Provenance Question-based AI Transparency & Accountable AI Governance

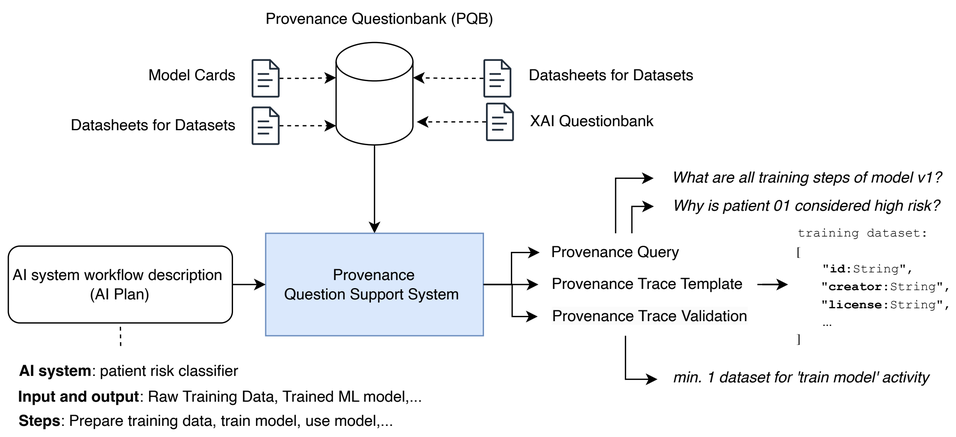
Laura Waltersdorfer[1,2], Dominique Hausler[3], Tanja Auge[3]

[1]Vienna University of Economics and Business, Austria
[2]University of Technology Vienna, Austria
[3]University of Regensburg, Germany

Objective & Contributions

Implementing technical governance and transparency in AI is hard:

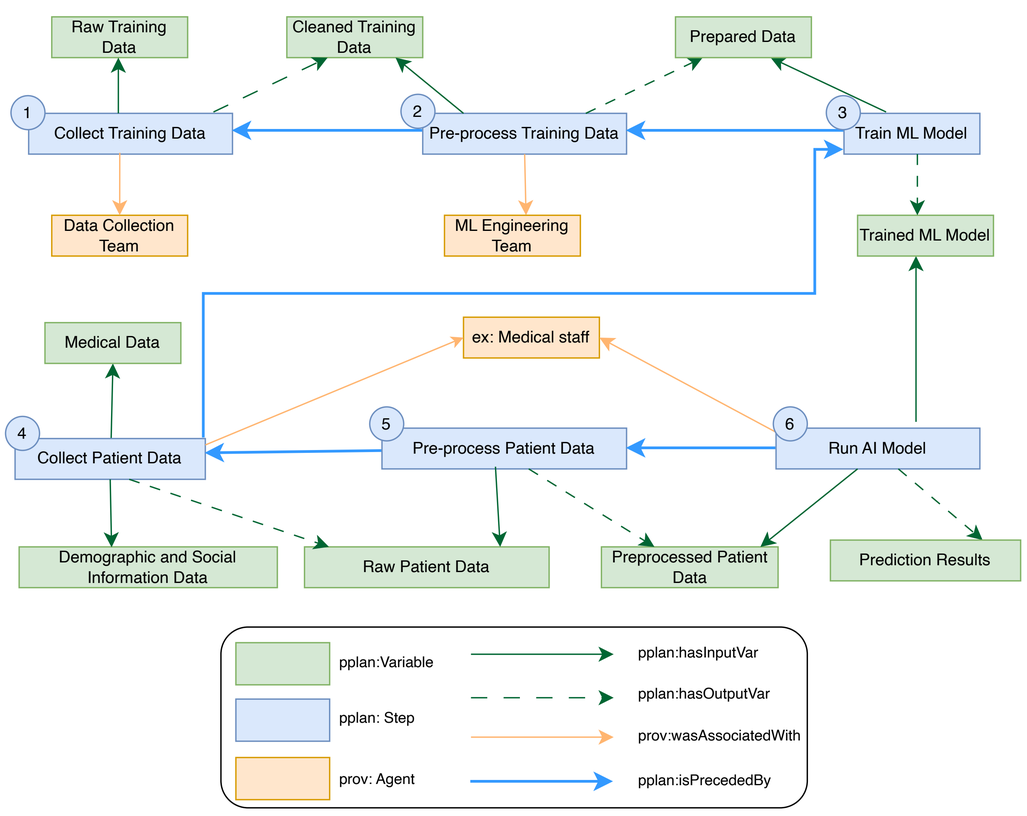
- **Challenges**: Vague requirements, missing know-how and time resources
- **Contribution**: Conceptual design for a *Provenance Question* (PQ) support framework
- **Aim**: Provide support for deriving provenance requirements



Example Scenario

A hospital uses an AI system to predict health-adverse effects based on high-risk patient's medical data and demographic-social information. Relevant PQs to increase the explainability of the AI system might be:

- *PQ1: What was the training process for the model?*
- *PQ2: Why is this patient considered to be high risk?*



Example Question	Answer type
Q_2: Who was involved in generating the data ?	agent, $\hookrightarrow Q'_2$
Q_8: What is the source of the training data ?	$\hookrightarrow Q'_8$ or Q''_8
Q'_2: Who generated the data ?	agent
Q'_8: Where does the training data come from ?	location
Q''_8: What kind of data was the system trained on ?	—

Table 1:Example PQs of different structures including a question word, a (main) verb, and a subject and/or a phrase. If necessary, additional conditions and refinement can be added and PQs can be rewritten ($\hookrightarrow$).

Provenance Question Analysis

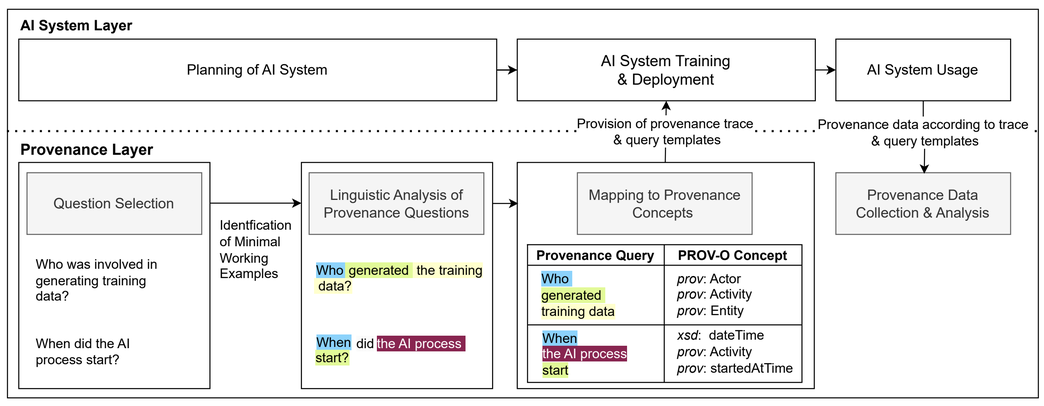
- **Question Selection.** Suitable PQs need to be selected from available sources (questionbank).
- **Linguistic Rewriting to Provenance Questions.** The rewriting consists of 2 steps: 1) Linguistic features *(question word, subject,..)* are analysed. 2) Provenance level between simple and complex PQ.
- **Mapping to Provenance Concepts.** Question answers are mapped to ontological provenance concepts (PROV-O) for adequate answering.

	Q_2			Q'_2		
	Who	was involved	in generating the data?	Who	generated	the data?
Linguistic L.	S	P	prepositional phrase	S	P	accusative object (NP_{Akk})
Provenance L.	question word	main verb	object of interest	question word	main verb	object of interest

Table 2:Example of provenance question rewriting.

Conceptual Reference Architecture

- **Question Selection.** Users can either load existing or provide custom PQs in natural language.
- **Linguistic analysis of provenance questions.** Selected PQs are analysed, using *Natural Language Processing (NLP)* and W7 model.
- **Mapping to provenance concepts.** Question components are mapped to P-Plan and PROV-O concepts, or other applicable ontologies.
- **Provenance data collection & analysis.** Finally, incoming trace data is integrated and stored for retrieval.



Acknowledgement

Laura Waltersdorfer is funded by the European Union's Horizon research grant 101120323;
Dominique Hausler is funded by the German Research Foundation research grant 385808805.

Next steps

- Construct provenance questionbank (survey and literature review)
- Implement software prototype (linguistic question rewriting, mapping to provenance concepts)