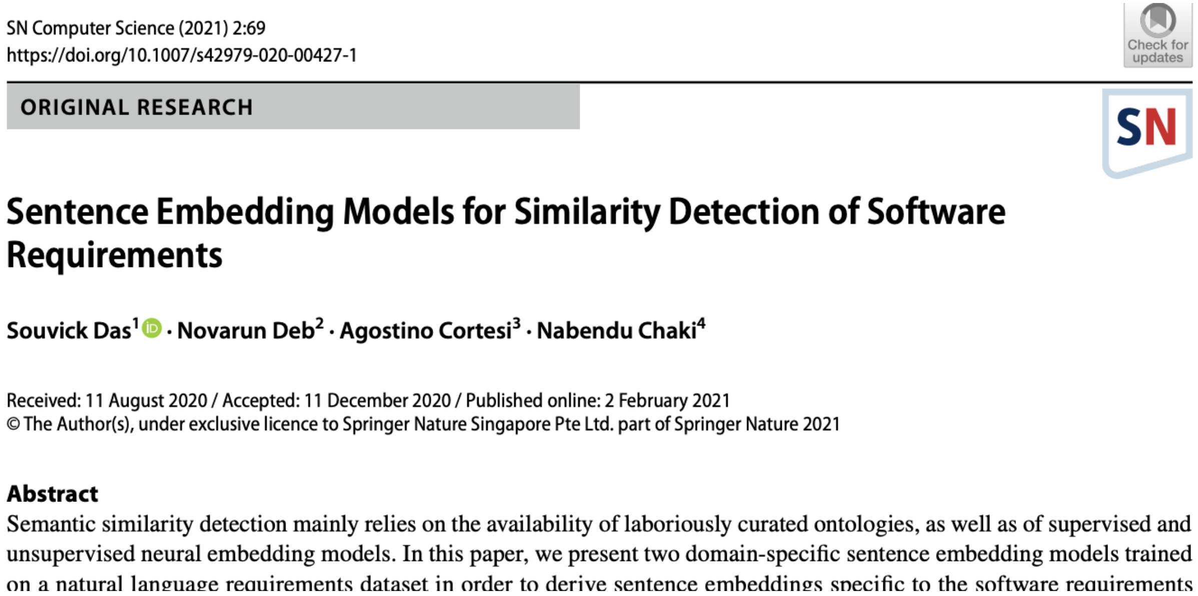
Adailton F. Araujo, Marcos P. S. Gôlo, Breno M. F. Viana,
Felipe P. Sanches, Roseli A. F. Romero, Ricardo M. Marcacini

Institute of Mathematics and Computer Sciences - University of São Paulo (USP)
PO Box 668 – 13.560-970 – São Carlos – SP – Brazil

{adailton.araujo,marcosgolo,brenov,fpadula}@usp.br

{ricardo.marcacini,rafrance}@icmc.usp.br

Abstract. *Popular mobile applications receive millions of user reviews. These reviews contain relevant information, such as problem reports and improvement suggestions. The reviews information is a valuable knowledge source for soft-*



SN Computer Science (2021) 2:69
https://doi.org/10.1007/s42979-020-00427-1

ORIGINAL RESEARCH

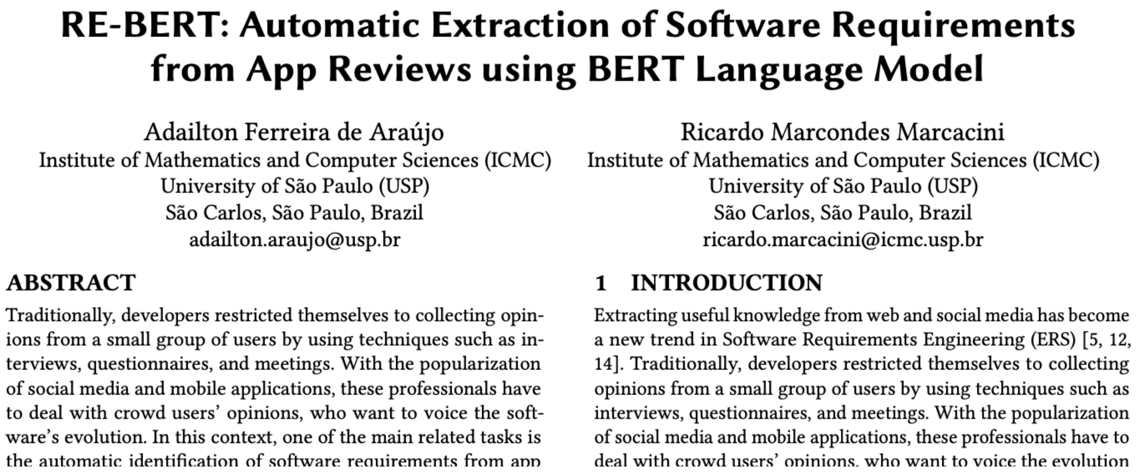
Sentence Embedding Models for Similarity Detection of Software Requirements

Souvick Das[1] · Novarun Deb[2] · Agostino Cortesi[3] · Nabendu Chaki[4]

Received: 11 August 2020 / Accepted: 11 December 2020 / Published online: 2 February 2021

Abstract

Semantic similarity detection mainly relies on the availability of laboriously curated ontologies, as well as of supervised and unsupervised neural embedding models. In this paper, we present two domain-specific sentence embedding models trained on a natural language requirements dataset in order to derive sentence embeddings specific to the software requirements



RE-BERT: Automatic Extraction of Software Requirements from App Reviews using BERT Language Model

Adailton Ferreira de Araújo
Institute of Mathematics and Computer Sciences (ICMC)
University of São Paulo (USP)
São Carlos, São Paulo, Brazil
adailton.araujo@usp.br

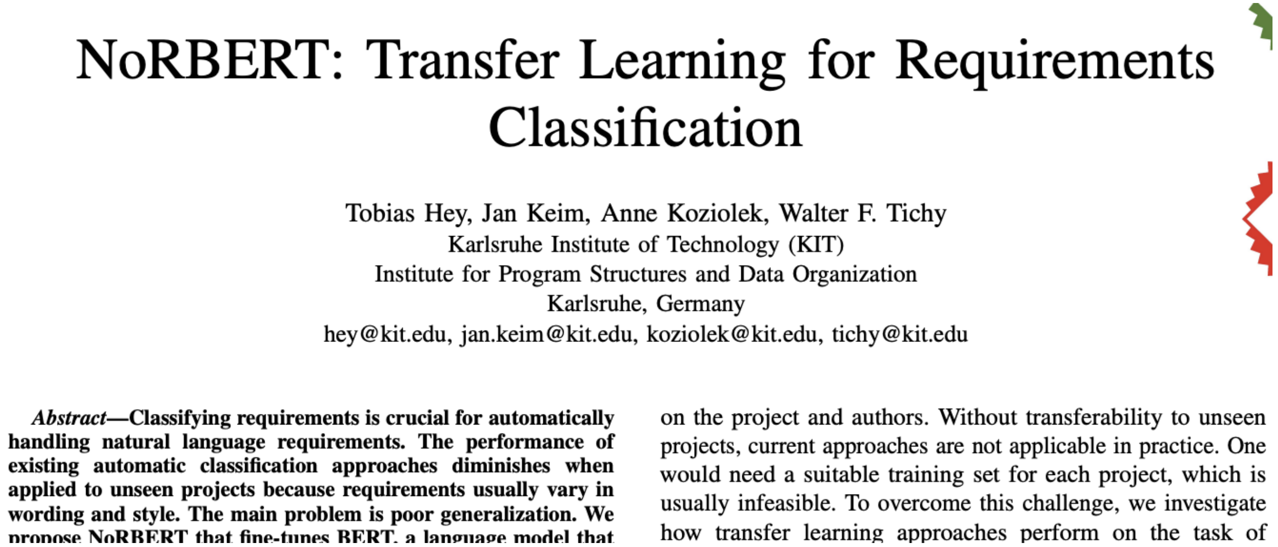
Ricardo Marcondes Marcacini
Institute of Mathematics and Computer Sciences (ICMC)
University of São Paulo (USP)
São Carlos, São Paulo, Brazil
ricardo.marcacini@icmc.usp.br

ABSTRACT

Traditionally, developers restricted themselves to collecting opinions from a small group of users by using techniques such as interviews, questionnaires, and meetings. With the popularization of social media and mobile applications, these professionals have to deal with crowd users' opinions, who want to voice the software's evolution. In this context, one of the main related tasks is the automatic identification of software requirements from app

1 INTRODUCTION

Extracting useful knowledge from web and social media has become a new trend in Software Requirements Engineering (ERS) [5, 12, 14]. Traditionally, developers restricted themselves to collecting opinions from a small group of users by using techniques such as interviews, questionnaires, and meetings. With the popularization of social media and mobile applications, these professionals have to deal with crowd users' opinions, who want to voice the evolution

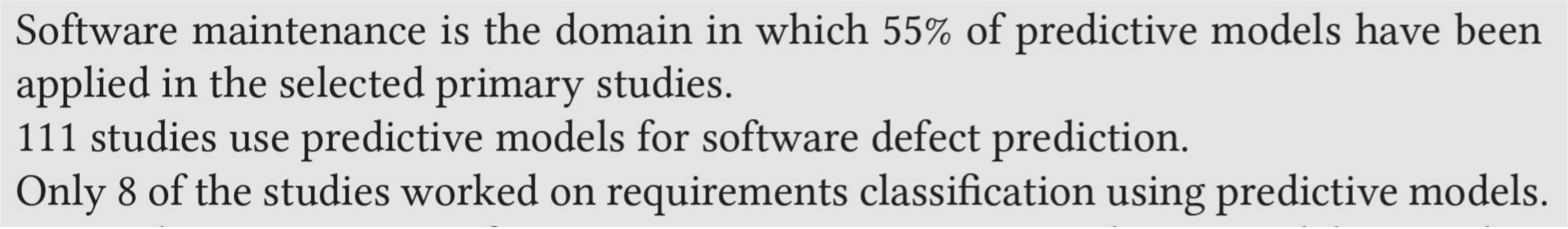


NoRBERT: Transfer Learning for Requirements Classification

Tobias Hey, Jan Keim, Anne Koziolek, Walter F. Tichy
Karlsruhe Institute of Technology (KIT)
Institute for Program Structures and Data Organization
Karlsruhe, Germany
hey@kit.edu, jan.keim@kit.edu, koziolek@kit.edu, tichy@kit.edu

***Abstract*—Classifying requirements is crucial for automatically handling natural language requirements. The performance of existing automatic classification approaches diminishes when applied to unseen projects because requirements usually vary in wording and style. The main problem is poor generalization. We propose NoRBERT that fine-tunes BERT, a language model that**

on the project and authors. Without transferability to unseen projects, current approaches are not applicable in practice. One would need a suitable training set for each project, which is usually infeasible. To overcome this challenge, we investigate how transfer learning approaches perform on the task of



Software maintenance is the domain in which 55% of predictive models have been applied in the selected primary studies.
111 studies use predictive models for software defect prediction.
Only 8 of the studies worked on requirements classification using predictive models.

What is SE for Machine Learning?

Plagiarism

□ Check http://www.plagiarism.org/

□ Link for the University of Calgary FGS Policy on plagiarism: http://www.ucalgary.ca/pubs/calendar/grad/current/gs-o.html

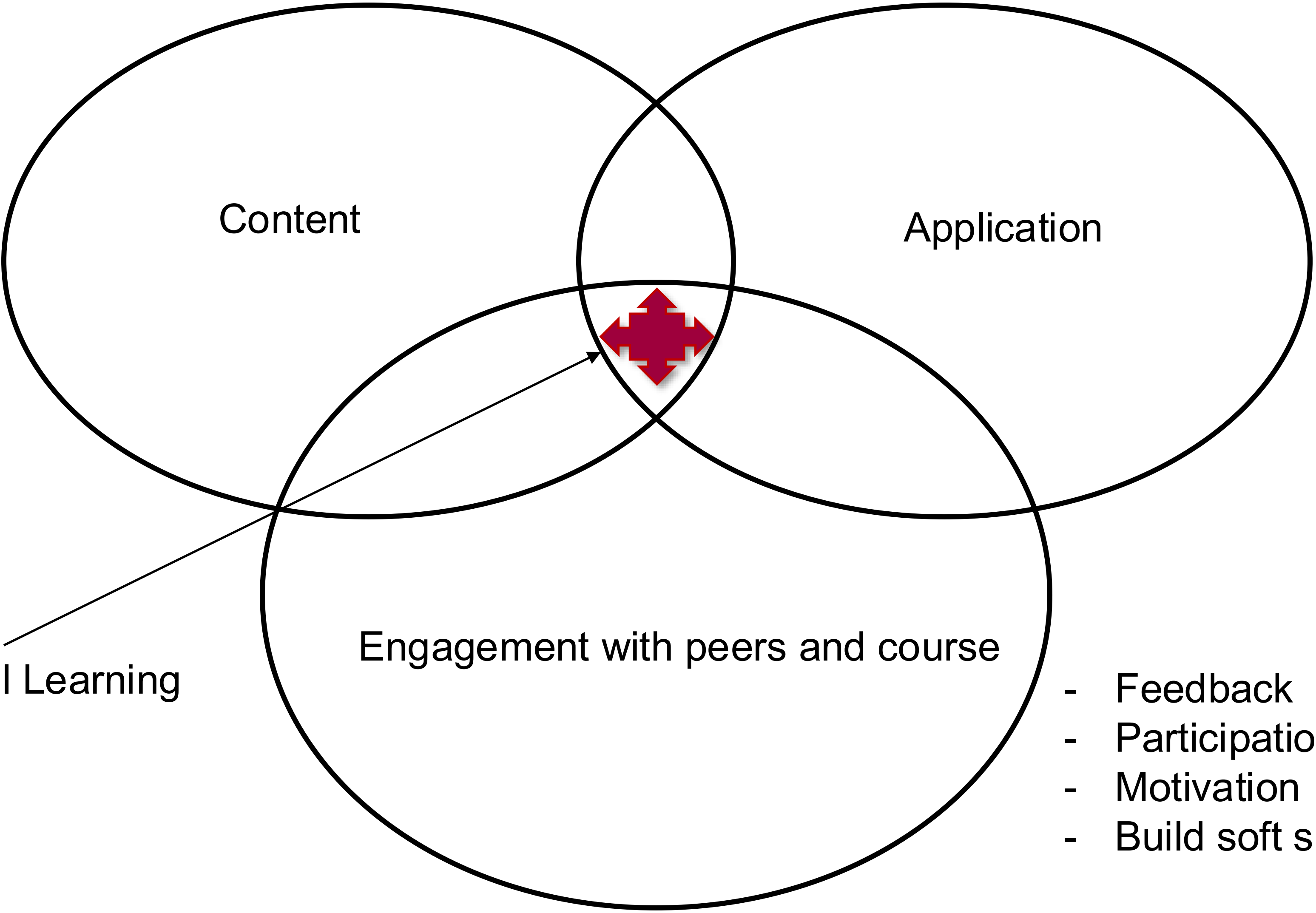
This Photo by Unknown Author is licensed under CC BY-SA-NC

Whaddya mean all my facts are wrong?!?
I copied everything straight off the internet!!
F
Pirillo Fitz
© BLAUGH.COM · FITZ & PIRILLO · SAVE 10% AT GODADDY BY USING THE COUPON CODE "BLAUG

Admin Notes of the course

Course components

- Reading
- Lectures

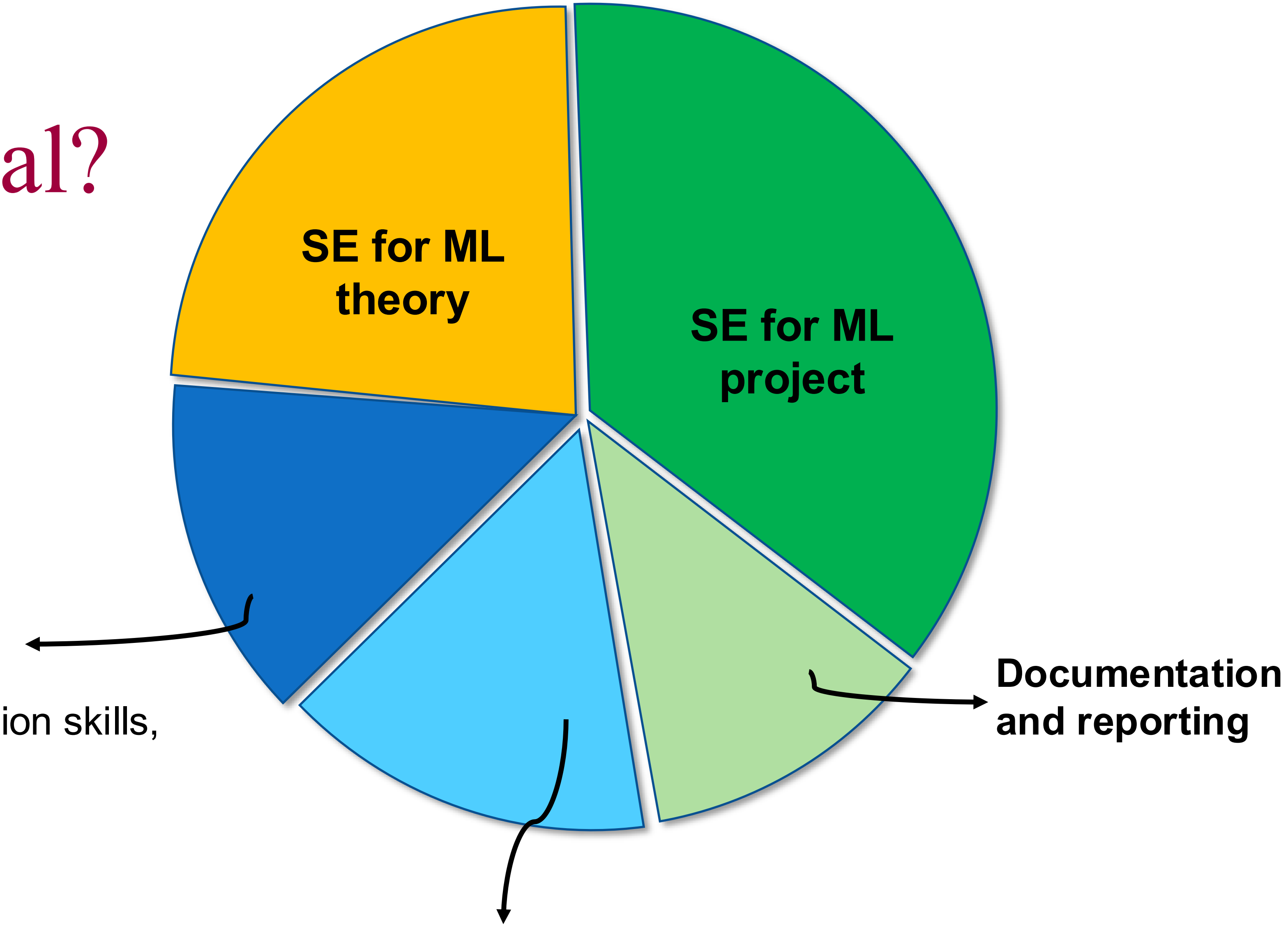


Meaningful Learning

- Assignment
- Group discussion
- One Quiz
- Presentations
- Activities/survey/polls
- Term Project

- Feedback
- Participation
- Motivation
- Build soft skills

What is the goal?



Soft skills
(Teamwork, thinking objectively, presentation skills, accountability)

Documentation and reporting

Reading and understanding scientific literature

Pedagogical approach

We will use an Active Learning approach:

Tuesday’s

- Lecture
- Reading papers and presenting: Group Assignment presentations
- There are no labs, but we will utilize classroom time to strike one-on-one coaching as and when needed.

Office hours – Fridays 2 - 3 pm (on a need basis)

Any additional meetings/one-on-ones and/or discussions can be scheduled during this time

Grading Scheme

NOTE: Attendance mandatory for all presentations and quizzes	Weight	Evaluation	Due dates
Assignment 1 (Notebook submission: ML refresher)	5%	Individual	5th Sept @Noon
Assignment 2 (presentation)	15%	Individual	19th Sept @Noon
Quiz	5%	Individual	7th Oct, In class
Assignment 3 (presentation)	15%	Individual	24th Oct @Noon
Term Project Problem scoping, modelling, problem definition	10%	Group	24th Sept
Course project (Problem, solution and results/demo)	20%	Group	21st Oct @Noon 18th Nov @Noon
Term project report (end of semester)	20%	Group	11th Dec @Noon
Participation	10%	Individual	Over the semester
Total	100%		

Observe that 50% of final grades are for group/team-based performance and 50% for individual performance

** A 2% bonus mark will be awarded to the top projects (based on collective evaluation from peers and instructor) **

** This is subject to change

Term project (50%)

- The term project is a group-based component (group of 2 in each team)

- There are 4 parts to it.

a) Project scoping: 2 pages proposal submission: 10%

b) Problem, solution approach and preliminary results presentation: 10%

c) Final results, demo & presentation: 10%

d) Project report: 20%

- The instructor and peers will evaluate in-class presentations.

Participation (10%)

I track participation in every class.

Expectations:

- Active Engagement in Discussions

Thoughtful and relevant comments during class discussions, asking questions that **demonstrate a deep engagement** with the material, and responding constructively to the ideas of others. **Remember, it's the quality of your contributions that will be valued over quantity.**

- Collaboration and Respect

Working effectively and respectfully with peers during group activities or discussions. This also involves listening to others, considering different perspectives, and contributing positively to the collaborative learning environment.

Textbook

There is no required textbook for this course. However, I will be sharing the scholarly articles and content as and when needed.

You can refer to the following two books for additional material

- *Wilson, Ben. Machine Learning Engineering in Action. Manning Publications, 2022.*

- *Software Engineering: A Practitioner's Approach by Roger S. Pressman*

RESEARCH PROJECT

- Covers study of literature and doing some own (independent) work
- Developing an AI/ML-based tool/solution following SE fundamentals
- Group work is encouraged, but not mandatory
- Option to enhance research into joint publication *(I shall be one of the co-authors in on the papers stemming from this coursework)*
- Inspiration for project themes (you can bring your project from your ongoing work with your supervisor)

on

A. Begel, T. Zimmermann, "Analyze This! 145 Questions for Data Scientists in Software Engineering",
In: Proceedings of the 36th International Conference on Software Engineering (ICSE 2014).

Course software

- For architecture diagrams, we recommend you use MS-Visio (from onthehub)
- But you can use others, such as Dia, or …
- Document editor (any of the following)
  - MS-word
  - Or LateX (recommended)
- Presentations
  - OpenOffice ppt
  - Or MS PowerPoint

More specific details will be provided later.

All the documents should be submitted in PDF format in D2L

Individual introductions
LEGO game

Time for a survey!
Please go to D2L

Any other questions?

The “Best Audience Ever” clause

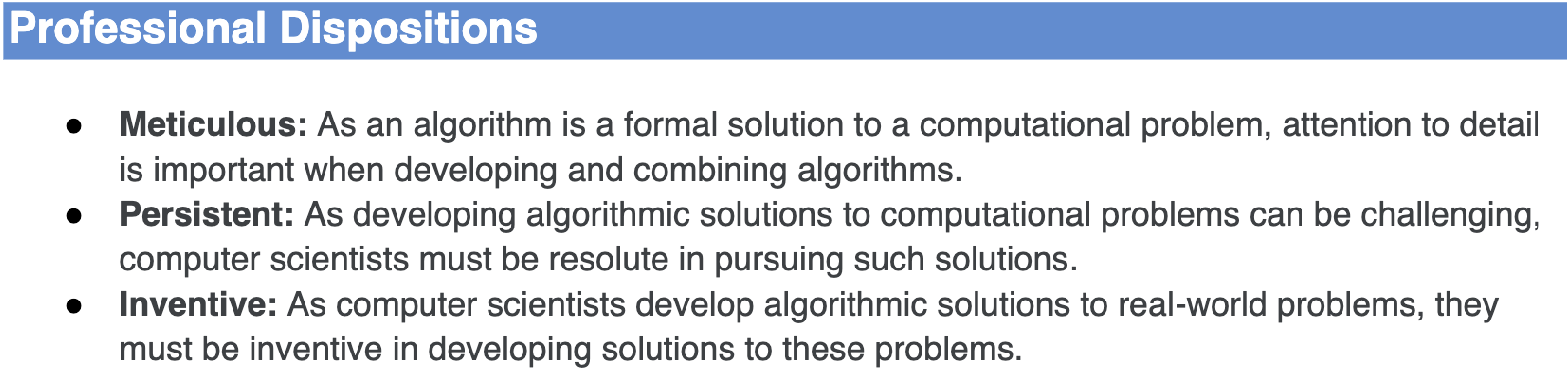
A positive atmosphere includes students who are

Attentive

Supportive

Smiling!

Focus is to also develop



Professional Dispositions

- **Meticulous:** As an algorithm is a formal solution to a computational problem, attention to detail is important when developing and combining algorithms.
- **Persistent:** As developing algorithmic solutions to computational problems can be challenging, computer scientists must be resolute in pursuing such solutions.
- **Inventive:** As computer scientists develop algorithmic solutions to real-world problems, they must be inventive in developing solutions to these problems.

Assignment 1 Kickoff

ENSF 619
Software Engineering for ML-based Systems

Overview and Admin Notes

Fall 2025

Gouri Ginde