Sebis Day 2023

Prof. Florian Matthes, 29.06.2023

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Time	Presentation	Speaker
16:00	Introduction and overview of current research	Prof. Dr. Florian Matthes
16:05	**Recent Work in Blockchain-Based System Engineering** Adoption of decentralized identities. Experimenting with Tezos and the GAIA-X Ecosystem. Detecting NFT Wash Trades on Ethereum and Solana. An exciting collaboration with the Algorand Foundation.	Burak Öz, Felix Hoops
16:15	**Metric-Driven Large Agile Organizations** We present our experience applying and improving a method for supporting the selection and operation of metrics in large-scale agile software development at SAP and an outlook on future work.	Franziska Tobisch, Pascal Philipp
16:25	**CreateData4AI** We present a novel approach to help domain experts efficiently annotate large textual training data for NLP.	Stephen Meisenbacher, Tim Schopf
16:35	**Engineering Conversational Interfaces** We provide an overview of our ongoing industry projects with SAP, Springer Nature, Alpha KI, in domains of human resources, scientific publishing, and healthcare.	Phillip Schneider, Anum Afzal, Juraj Vladika
16:45	**Supporting the Adoption of Privacy-Enhancing Technologies** We present an approach and learning materials to support managers, legal and technical experts in the collaborative adoption of privacy-enhancing technologies in their organization.	Alexandra Klymenko
17:00	**Poster Session and Stammtisch** We provide food and beverages and give you the opportunity to network and talk with all our research assistants about their research projects.	All participants
19:00	**End of the event**	

Sebis Research Areas & Current Research Projects

Next-Generation IT Governance

- Scaled Agility in Large IT Organizations
- Interorganizational EAM
- Enterprise Architecture Discovery at Runtime
- Autonomous Company
- Adopting Metrics in Large-Scale Agile Software Development (AMSAS)
- Security in Scaled Agile Software Development (SISAS)

Natural Language Processing and Legal Tech

- CreateData4AI (CD4AI)
- Applications of Text Generation through Semi-supervised Learning
- Abstractive Text Summarization for Domain-Specific Documents (ATESD)
- AI-Based Digital Health Assistant (ALPHA-KI)
- Conversational Graph-Based Navigation Over Semantically Connected Content (COGNOSCO)
- Research Institution Knowledge Graph (RIKG)
- NLawP – Natural Language Processing and Legal Tech
- Scientific Claim Verification with Evidence from Text and Structured Knowledge (VeriSci)

Digital Platforms and Ecosystems

- Ecosystem for a Knowledge-Based Platform Supporting In-Store Logistics
- Differential Privacy
- Learn, Apply, Comply: Development of Continuing Education Materials on Privacy-Enhancing Technologies (LACE)
- Digital Platform Engineering in the Government Domain

Blockchain-Based Systems Engineering

- Digital Credentials for Higher Education Institutions (DiBiHo)
- GAIA-X 4 Production, After-Sales and PLC - Across Automated Driving
- Algorand Centres of Excellence (ACE) - SUPPRA
- Blockchain Technology for International Student Mobility
- Management of Education Certificates

Sebis Team

Sebis Team Members

TLM

Sascha Nägele
Large-Scale Agile
Development & Security

Alexandra Klymenko
Privacy

Prof. Dr. Florian Matthes
Head of sebis

Tri Huynh
Domain Specific Languages

Pascal Philipp
Large-Scale Agile
Development & Metrics

Nektarios Machner
Software Engineering

Tim Schopf
NLP & Knowledge
Representation

Phillip Schneider
NLP

Felix Hoops
Blockchain & SSI

Tobias Müller
Privacy & ML

Peter Kuhn
Platform Engineering

Burak Öz
Blockchain

Anum Afzal
NLP

Mahdi Dhaini
Explainable NLP

Juraj Vladika
NLP

Stephen Meisenbacher
NLP & Privacy

Franziska Tobisch
Large-Scale Agile
Development
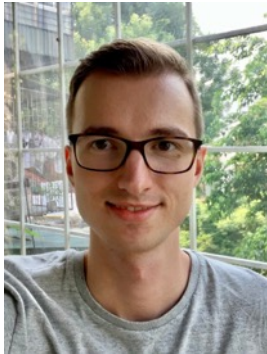
Oliver Wardas
NLP & LegalTech

Wessel Poelman
NLP

Recent Work in Blockchain-Based System Engineering

Burak Öz, Felix Hoops 29.06.2023, sebis Day 2023

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
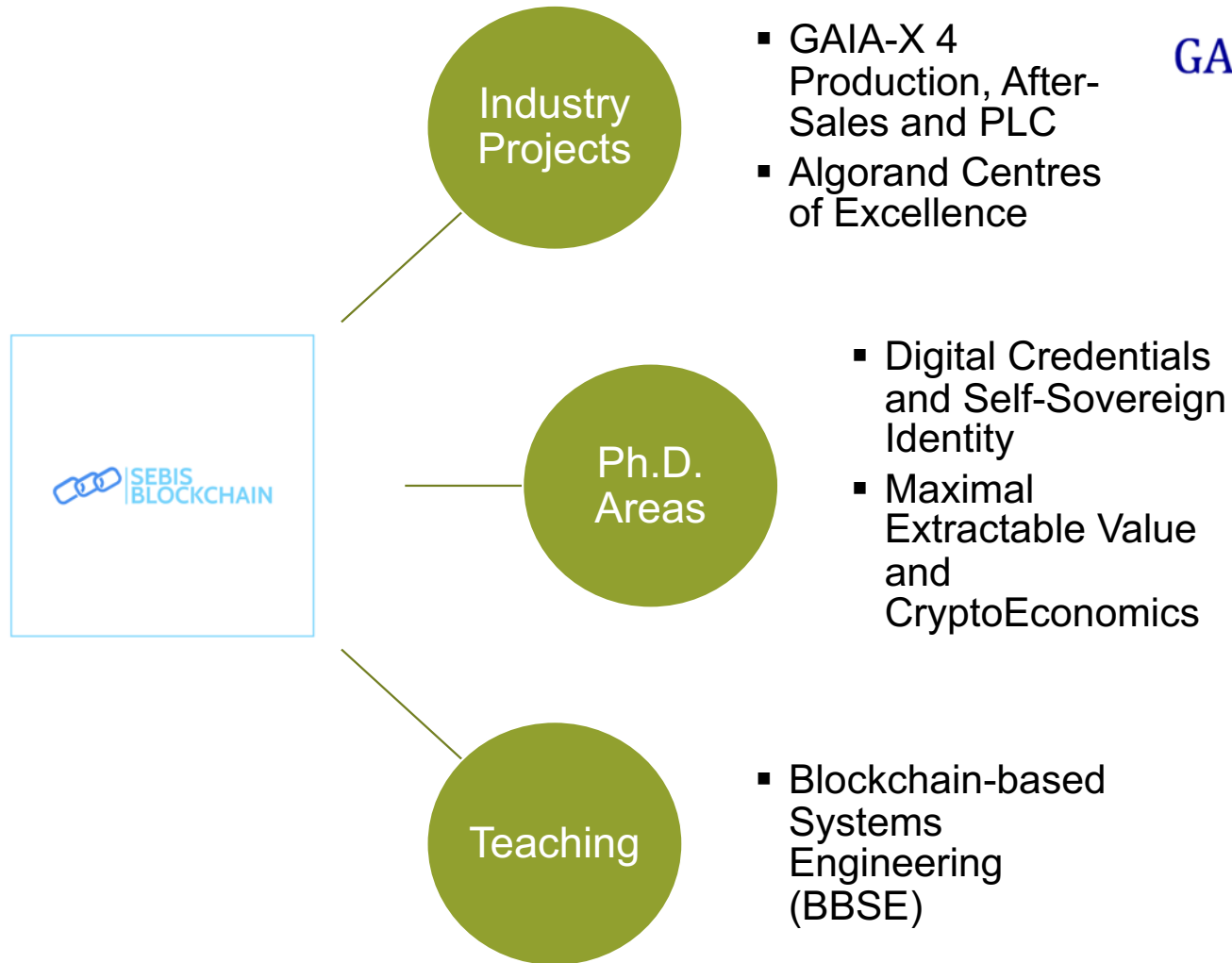Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Overview

Felix Hoops

Burak Öz

SEBIS BLOCKCHAIN

Industry Projects
- GAIA-X 4 Production, After-Sales and PLC
- Algorand Centres of Excellence

GAIA-X 4 PLC-AAD

Algorand Centres of Excellence

Ph.D. Areas
- Digital Credentials and Self-Sovereign Identity
- Maximal Extractable Value and CryptoEconomics

Teaching
- Blockchain-based Systems Engineering (BBSE)
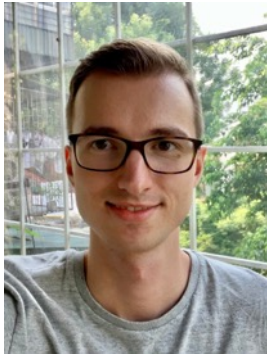
Gaia-X: https://wwwmatthes.in.tum.de/pages/va5wdn40w6q2/GAIA-X-4-Production-After-Sales-and-PLC-Across-Automated-Driving
Algorand: https://wwwmatthes.in.tum.de/pages/jfvlv68q7m4a/Algorand-Centres-of-Excellence-ACE-SUPPRA
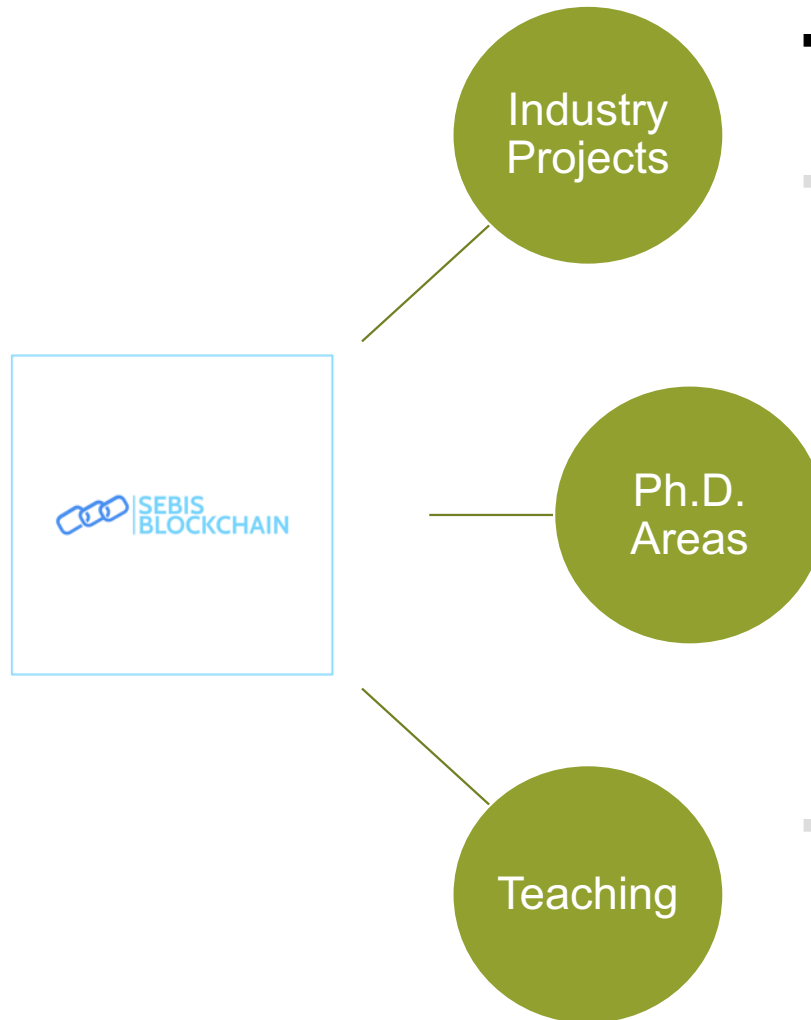BBSE: https://wwwmatthes.in.tum.de/pages/enf3vo4lqv74/Blockchain-based-Systems-Engineering-BBSE

GAIA-X 4 Production, After-Sales and PLC

TLN

Felix Hoops

Burak Öz

SEBIS BLOCKCHAIN

Industry Projects

Ph.D. Areas

Teaching

- GAIA-X 4 Production, After-Sales and PLC
- Algorand Centres of Excellence

GAIA-X 4 PLC-AAD

Algorand
Centres of Excellence

- Digital Credentials and Self-Sovereign Identity
- Maximal Extractable Value and CryptoEconomics

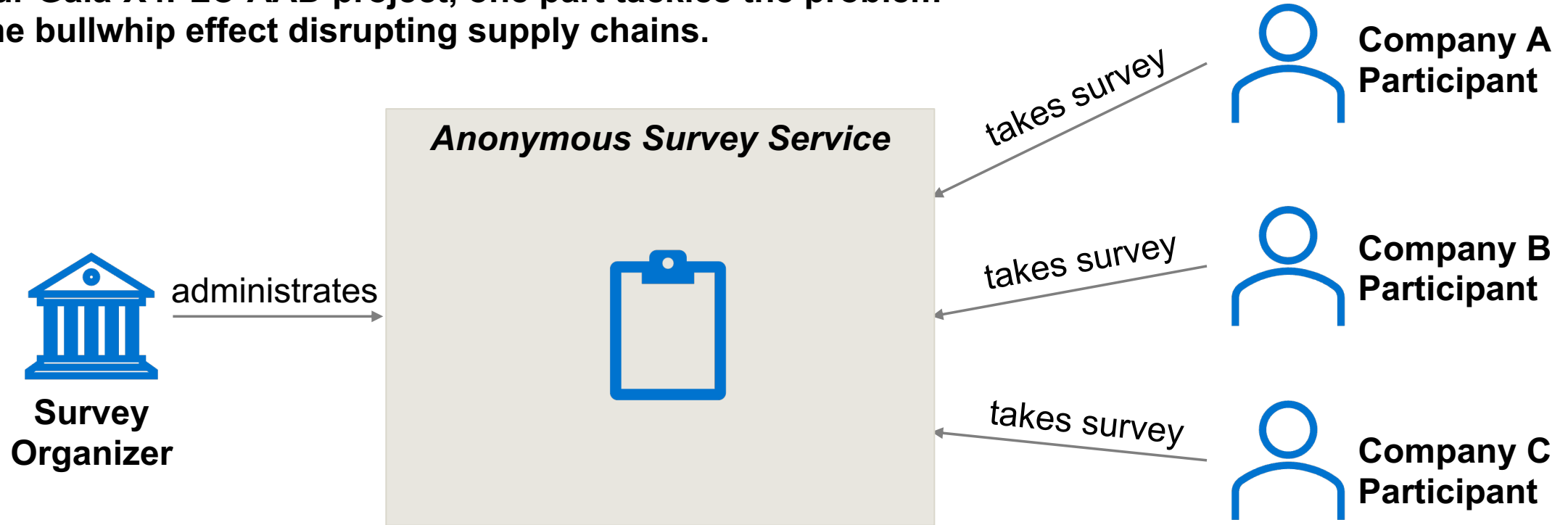- Blockchain-based Systems Engineering (BBSE)

Gaia-X: https://wwwmatthes.in.tum.de/pages/va5wdn40w6q2/GAIA-X-4-Production-After-Sales-and-PLC-Across-Automated-Driving
Algorand: https://wwwmatthes.in.tum.de/pages/jfvlv68q7m4a/Algorand-Centres-of-Excellence-ACE-SUPPRA
BBSE: https://wwwmatthes.in.tum.de/pages/enf3vo4lqv74/Blockchain-based-Systems-Engineering-BBSE

The Bullwhip Effect Survey

In our Gaia-X4PLC-AAD project, one part tackles the problem of the bullwhip effect disrupting supply chains.

Anonymous Survey Service

Survey Organizer *administrates* → **Anonymous Survey Service**

takes survey ← **Company A Participant**

takes survey ← **Company B Participant**

takes survey ← **Company C Participant**

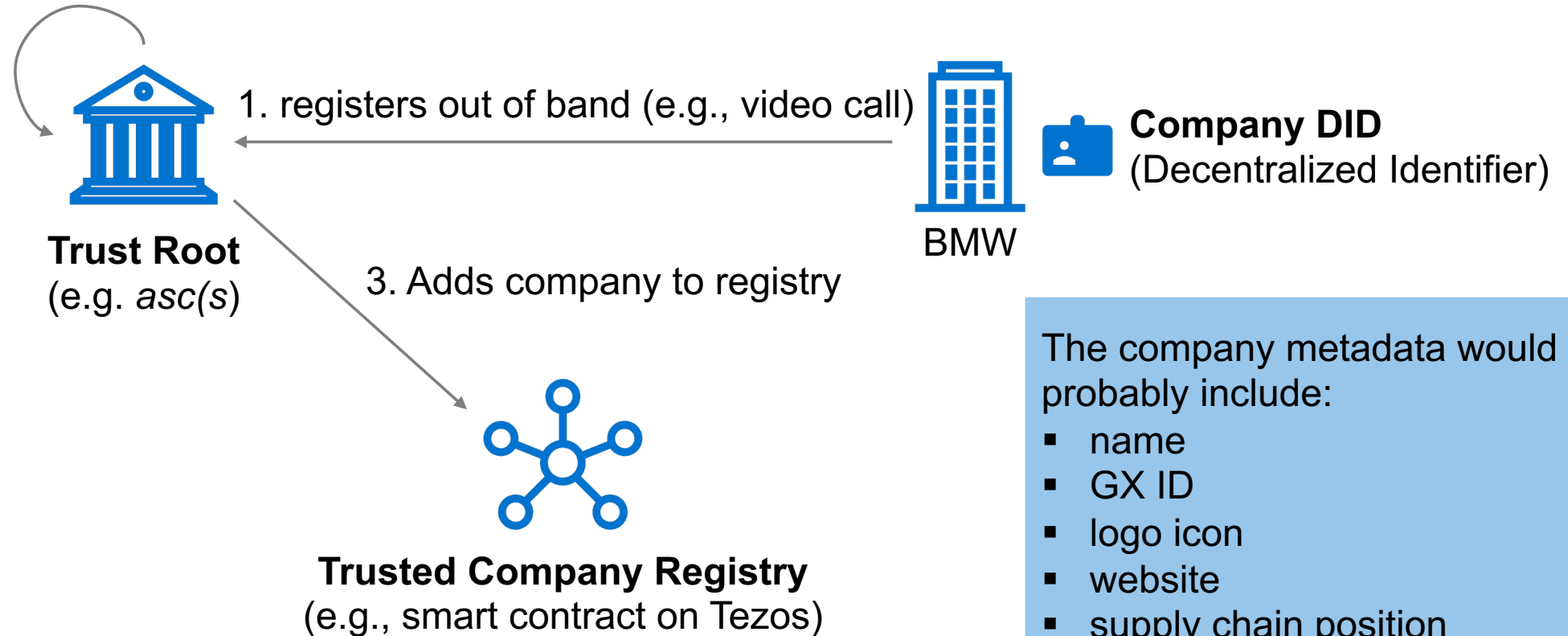*Who is allowed to take part in the survey?*

How do participants get access?

How does the survey know the role of a participant's company in the supply chain?

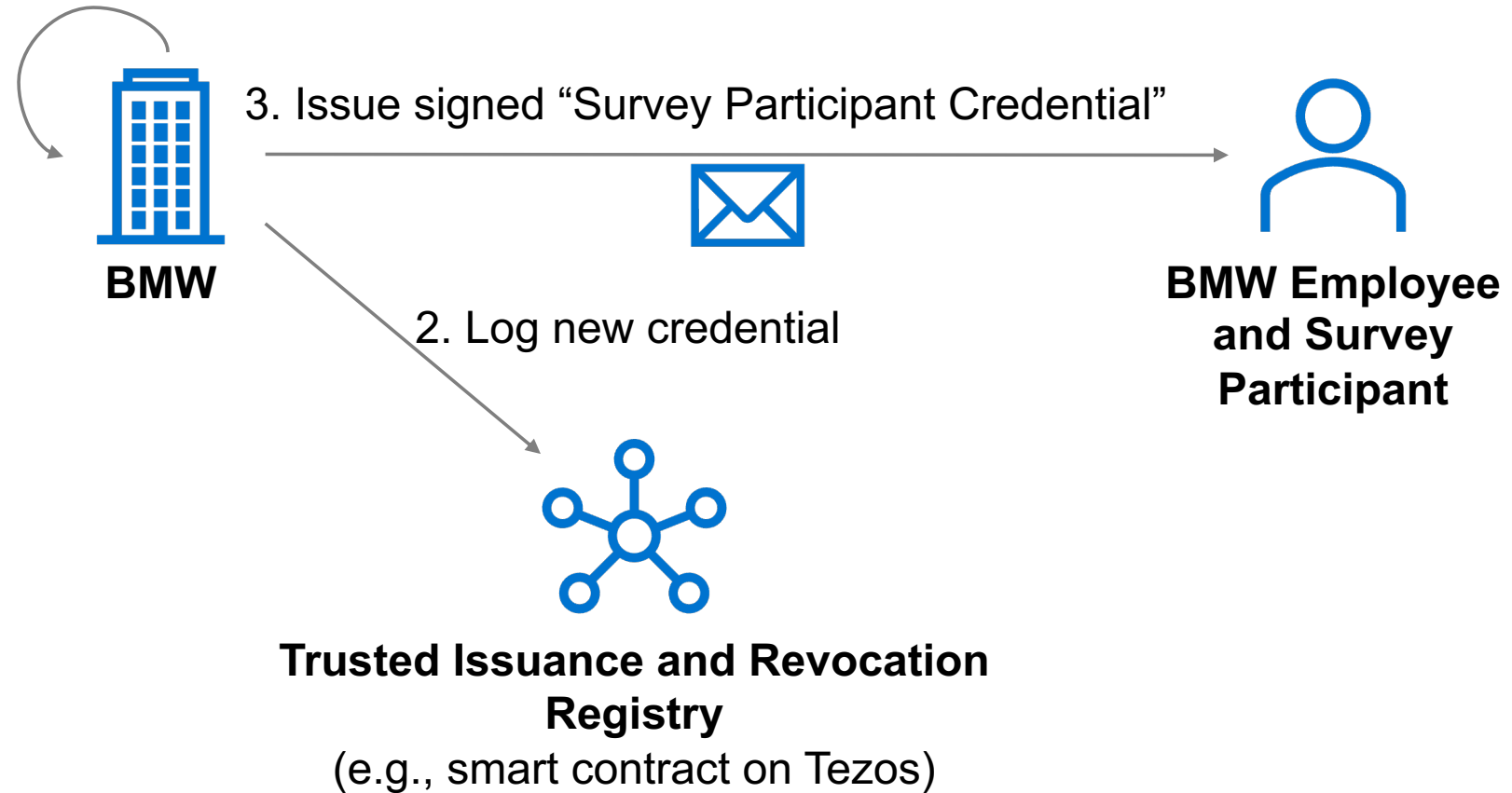# Customer Journey through the Bullwhip Effect with SSI
Company Setup

2. Confirms company metadata and relevance

Trust Root
(e.g. *asc(s)*)

1. registers out of band (e.g., video call)

BMW

Company DID
(Decentralized Identifier)

3. Adds company to registry

Trusted Company Registry
(e.g., smart contract on Tezos)

The company metadata would probably include:
- name
- GX ID
- logo icon
- website
- supply chain position
- …

1. Generate "Survey Participant Credential"

3. Issue signed "Survey Participant Credential"

BMW

2. Log new credential

BMW Employee and Survey Participant

Trusted Issuance and Revocation Registry
(e.g., smart contract on Tezos)

Survey Participation

Anonymous Survey Service

Survey Organizer — administrates →

1) Presents "Survey Participant Credential"

2) Shows OEM survey with login as BMW

3) fills survey

BMW Employee and Survey Participant

Trusted Issuance and Revocation Registry
(e.g., smart contract on Tezos)

Self-Sovereign Identity

- **Self-Sovereign Identity** (SSI) is the term used to describe a decentralized goal architecture using W3C Verifiable Credentials.
- With SSI you can instantly create an account (i.e., identifier) without anyone being able to prevent that.
  - You alone control that account, which means no one can shut it down or take it over.
- Accounts can make statements about other accounts in the form of **Verifiable Credentials**
  - Including but not limited to Name, Date of Birth, Education Level, Company Affiliation

Current Model

SSI Model
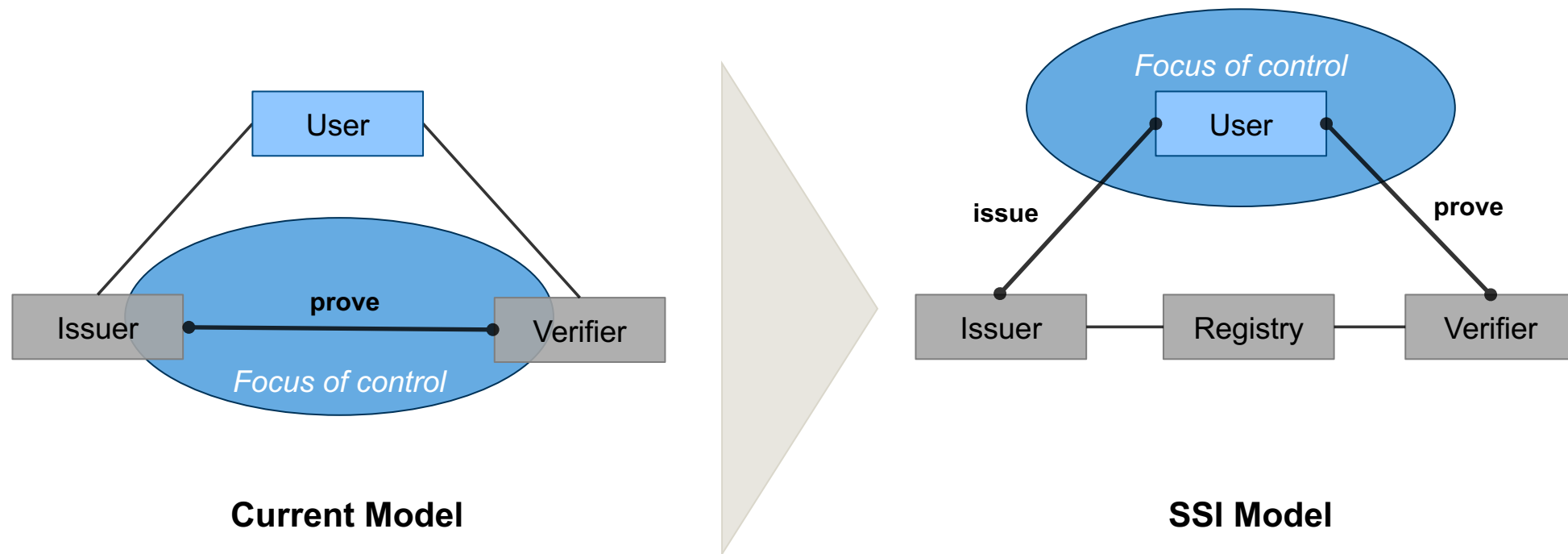
Blockchain technology is a solid choice to publish and administrate such an account.

Maximal Extractable Value (MEV) and CryptoEconomics

Felix Hoops

Burak Öz

Industry Projects
- GAIA-X 4 Production, After-Sales and PLC
- **Algorand Centres of Excellence**

Ph.D. Areas
- Digital Credentials and Self-Sovereign Identity
- **Maximal Extractable Value and CryptoEconomics**

Teaching
- Blockchain-based Systems Engineering (BBSE)

Gaia-X: https://wwwmatthes.in.tum.de/pages/va5wdn40w6q2/GAIA-X-4-Production-After-Sales-and-PLC-Across-Automated-Driving
Algorand: https://wwwmatthes.in.tum.de/pages/jfvlv68q7m4a/Algorand-Centres-of-Excellence-ACE-SUPPRA
BBSE: https://wwwmatthes.in.tum.de/pages/enf3vo4lqv74/Blockchain-based-Systems-Engineering-BBSE

Maximal Extractable Value (MEV) and CryptoEconomics

CryptoEconomics		Consensus Security
MEV		
Data Forensics		Incentive Mechanisms

"When you're looking at **a system in a certain state**, with some **transactions pending**, and you've got the **power to determine** which ones get the green light and in what order, plus the ability to throw some new transactions into the mix, then **MEV is the most value you can possibly pocket** from this setup."

On-Going Research

- **MEV x Algorand Blockchain**

- **Waiting Games in Ethereum Blockchain (AFT 23')**

Completed Research

- **Algorand MEV Analytics**
  - `https://algorand.streamlit.app/`
- **NFT Wash Trading Detection**
  - `https://xyz.pizza/`

Posters Today

- **An Overview of Maximal Extractable Value in Blockchains**
- **An Analysis of Maximal Extractable Value in Algorand Blockchain**
- **CowSwap to Mitigate Maximal Extractable Value**
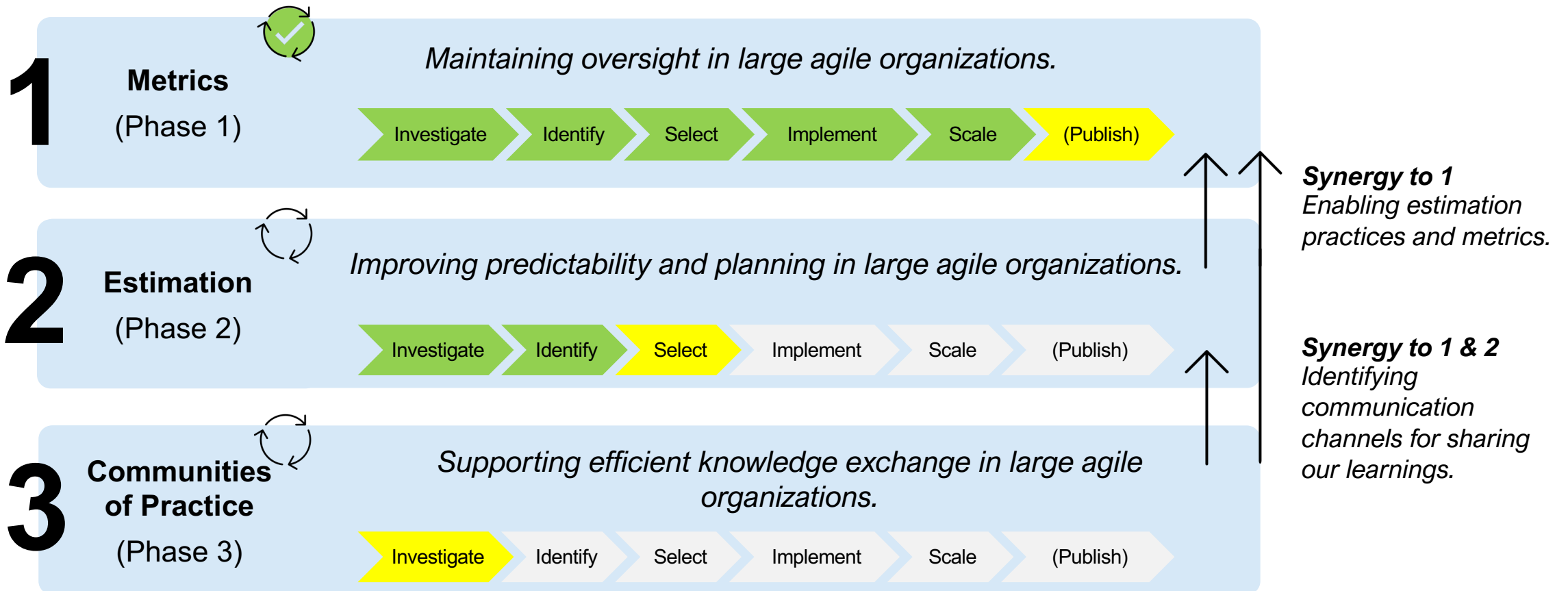
Metric-Driven Large Agile Organizations

Pascal Philipp, Franziska Tobisch

29.06.2023, Sebis Day

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
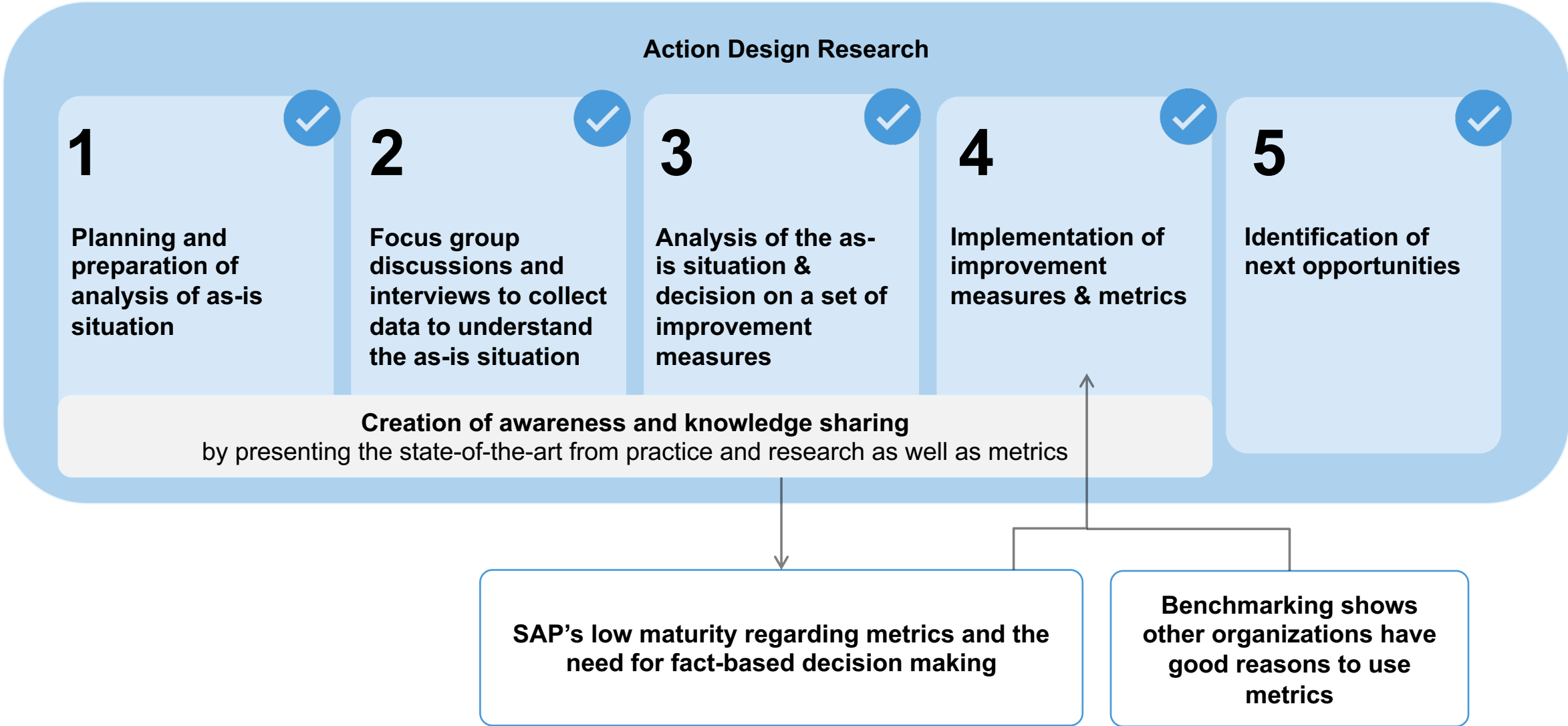Technical University of Munich (TUM)
wwwmatthes.in.tum.de

We follow a multi-pronged approach and facilitate synergy effects for optimal project performance

Scaling Agile Topics

1 **Metrics** (Phase 1)

Maintaining oversight in large agile organizations.

Investigate → Identify → Select → Implement → Scale → (Publish)

2 **Estimation** (Phase 2)

Improving predictability and planning in large agile organizations.

Investigate → Identify → Select → Implement → Scale → (Publish)

3 **Communities of Practice** (Phase 3)

Supporting efficient knowledge exchange in large agile organizations.

Investigate → Identify → Select → Implement → Scale → (Publish)

Synergy to 1
Enabling estimation practices and metrics.

Synergy to 1 & 2
*Identifying communication channels for sharing our learnings.*

We successfully completed phase 1 and already initiated phases 2 & 3

We used Action Design Research (ADR) as method to co-create solutions fitted to the requirements of SAP

Action Design Research

1 ✓ Planning and preparation of analysis of as-is situation

2 ✓ Focus group discussions and interviews to collect data to understand the as-is situation

3 ✓ Analysis of the as-is situation & decision on a set of improvement measures

4 ✓ Implementation of improvement measures & metrics

5 ✓ Identification of next opportunities

Creation of awareness and knowledge sharing
by presenting the state-of-the-art from practice and research as well as metrics

SAP's low maturity regarding metrics and the need for fact-based decision making

Benchmarking shows other organizations have good reasons to use metrics

We used our initial solution designed at an insurance service provider as a basis for our ADR efforts at SAP (1/2)
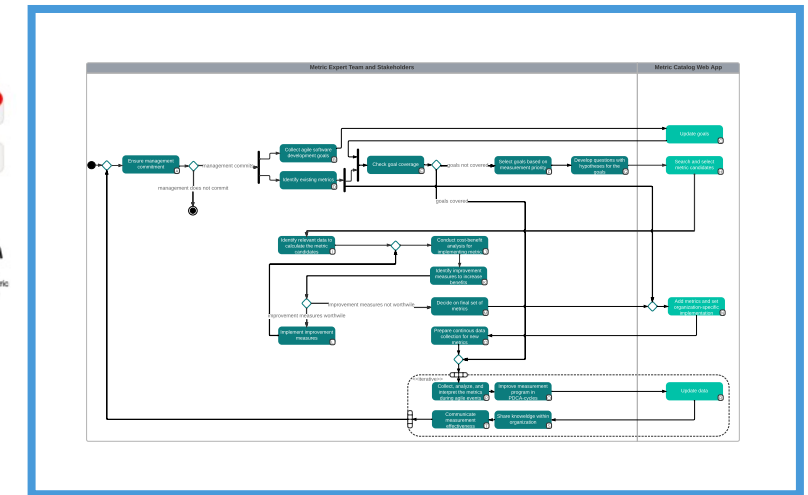

Method design for an insurance provider
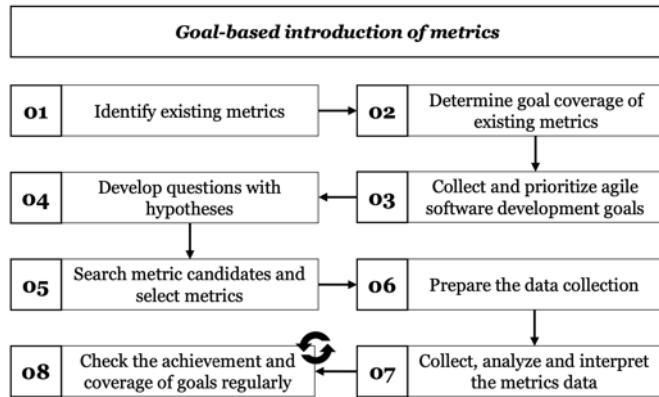
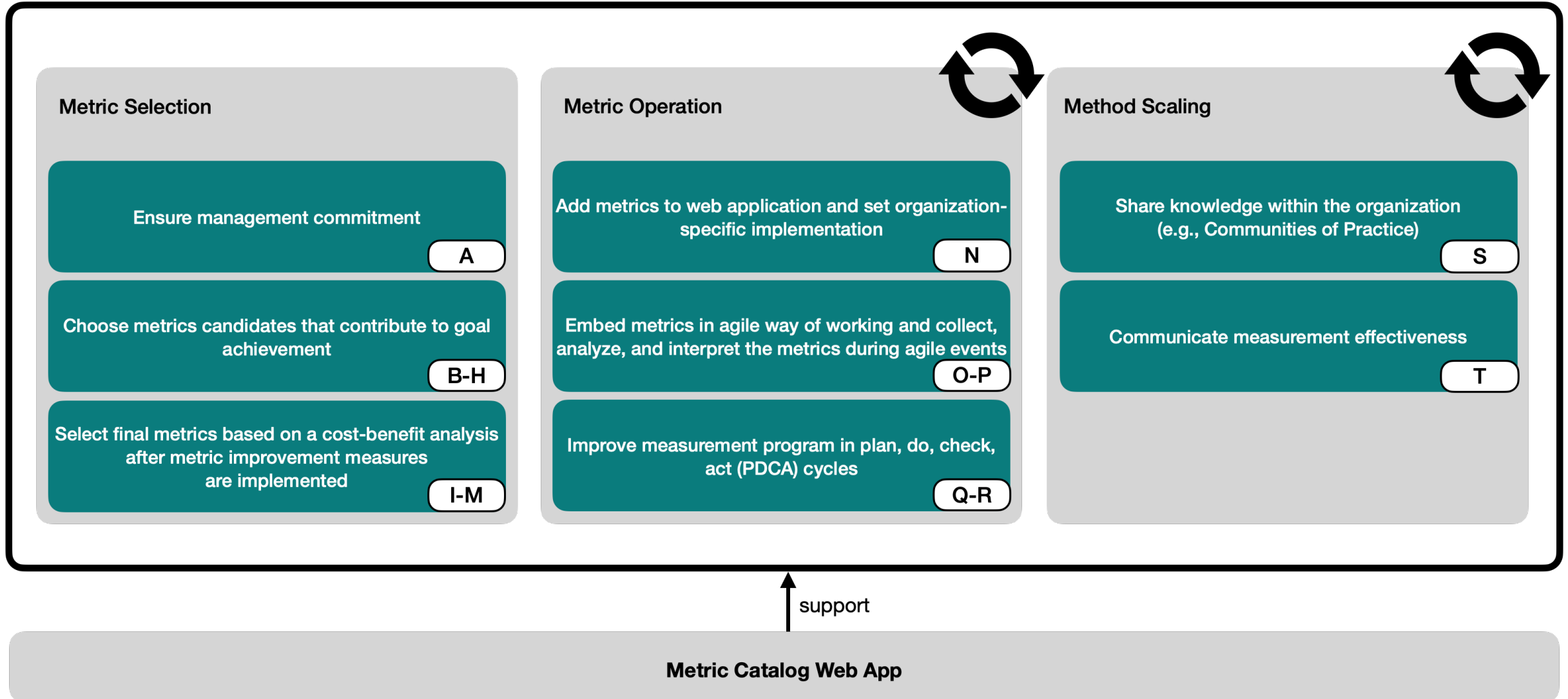First method design for SAP
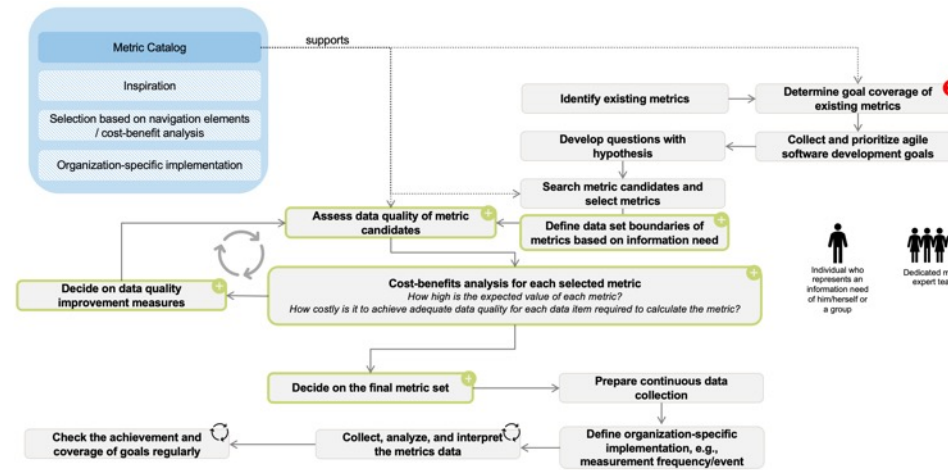
Final method design for SAP

First Evaluation

Second Evaluation

We used our initial solution designed at an insurance service provider as a basis for our ADR efforts at SAP (2/2)

Metric Selection

Ensure management commitment
A

Choose metrics candidates that contribute to goal achievement
B-H

Select final metrics based on a cost-benefit analysis after metric improvement measures are implemented
I-M

Metric Operation

Add metrics to web application and set organization-specific implementation
N

Embed metrics in agile way of working and collect, analyze, and interpret the metrics during agile events
O-P

Improve measurement program in plan, do, check, act (PDCA) cycles
Q-R

Method Scaling

Share knowledge within the organization (e.g., Communities of Practice)
S

Communicate measurement effectiveness
T

support

Metric Catalog Web App

We investigated how effort estimation was conducted within a successful large-scaling agile program at an energy provider

S/4HANA transformation program with the goal to build one commercial harmonized ERP core as the backbone for operational business "in time & in budget"

12 months from kick-off to first go live	20 scrum teams in charge to deliver functional scope	~ 350 people involved (from there different organizations)
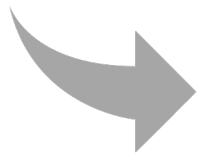
1. **Interviews with experts** from the case program to identify how effort estimation is performed and what challenges exist
2. Identification of mitigation propositions to counteract the challenges based on the interviews and literature
3. **Evaluation interviews & survey replies** by the experts to evaluate the criticality of the challenges and the effectiveness of the identified mitigation propositions

Synergy with metrics topic: Applying the knowledge gained in the bachelor thesis helps us improve the estimation practices at SAP, consequently leading to better estimates (e.g., estimation metrics)

We identified 25 effort estimation challenges in interviews with 20 experts from the case program and evaluated them together afterwards

	Challenge	Avg. criticality
C23	**Unclear and incomplete specification of the requirements**	**1,50**
C21	**Information deficit in the initial estimation of large, complex requirements**	**1,58**
C4	**Pressure and control by management**	**1,58**
C20	**Adjustment of effort estimates**	**1,58**
C14	**Considering dependencies to other teams, workstreams, and systems in the estimation**	**1,67**
C1	Project setting	1,75
C12	Having a correct and common understanding of requirements	1,83
C15	Subjectivity of estimates	1,92
C22	Information deficit regarding new requirements	1,92
C25	Unforeseen changes	2,08
C7	Monitoring of estimations and actual efforts	2,25
C19	Neglection of relevant factors when estimating	2,25
C13	Difficulty to estimate additional overhead (e.g., meetings and explanations)	2,25
C9	Lack of (team) commitment	2,42
C3	Unclear responsibilities	2,42
C11	Efficient communication despite spatial distribution and language barriers	2,50
C2	Time restrictions	2,58
C16	Lack of experts involved in estimates	2,67
C17	Lack of involvement of experts in top-level estimations	2,75
C10	Lack of knowledge about contact partners in the beginning	2,83
C18	Lack of knowledge and experience regarding effort estimation	2,92
C6	Inappropriate tool support	3
C5	Lack of measures to improve estimations	3
C8	Difficulties to estimate in story points units	3,08
C24	Missing knowledge about resources in terms of people involved in the implementation	3,08

1 = very critical, … , 5 = very uncritical

We identified 21 mitigation propositions to counteract the effort estimation challenges and evaluated them together with experts from the case program
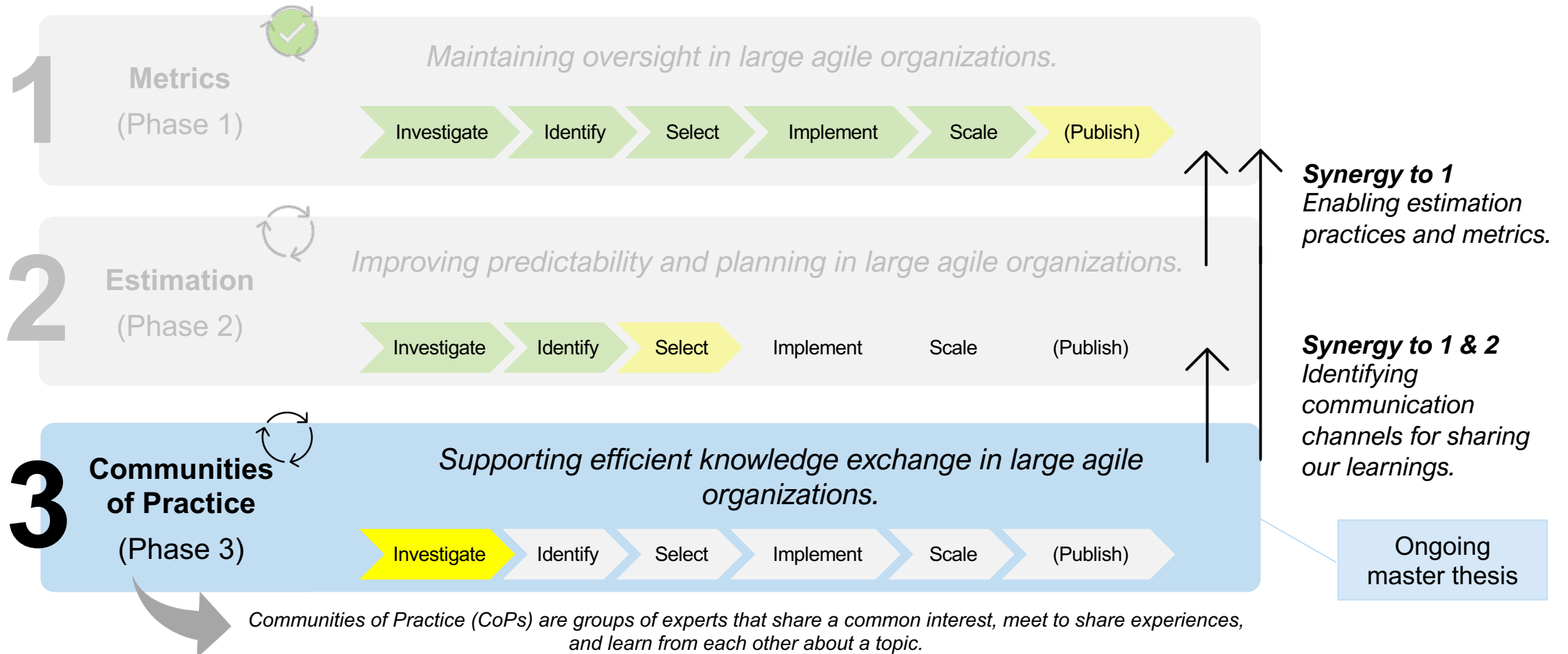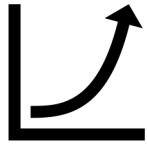
	Mitigation proposition	Avg. estimated effectiveness*
M15	**Support and motivation by scrum masters and agile coaches**	**1,58**
M17	**Tool/feature to support the estimation process**	**1,83**
M4	**Deep dive sessions and workshops to clarify requirements and needed resources**	**1,92**
M7	**Guidelines and standard estimates for tasks uniform across all teams**	**1,92**
M3	Cultivation of a dependency list	2,08
M13	Platforms and meetings to identify and discuss dependencies	2,08
M6	Events to recap on learning and document those	2,17
M8	Consider additional effort regarding organizational and process factors during estimation	2,17
M11	Normalization of story points (to person days)	2,17
M12	Plan requirements for an appropriate time frame to improve the accuracy of estimates	2,17
M21	Use agile metrics and store estimates for improvement and future use	2,17
M5	Early and continuous communication between all levels: teams, workstreams, and program management	2,25
M9	Include people with experience in estimating effort and reacting to unforeseen in the estimation process	2,25
M14	Reduce pressure for employees	2,33
M1	Adding a buffer to estimations in case of uncertainty	2,5
M16	Tool support to automate the estimation process	2,5
M10	Measures to convince the team that effort estimation is a group/team activity	2,67
M2	Additional phase before the implementation phase to check requirements in detail (feasibility and quality)	2,92
M18	Tracking of actual efforts	2,92
M19	Use of supporting techniques during the estimation process	2,92
M20	Use T-Shirt size as an estimation unit to avoid difficulties to estimate in story points and trust the team more	3,25

1 = highly effective, … , 5 = not highly effective

Outlook: Communities of Practice (CoP)

Scaling Agile Topics

1

Metrics (Phase 1)

Maintaining oversight in large agile organizations.

Investigate → Identify → Select → Implement → Scale → (Publish)

2

Estimation (Phase 2)

Improving predictability and planning in large agile organizations.

Investigate → Identify → Select → Implement → Scale → (Publish)

3

Communities of Practice (Phase 3)

Supporting efficient knowledge exchange in large agile organizations.

Investigate → Identify → Select → Implement → Scale → (Publish)

Communities of Practice (CoPs) are groups of experts that share a common interest, meet to share experiences, and learn from each other about a topic.

Synergy to 1
Enabling estimation practices and metrics.

Synergy to 1 & 2
Identifying communication channels for sharing our learnings.

Ongoing master thesis

Outlook

Scale the metric approach further at SAP
- Identify more team/stakeholders
- Introduce estimation metrics

Decide on the topic(s) for the next phases together with stakeholders from SAP
- Identify and talk to stakeholders
- Potential topics: Estimation, CoPs
- Define work items together with stakeholders

Publication in progress
- Title: "*A Method to Support the Selection and Operation of Metrics in Large-Scale Agile Software Development*"
- Year: 2023

**Prof. Florian Matthes,
Pascal Philipp,
Franziska Tobisch**

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.17132
matthes@in.tum.de
wwwmatthes.in.tum.de

CreateData4AI

Stephen Meisenbacher, Tim Schopf

29.06.2023, sebis Day

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Unstructured Data

What is it?

Simply:
Data that cannot be stored in a traditional row-by-column database

Examples:
Text files, emails, web pages, social media, transcripts, presentations, etc.

Scale:
Roughly 80-90% of enterprise data is unstructured! [1]

Volume:
The volume of unstructured data is growing at a rate of 62% per year [2]

Global Data Generated Annually

Unstructured Data

The Gap

FORBES > INNOVATION

The Big (Unstructured) Data Problem

Juliette Rizkallah Forbes Councils Member
Forbes Technology Council COUNCIL POST | Membership (Fee-Based)
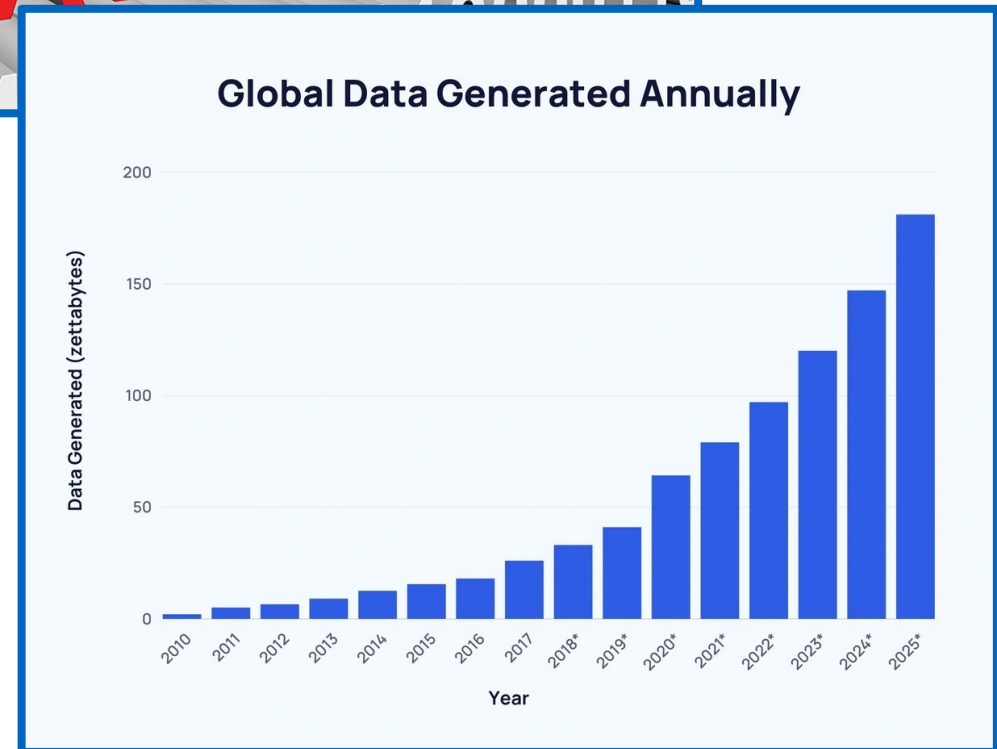
IDEAS MADE TO MATTER | ANALYTICS

Tapping the power of unstructured data

by **Tam Harbert** | Feb 1, 2021

Why It Matters

Data locked away in text, audio, social media, and other unstructured sources can be a competitive advantage for firms that figure out how to use it.

But…

Only 18% of organizations in a 2019 survey by Deloitte reported being able to take advantage of such data!* [3]

Left: https://mitsloan.mit.edu/ideas-made-to-matter/tapping-power-unstructured-data

Right: https://www.forbes.com/sites/forbestechcouncil/2017/06/05/the-big-unstructured-data-problem/

*With some estimates as low as **0.5%** [4]

Some Definitions (1/2)

Keyword: the atomic unit of language carrying meaning, which in this case serves as a base indicator for a piece of text belonging to a specific class
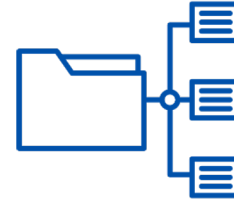
Class: also *tag*, refers to an expert-defined classification scheme, in which each class points to a distinct group of meaning

Context Window: the relevant window of context surrounding class-specific keywords

Context Rule: a context rule deemed by a domain expert to be representative of a specific class. Each class may have many context rules

Classify texts according to expert defined rules, based upon windows of contexts surrounding keywords

Domain Expert: the domain expert is at the center of our proposed approach. Given unstructured data, the domain expert is tasked with conceptualizing classes that reside in the data. Furthermore, key information, such as keywords and class descriptions, are provided by the expert to kickstart our pipeline. The domain expert is also vital to validation in the development of our proposed pipeline.

The expert…
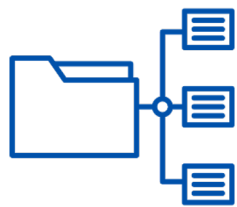
Conceptualizes Classes

Creates Keywords
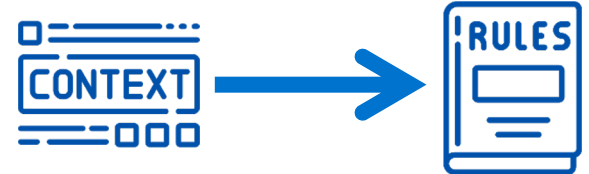
Validates Context Windows into Rules

The Proposed Pipeline
Finding the diamond in the rough

Text Corpus — A corpus of unstructured text (i.e. documents) to be annotated. **0**

Keyword Extraction — Keyword sets for each class are created. **1**

Context Window Extraction — Centered around keywords, context windows are extracted from a sample of documents. **2**

Context Rule Creation — With the help of a domain expert, context rules are derived from the extracted context windows. **3**

Extrapolation — The previously created context rules are then used to annotate the remainder of the unstructured text data. **4**

Dataset — The result is a structured, annotated dataset which embodies the worldview of the domain expert. **5**

Ready for AI Applications!

In what way can current state-of-the-art Natural Language Processing techniques be augmented to incorporate specific domain knowledge, with the goal of transforming unstructured text to structured datasets?

1 How can domain experts be supported in the definition of classes for characterizing large text corpora, particularly in the creation of keywords?

2 Which NLP techniques are best suited for the extraction of coherent windows of context centered around predefined keywords?

3 In which way can a set of context rules be most efficiently and accurately applied to large-scale text corpora?

4 How can the accuracy and usability of the proposed pipeline be validated and evaluated?

5 What is the best way to present the resulting research, such that users of varying backgrounds and expertise can utilize its capabilities?

Current Progress: Keyword Extraction

| Text Corpus | A corpus of unstructured text (i.e. documents) to be annotated. | 0 |
| Keyword Extraction | Keyword sets for each class are created. | 1 |

Subset data (~10k)

Guided Keyword extraction

Seed keywords

List of extracted keywords

discarded

Keyword generation

List of Generated keywords

Filtering for all classes (e.g. deduplication, clustering etc.)

(Reasonably) complete keyword set per class
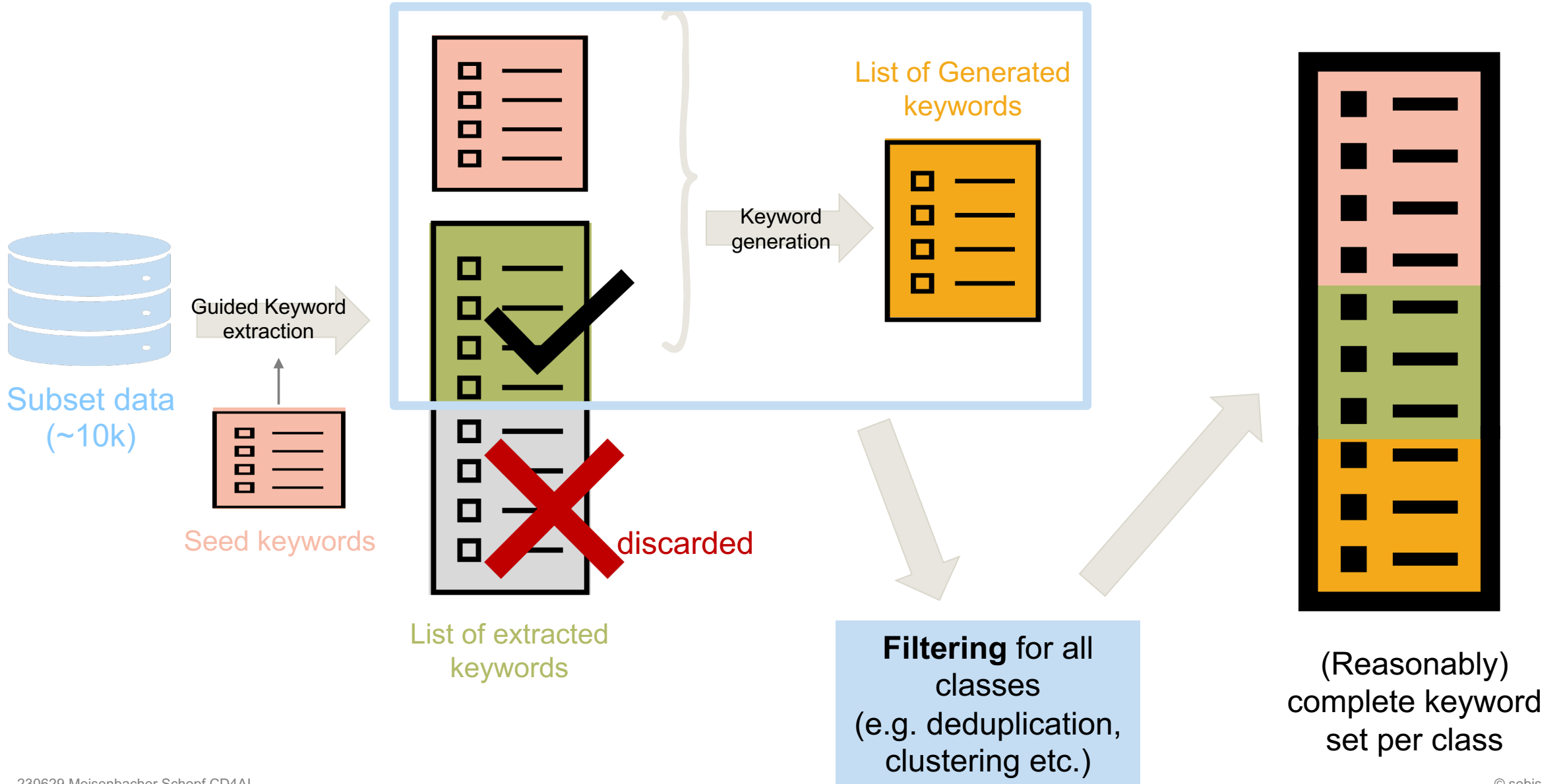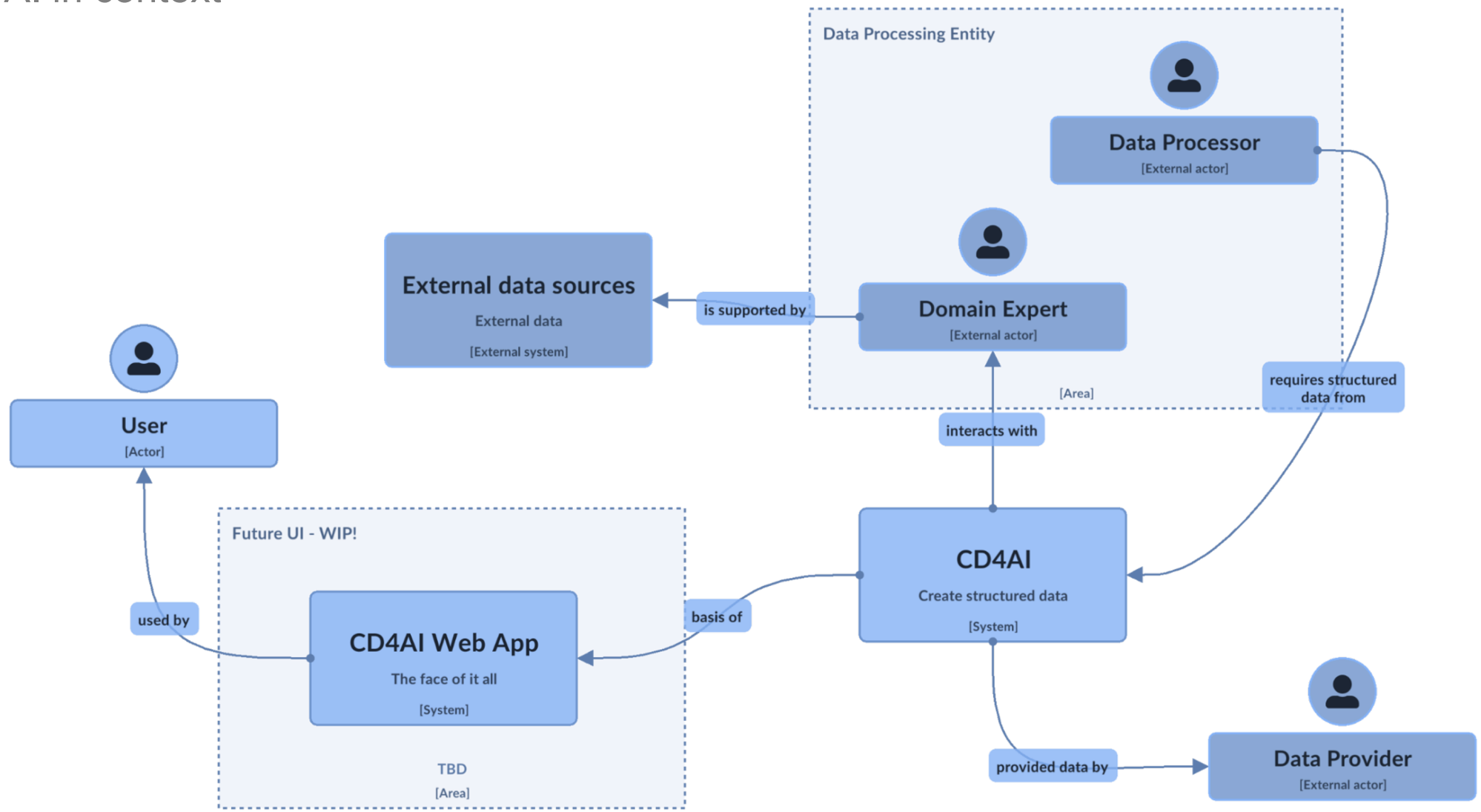
CD4AI in context

CreateData4AI

Context Rule Embedding-Assisted Annotation of Textual Data for AI Applications

In cooperation with:

M.Sc.

Stephen Meisenbacher
stephen.meisenbacher@tum.de

M.Sc.

Tim Schopf
tim.schopf@tum.de

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.17132
matthes@in.tum.de
wwwmatthes.in.tum.de

References

[1] https://www.gartner.com/en/documents/3989657

[2] https://www.ciklum.com/blog/big-data-and-the-challenge-of-unstructured-data

[3] https://www2.deloitte.com/us/en/insights/topics/analytics/insight-driven-organization.html

[4] https://deep-talk.medium.com/80-of-the-worlds-data-is-unstructured-7278e2ba6b73

Engineering Conversational Interfaces for Information Systems

Anum Afzal, Phillip Schneider, Juraj Vladika

29.06.2023, SEBIS Day

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Vision: Conversational Interfaces for Accessing Information Systems

→ Advances in natural language processing (NLP) have brought novel ways of accessing information

- Conversational interfaces make information retrieval more intuitive and interactive, bridging the gap between different data structures

- They enable users to ask questions in plain language and get concise responses, eliminating the need to learn complex query languages or data schemas

- Emergent capabilities of large language models (LLMS) offer a range of capabilities from text analyses, summaries, translation, question answering, or explanations
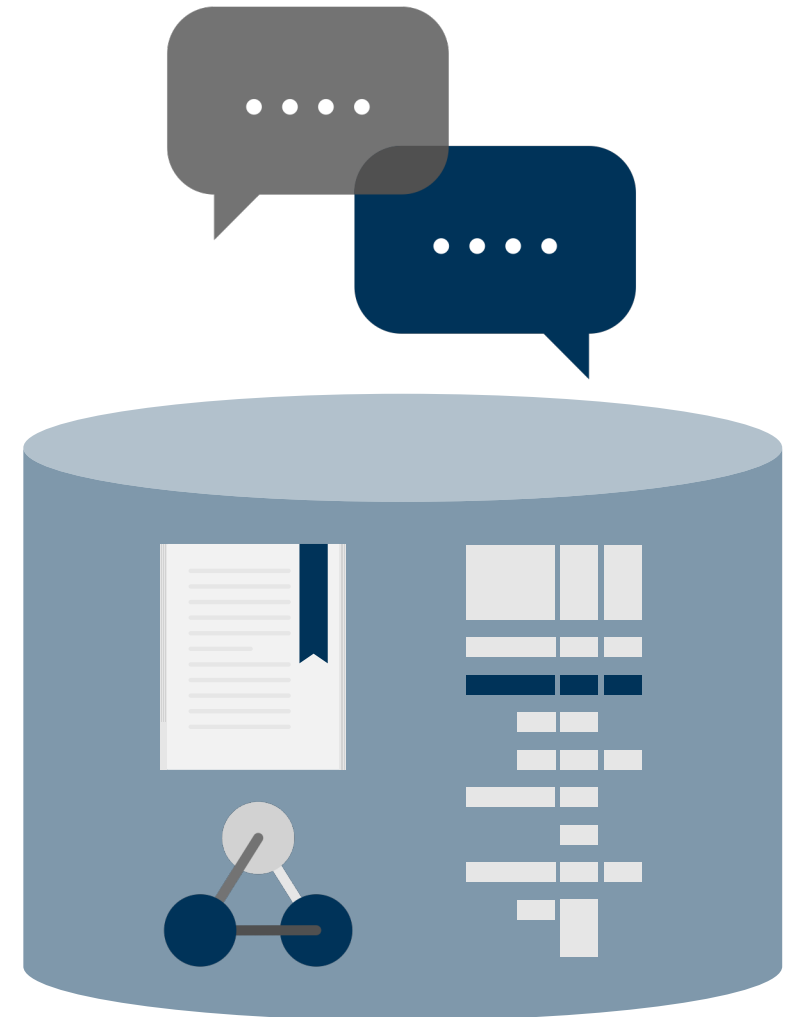
ALPHA-KI Project: Building Conversational Interfaces for Health Assistance

- Project aim is to develop an intelligent voice-based conversational agent for geriatric care

- Skill set has three core areas:
  (1) health skills
  (2) emergency skills
  (3) informative skills
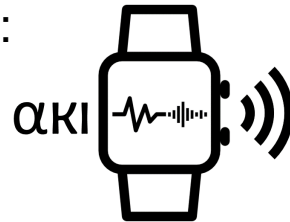
- Main challenge relates to trade-off between dialogue controllability versus flexibility

- Research focuses on grounding conversations in knowledge representations, which is also crucial for the integration of LLMs

ALPHA-KI Project: Building Conversational Interfaces for Health Assistance
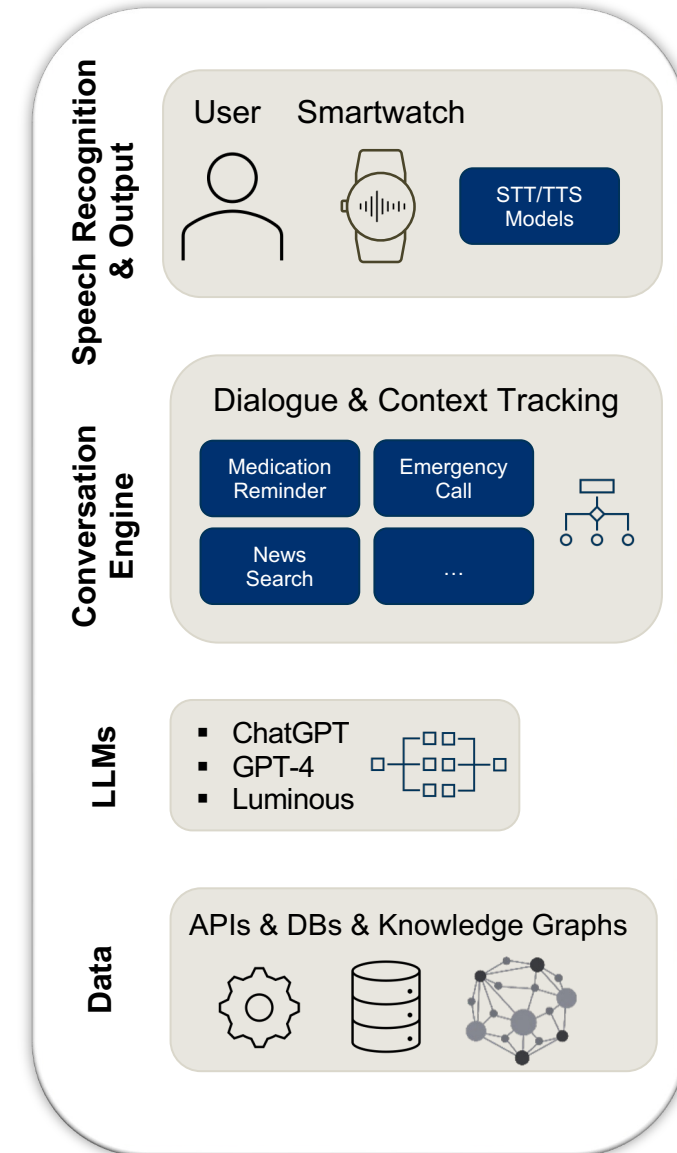
Knowledge Graph with News Articles
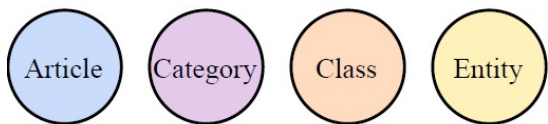
Number of Nodes

Article: 2118

Category: 6

Class: 1264

Entity: 2874

Number of Edges

INSTANCE_OF: 4406

IS_PART_OF: 2152

HAS_ENTITY: 12567
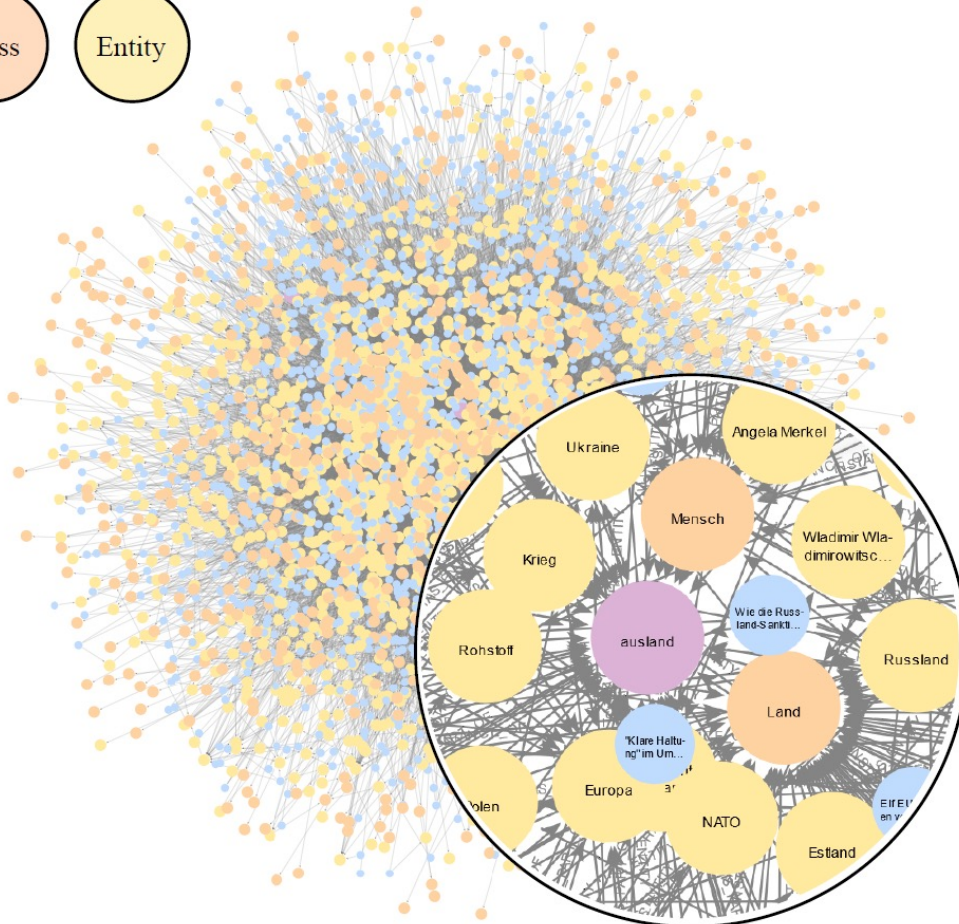
Excerpt from Conversation Logs

State (S)	User Utterance (U) and Agent Reponse (A)
S6: Entity Search	U: News about France.
	A: Here are articles about France. First: Six people killed in ...
S7: Navigation	U: Read the third article.
	A: Building collapsed in Marseille: In the center of the French ...
S7: Navigation	U: Give me more articles.
	A: Okay! First: More protests in France ... Second: Supplier for ...
S7: Navigation	U: I want the second article.
	A: Supplier for ... For news about related topics like Ukraine or ...
S6: Entity Search	U: I want news about Russia.
	A: Here are articles about Russia. First: Clear stance in dealing ...
S7: Navigation	U: Now, read the article number three.
	A: EU sanctions ... For news about related topics like Germany ...
S6: Entity Search	U: Then, about Ukraine.
	A: Here are articles about Ukraine. First: Horror over ...
S7: Navigation	U: Read the article about the Ukranian economy.
	A: Ukrainian economy slumped further: The Russian war ...
S7: Navigation	U: Next article.
	A: Mined for decades: In Ukraine, an area twice the size of ...
S7: Navigation	U: Again, next.
	A: Horror over ... For news about related topics like Putin.
S6: Entity Search	U: Okay, news about Putin.
	A: Here are articles about Vladimir Putin. First: Putin's nuclear ...

COGNOSCO: Conversational Graph-Based Navigation of Scientific Content

- How to develop conversational interfaces for exploratory search in the scholarly domain?

- Enable search-based dialogues with agent that has access to knowledge graph, including a taxonomy of more than 80 NLP subfields
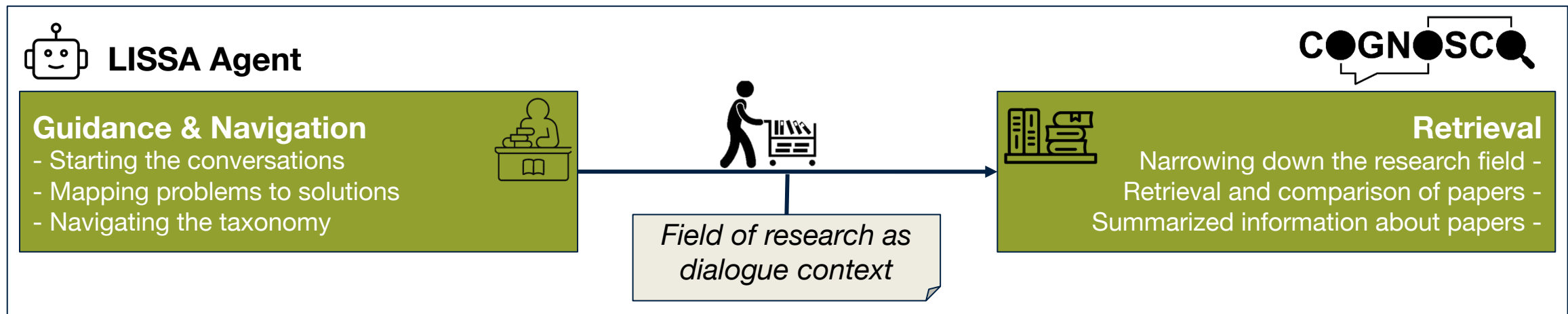
NLP Taxonomy
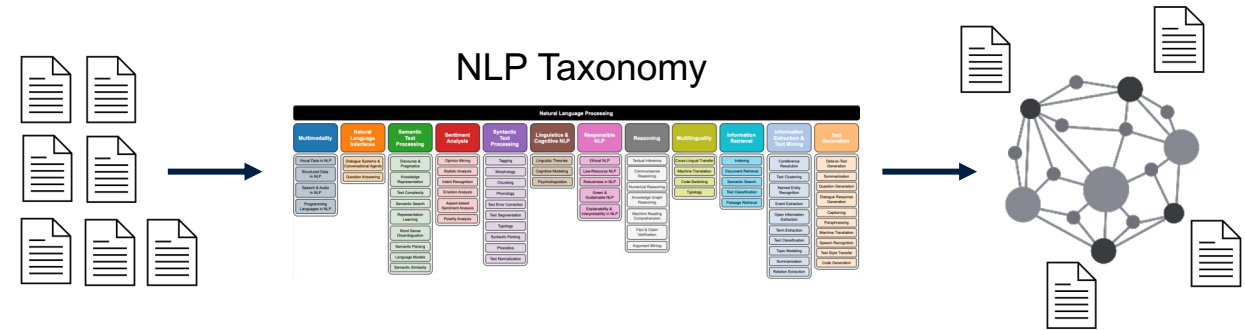
How to build a talking robot?

Is ChatGPT ethical?

What models can automatically summarize laboratory reports?

LISSA Agent

COGNOSCO

Guidance & Navigation
- Starting the conversations
- Mapping problems to solutions
- Navigating the taxonomy

Field of research as dialogue context

Retrieval
Narrowing down the research field -
Retrieval and comparison of papers -
Summarized information about papers -

A Generative Question Answering Approach for a Human Resource FAQ chatbot.

Anum Afzal 29.06.2023

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Motivation and Relevance

Vision: Boost productivity for every user in the intelligent enterprise

Provide an enterprise-grade Conversational Interface

Vision
for Digital Assistant

For (Target Group)	… **all employees**
who (Need)	… **need to be more productive in the office and on the go,**
we offer (Product)	… **the virtual assistant for the enterprise**
that (Key Benefit)	… **saves time during business task fulfillment,** … **provides easy access to information &** … **helps them to find what they need in the business context they are in.**

Value Proposition
of Digital Assistant

Boost Employee Productivity (by)	… **saving time during business task fulfillment** … **reduced complexity by providing relevant information**
Service Cost Reduction (through)	… **avoided support cases** … **enabling employee self-service**

Differentiation
vs. similar solutions

Unlike (Alternative)	… **simplistic chatbots,**
our solution (USP)	… **scales across the intelligent suite &** … **provides value through a consistent conversational user experience.**

Motivation and Relevance

Application: HR Support Chatbot

- Currently > 330.000 HR tickets per year

- Target is to cover ~30% of the ticket volume via bot functionality

14 Business Days
Saved of ticket processing time for HR support agents
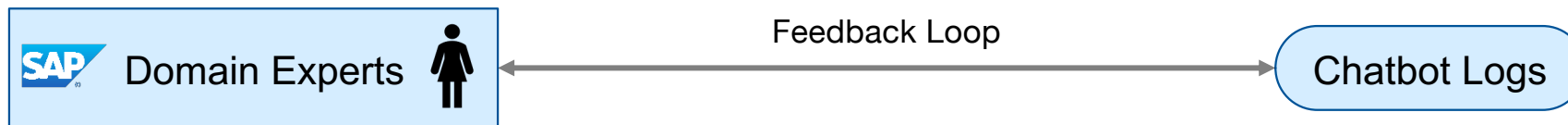
1540 Business Days
Saved of waiting time for ticket requestors

- Uses simplistic **Natural Language Processing** techniques.
  - Current system is already **60% accurate.**
  - **why stop here?**

Project Overview

Contributions & Limitations of the current system.

- Additional data sources such as **chatbot logs**. How to use them?

- **User Utterance:** I am trying to log into Benefitfocus for the first time, but it is not granting me access
  → **Mapped Question from FAQ:** How can I access the BenefitFocus tool?

- **User Utterance:** How do I enter a progression for an employee?
  → **Mapped Question from FAQ :** How do I submit salary changes for my employee?
  → **Correct Question:** How do I promote an employee in the tool?

- Semi-Supervised Learning to include chatbot logs
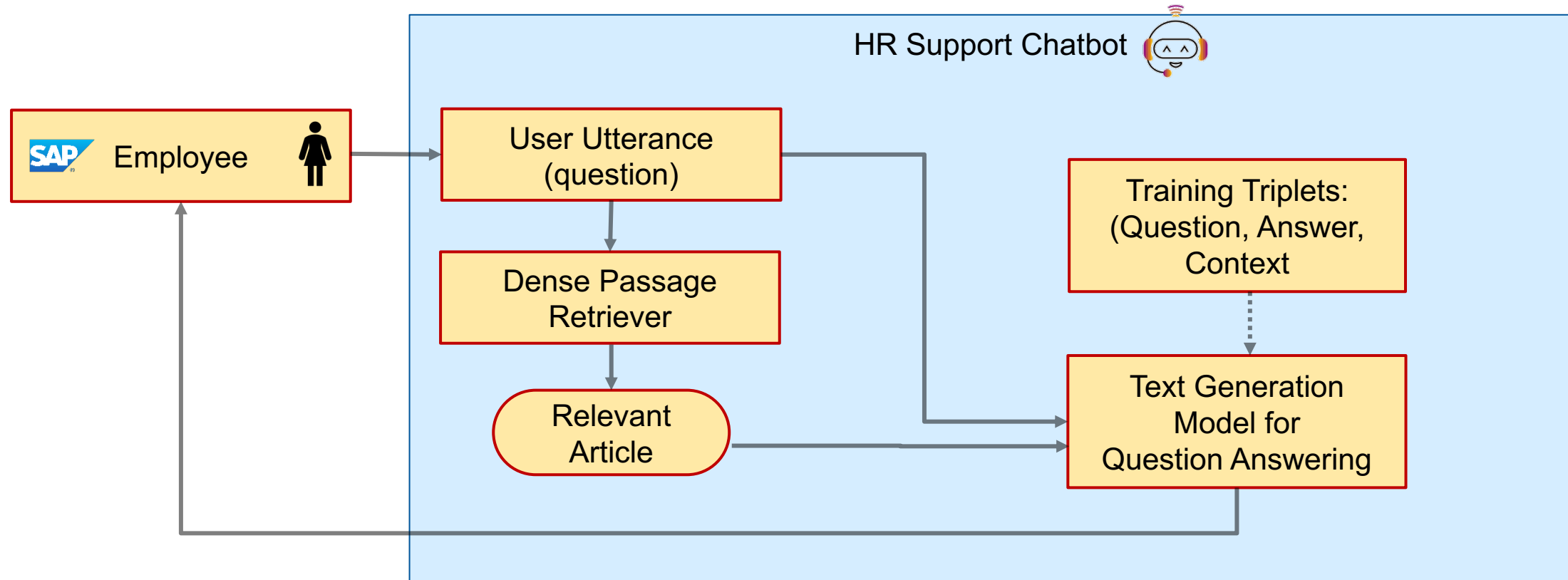- Natural Language Generation to cover edge cases and generate chat-like answers

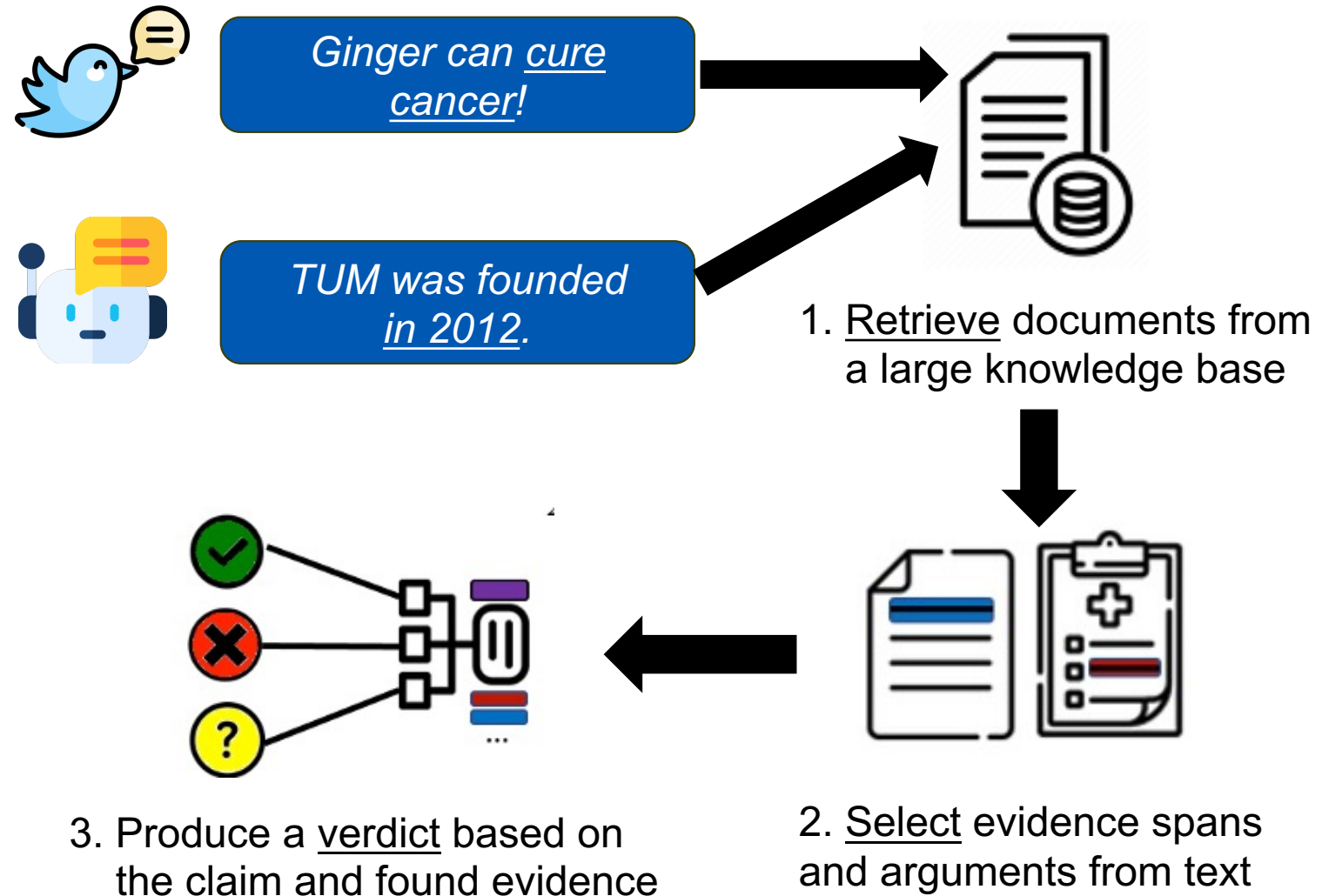# Project Overview

Previous Solution
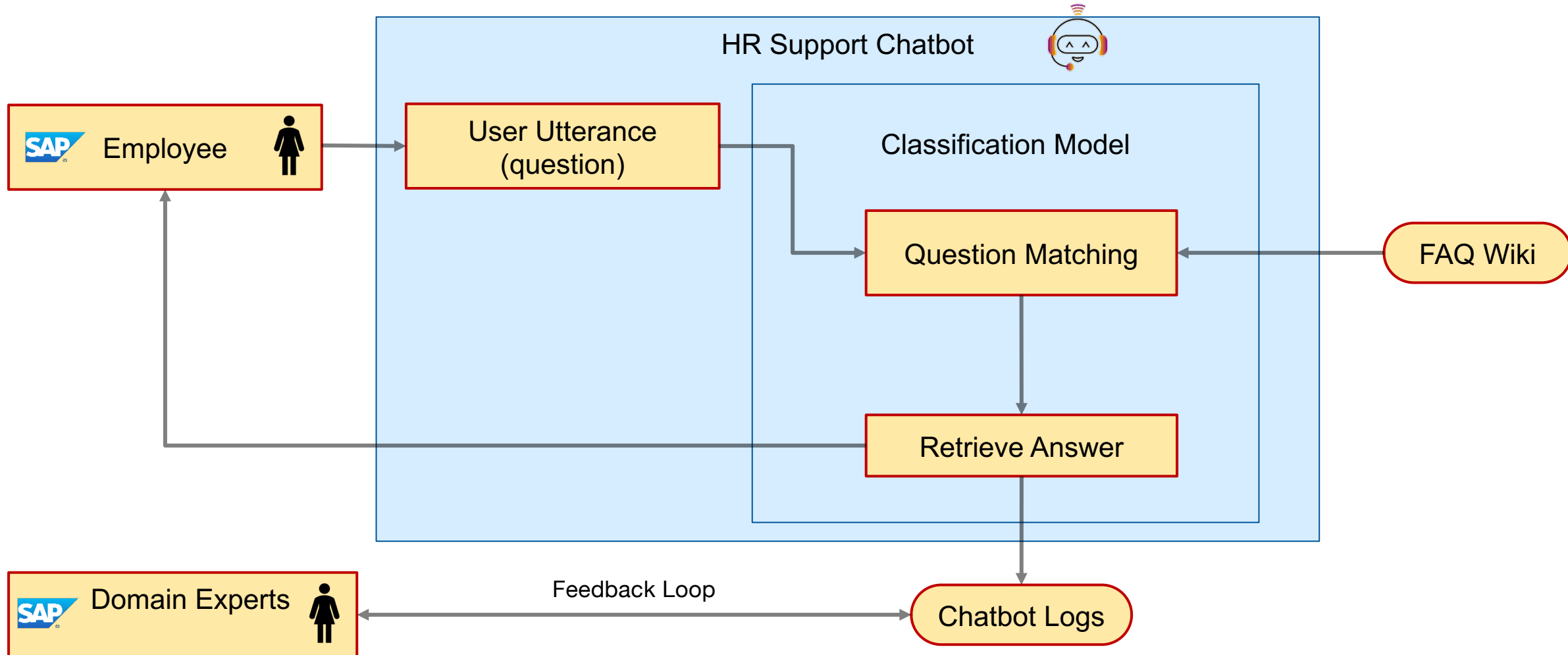
Project Overview

Proposed Solution

NLP Methods for Automated Fact-Checking and Factual Generation

- **Fact-checking** is the task of assessing the veracity of a claim based on background evidence
- **User-written content** → online misinformation detection
- **Model-generated content** → LLMs generate coherent text that can be factually incorrect

- Improved factuality in LLMs:
  - Structured prompts
  - Knowledge-grounding to external sources (knowledge graphs, ontologies, documents…)
  - Factually aware pre-training
  - Generation controlability with parameters like temperature

Ginger can cure cancer!

TUM was founded in 2012.

1. Retrieve documents from a large knowledge base

2. Select evidence spans and arguments from text

3. Produce a verdict based on the claim and found evidence

Supporting the Informed Adoption of Privacy-Enhancing Technologies in the Process of Data Privacy Compliance

Alexandra Klymenko, 29.06.2023

Chair of Software Engineering for Business Information Systems (sebis)
Department of Computer Science
School of Computation, Information and Technology (CIT)
Technical University of Munich (TUM)
wwwmatthes.in.tum.de

Privacy-Enhancing Technologies

Privacy-Enhancing Technologies (PETs) "protect privacy by eliminating or reducing personal data or by preventing unnecessary and/or undesired processing of personal data, all without losing the functionality of the information system"

Homomorphic Encryption
Performing computation on encrypted data without the need for decryption

Differential Privacy
Adding noise to a dataset so that it is impossible to reveal information about any specific individual

Secure multiparty computation
Spreading data analysis across multiple parties such that no individual party can see the other parties' data

Federated Learning
Training models collaboratively on decentralized devices or servers while keeping the private data locally

Zero-knowledge Proofs
Proving the knowledge of a value to another user without revealing the value itself

Motivation

TLTI

> ## Support the informed adoption of Privacy-Enhancing Technologies in the process of data privacy compliance

Data Privacy Concerns

- An ever-increasing amount of personal data is being collected, shared and analyzed
- Governments, researchers, and businesses use sensitive data to facilitate research or create business value
- Managing and sharing sensitive data that contains PII involves ethical, legal, and technical aspects

Regulatory Compliance

- Organizations are required to handle data in strict compliance with data protection laws such as GDPR
- According to GDPR, "appropriate technical measures" for data protection must be implemented

Privacy-Enhancing Technologies

- Privacy-Enhancing Technologies (PETs) have emerged to tackle the technical side of privacy compliance
- Many organizations are still reluctant to use modern PETs
  - Main reasons (among others): *complexity*, *lack of understanding*, *awareness*, and *incentive*

The Privacy Compliance Structure

- There is a multitude of roles involved in the privacy compliance process

- Amongst the main categories:
  - Management
    - *CISO, CTO, privacy manager,…*
  - Development
    - *Software engineers, architects,…*
  - Legal
    - *Privacy lawyers, compliance officers*
  - Go-betweens
    - *Data Protection Officer, privacy engineer*

- Different roles in the compliance process have distinct responsibilities, interests, and concerns
- The task of increasing the adoption of PETs should take into consideration such diversity

Target User Groups for the Research

TLTl
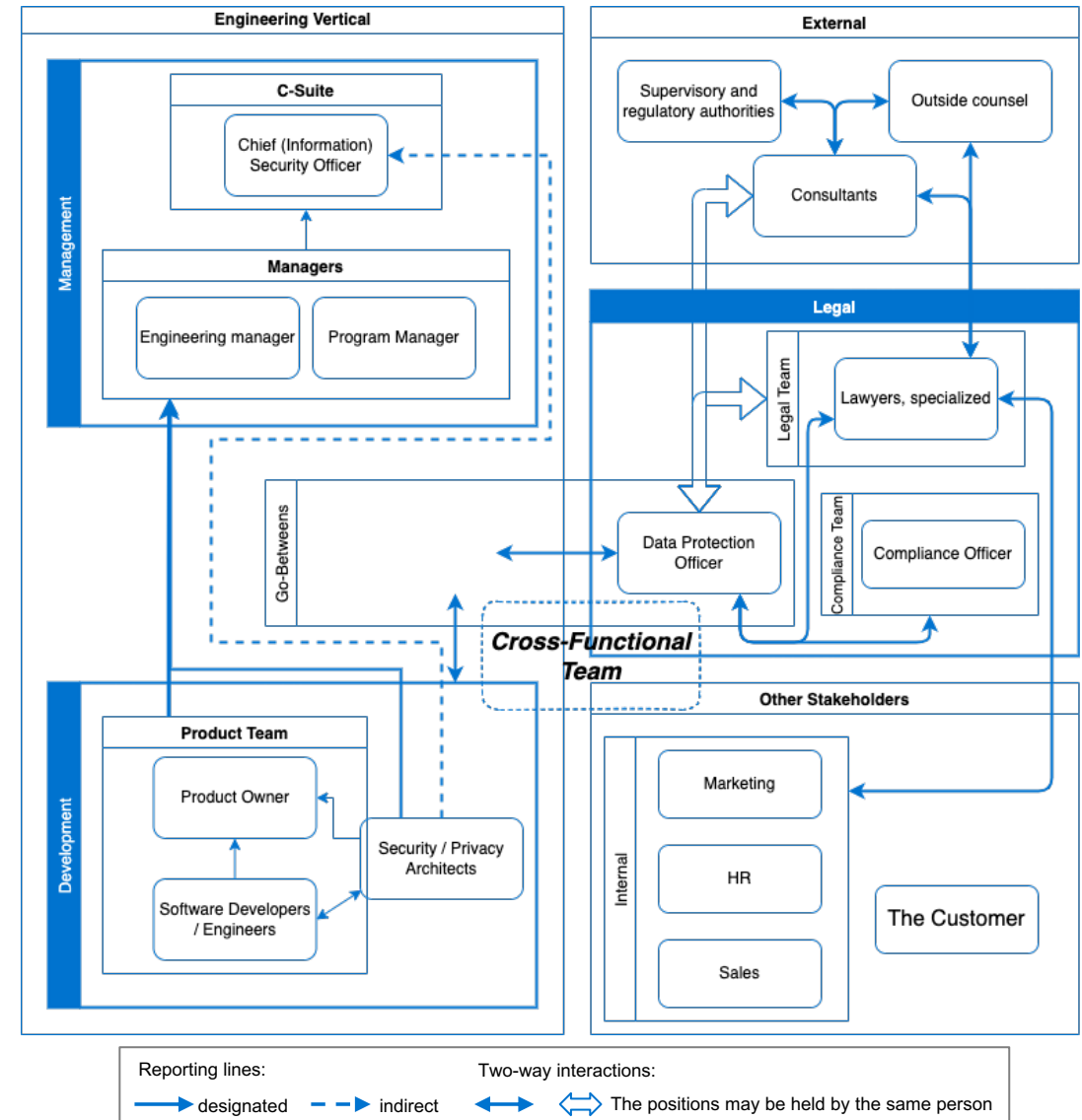
Management

"The decision to interpret PETs as appropriate technical measures is one left to upper management".

"Decisions making [regarding compliance] come from a couple levels above", i.e. management.

⬇

How can management's **awareness** on privacy topics and **motivation to adopt** PETs be raised?

Legal

"There is a different type of language that lawyers speak, than the technicians speak, we are always not 100% sure, do we understand each other?"

"It's really an interdisciplinary challenge…"

⬇

How can PETs be explained to **ease** legal-technical **interaction**?

Development

"The technology is very complicated… this is a very important issue that the most of the PETs are very hard to use in practice."

"All technical measures are worth nothing, if they are not implemented right"
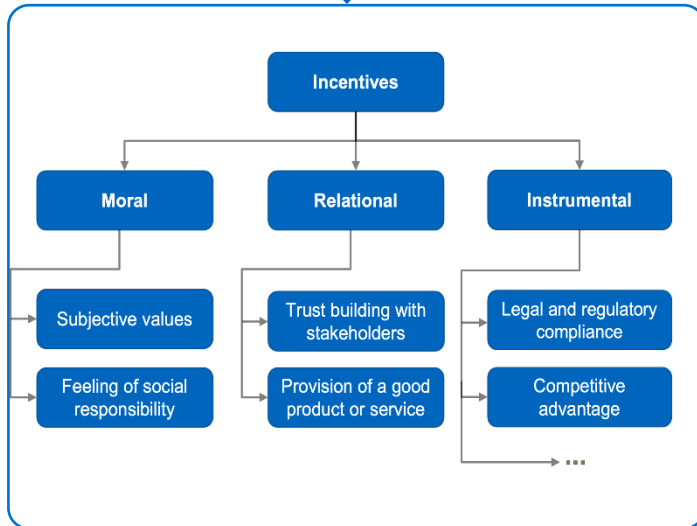
⬇

How can development teams be supported in **choosing & implementing** appropriate PET?

Target User Groups for the Research

Management

How can management's **awareness** of privacy topics and **motivation to adopt** PETs be raised?

- Incentives
  - Moral
    - Subjective values
    - Feeling of social responsibility
  - Relational
    - Trust building with stakeholders
    - Provision of a good product or service
  - Instrumental
    - Legal and regulatory compliance
    - Competitive advantage
    - …

Legal

How can PETs be easily explained to **ease** legal-technical **interaction**?

- Provide a detailed definition of PET and discuss its unique aspects and characteristics
- Present the benefits and challenges associated with the PET
- Examine the broad range of industries and application areas
- Briefly explain of the characteristics that make a problem ideal for this PET
- Provide an example of how this PET could have prevented a real privacy breach
- Showcase two successful real-world implementation examples of this PET

Development

How can development teams be supported in **choosing & implementing** appropriate PET?

Is statement s0 true for your data processing scenario?
- Yes
  - Is statement s1 true for your data processing scenario?
    - Yes → … Differential Privacy
    - No → … Federated Learning
- No
  - Is statement s2 true for your data processing scenario?
    - Yes → … Zero-Knowledge Proof
    - No → … Secure multi-party computation
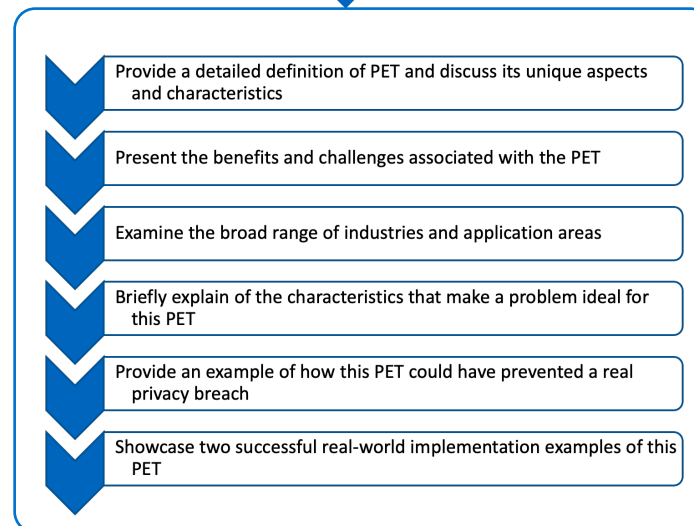
Target User Groups for the Research

Management

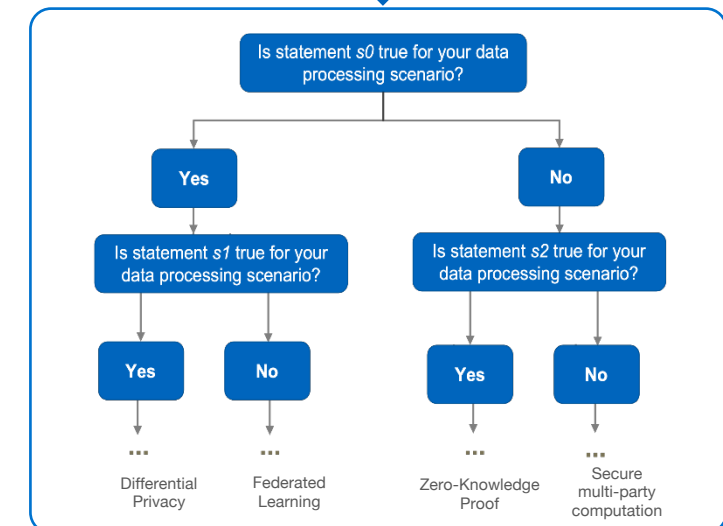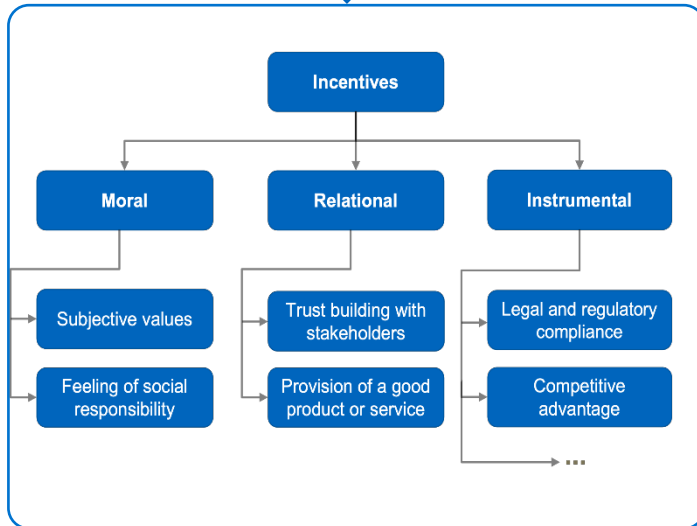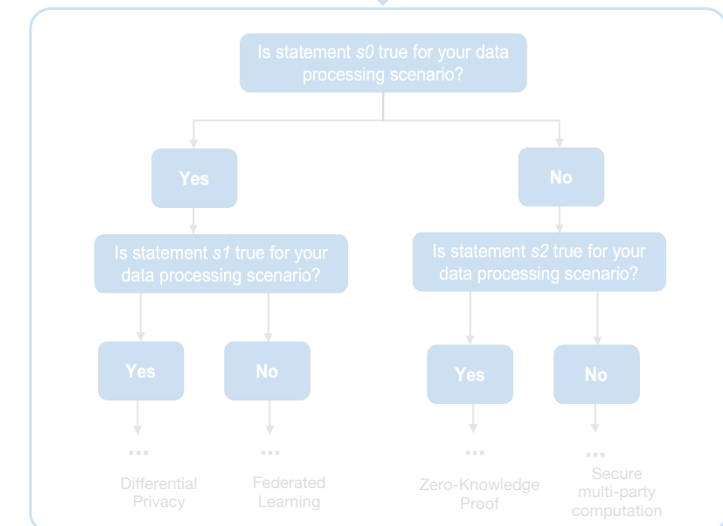How can management's **awareness** of privacy topics and **motivation to adopt** PETs be raised?

Legal

How can PETs be easily explained to **ease** legal-technical **interaction**?

- Provide a detailed definition of PET and discuss its unique aspects and characteristics
- Present the benefits and challenges associated with the PET
- Examine the broad range of industries and application areas
- Briefly explain of the characteristics that make a problem ideal for this PET
- Provide an example of how this PET could have prevented a real privacy breach
- Showcase two successful real-world implementation examples of this PET

Development

How can development teams be supported in **choosing & implementing** appropriate PET?

Management Perspective: Adoption of PETs as a CSR Initiative

TLTT

Problem

Implementation of PETs requires a clear organizational incentive to do so

"Obvious" reason:
the demonstration of compliance to avoid fines

However:
data protection laws and regulations do not explicitly require the use of advanced state-of-the-art technologies such as PETs

$\Rightarrow$ ***What might motivate managers to surpass the bare minimum and invest in PETs?***

Approach

Corporate Social Responsibility (CSR):
the firm's responses to issues *beyond* its narrow economic, technical, and legal requirements in order to achieve social benefits in addition to the traditional economic gains that it seeks [5]

$\Rightarrow$ Privacy can be categorized as CSR initiative
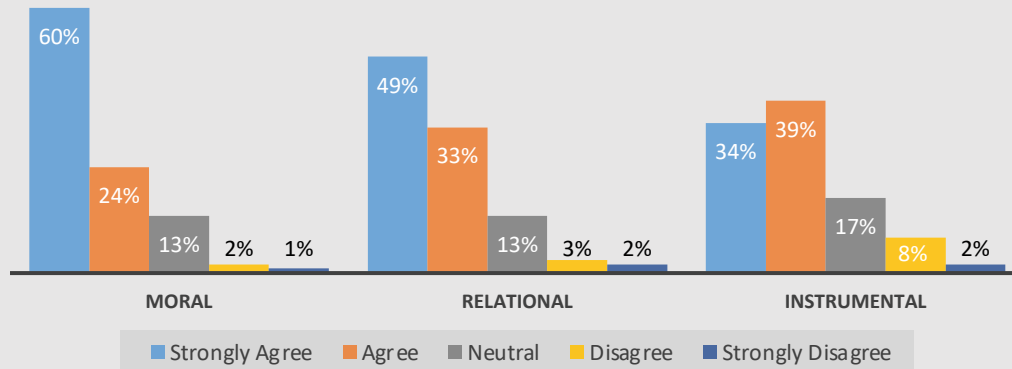$\Rightarrow$ The adoption of PETs can be categorized as a CSR activity

Results

A taxonomy comprising a total of **54 incentives** to adopt PETs, divided into 3 main categories:

1. **Moral** reasons determined by morality-driven values
2. **Relational** reasons driven by the company's concern about stakeholder relationships
3. **Instrumental** reasons driven by corporate self-interest

Management Perspective: Adoption of PETs as a CSR Initiative

TLT
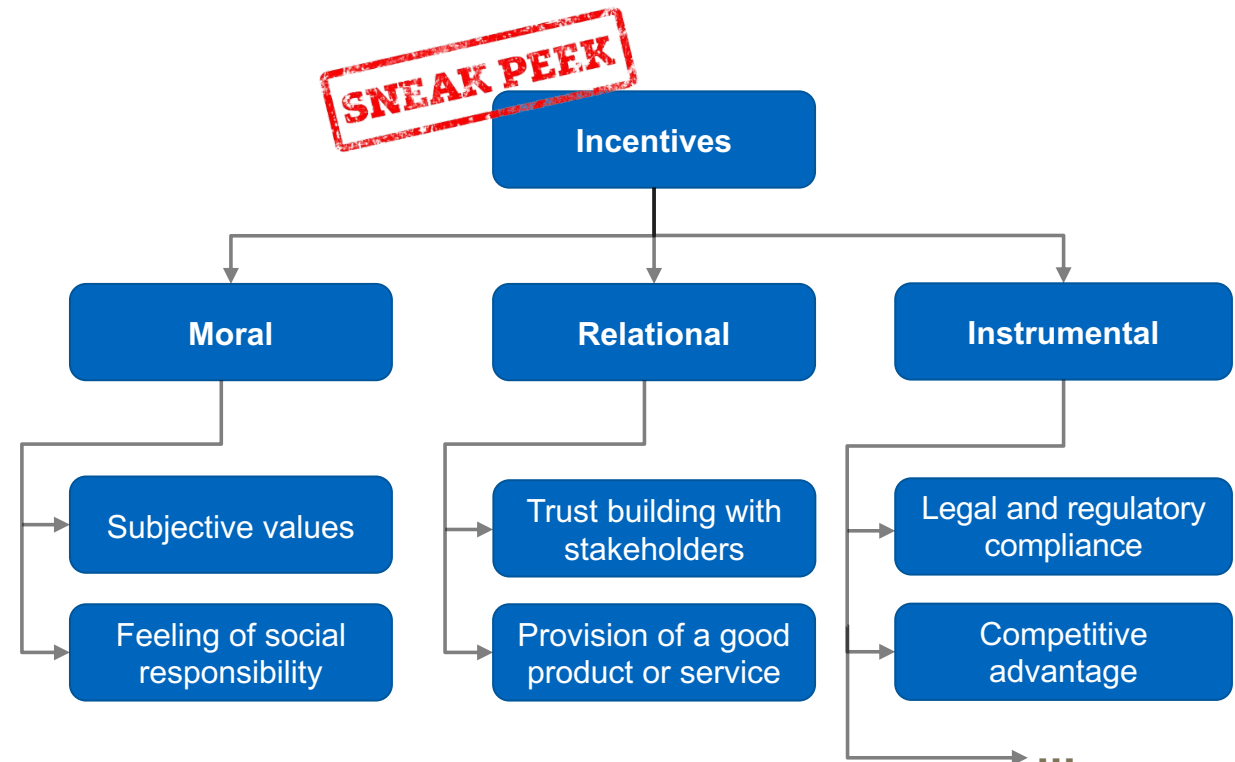
Agreement at the Category Level

Agreement with the Categories

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree

Agreement with the
- Moral Category (84%),
- Relational Category (82%),
- Instrumental Category (73%);

Despite what one might initially assume:
⇒ At the category level, the decision-makers are *more strongly motivated* to adopt PETs by the ethical implications of privacy rather than by enhancing key stakeholder relationships or solely serving the company's self-interest

Agreement at the Subcategory Level

Top Six Most-Agreed-Upon Subcategories

- Strongly Agree
- Agree
- Neutral
- Disagree
- Strongly Disagree
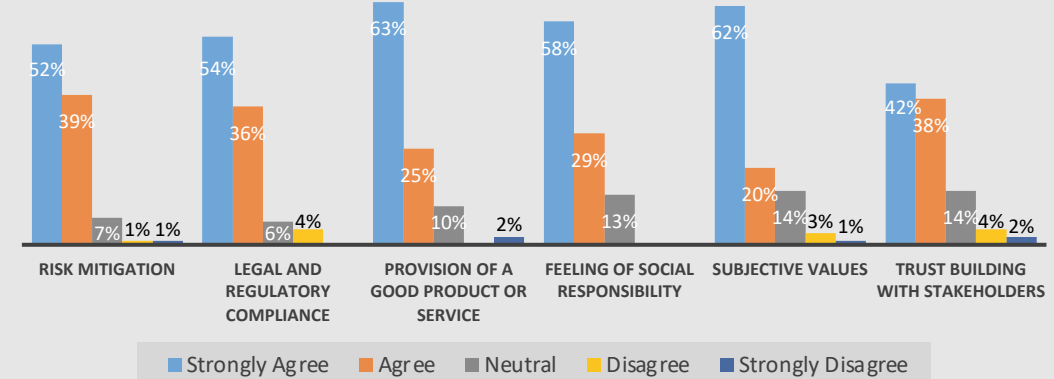
- Exactly two *moral*, two *relational*, and two *instrumental* subcategories among the top six most-agreed-upon subcategories of incentives

"Merely projecting privacy as an altruistic nice-to-have is a missed opportunity to deliver tangible benefits for your business."

Alexandra Klymenko

Technical University of Munich (TUM)
TUM School of CIT
Department of Computer Science (CS)
Chair of Software Engineering for Business
Information Systems (sebis)

Boltzmannstraße 3
85748 Garching bei München

+49.89.289.17114
alexandra.klymenko@tum.de
wwwmatthes.in.tum.de

Thank you for your attention!

Time for a poster session & get-together with food and drinks.