

Reference: QSE:SE4AI

Topic: SE4AI: Software Engineering for AI Workflows ([dataTUdiscovery](#))

Course Type: Bakk-/Master Thesis, ASE or CS Project

Start: As soon as possible

End: To be defined

Contact: Kristof Meixner (kristof.meixner@tuwien.ac.at)

Stefan Biffl (stefan.biffl@tuwien.ac.at)

Background

In data-driven AI engineering, machine learning models depend entirely on the quality, structure, and accessibility of input datasets. In empirical science domains across TU Wien research groups, valuable experimental data remains severely fragmented. Currently, research data is trapped in isolated spreadsheets, unstandardized CSV files, and proprietary binary outputs generated by lab instruments. Data transfer between researchers relies on manual ad-hoc methods, while transformation depends on unmaintained legacy scripts. Crucially, missing metadata standards, insufficient automated verification workflows, and unvalidated inputs create "data debt", resulting in insufficient machine learning readiness and forcing domain researchers to waste significant time on manual data cleanup. The project *dataTUdiscovery* [1] aims at improving this situation.

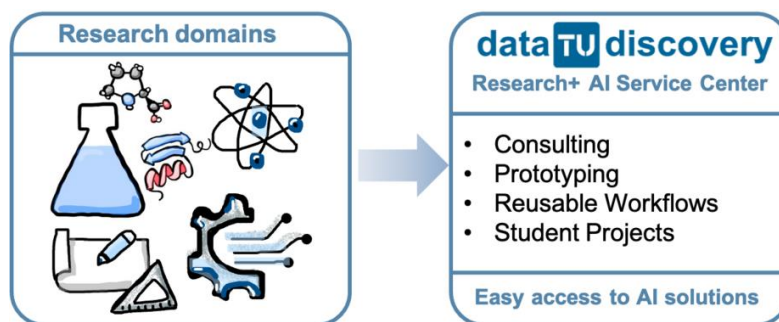


Figure 1: SE solutions for AI workflows to strengthen research at TU Wien.

Goal

The central goal of this topic is to **bridge software engineering methods with AI workflows** for domain-specific scientific research by establishing automated, reliable, and FAIR-compliant (Findable, Accessible, Interoperable, Reusable) data management pipelines (dataTUdiscovery [1]). By applying rigorous software engineering practices to domain data pipelines, this project aims at replacing spreadsheet chaos with structured knowledge repositories designed for AI model training.

Key Value Proposition

- *Data Sovereignty* by maintaining complete internal data hosting and ownership at TU Wien.
- *AI Readiness & Efficiency* to reduce redundant laboratory testing, save valuable researcher time, and establish standardized data streams to train predictive AI models.

Illustrative Use Cases

To ground these software engineering and information systems concepts in concrete applications, projects in this topic include (i) web applications to collect and validate metadata from chemical experiments as input to efficiently find promising materials to combine, (ii) simulation to efficiently plan experiments with limited resources, and (iii) software engineering for high-quality Jupyter notebooks for reuse (supporting the TU Cookbooks initiative).

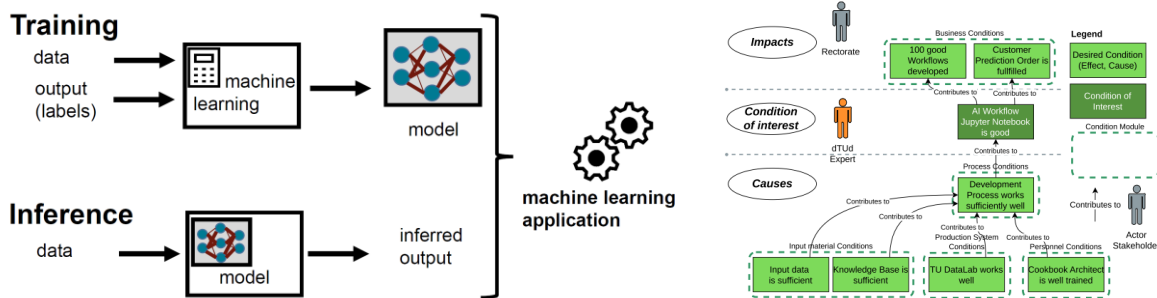


Figure 2: Application for AI workflows to optimize BioTech experimentation.

Tasks

Tasks will be adapted according to a specific selected project context; typical tasks include:

Requirements Engineering & Acceptance Criteria: Clarify stakeholder requirements and acceptance criteria using Goal-Question-Metric (GQM) frameworks [2] and Given-When-Then scenarios with a product owner or data science solution architect, e.g., a senior scientist in dataTUdiscovery [1].

Socio-Technical Process Analysis: Map and model human-in-the-loop workflows, data handoffs, and system dependencies using practical modeling methods (e.g., *procan.do Business Intelligence* [2]).

Architecture & Schema Design: Design modular system architectures integrating relational storage, time-series data handling, and microservice validation pipelines.

Full-Stack Implementation: Implement web-based input interfaces, automated file parsers, data verification microservices, and standardized export/import interfaces.

Quality Assurance & Verification: Conduct system tests, evaluate FAIR data compliance criteria, and validate system with domain users.

Documentation & Academic Thesis: Document system architecture, write technical documentation, and write the academic thesis following standard SE methodologies.

Expertise

For this topic, a set of skills is recommended (at least two are mandatory).

- Solid foundation in core Software Engineering principles and web application development.
- Working knowledge of relational databases and SQL.
- Interest in software engineering for AI/data pipelines and domain data management.

References

- [1] The TU strategic project [dataTUdiscovery](#). web link.
- [2] Biffl, S., Marhold, K., Musil, J., Meixner, K.: [Business intelligence for value-based process and systems engineering and operation with procan.do](#). In: Business Intelligence for Multi-Domain Systems Engineering and Operation. Springer Cham (2026). > Section 5.2.x
- [3] Pollinger M. (2026) SE and Modeling Approaches for Jupyter AI Workflows. BSc thesis, TU Wien.