Human-Centered Computing
Research Group

Freie Universität Berlin

«Human-Centered Data Science»

Exercise 11

Lars Sipos

Human-Centered Computing, Institute of Computer Science

Freie Universität Berlin

05.07.2022

Assignment 5 Presentation

LIME Introduction Discussion

» What are your generell impressions of LIME?

» How difficult was it to use LIME?

» What insights have you gained about the model?

» How useful is LIME for people without a data science background?

» What suggestions do you have as an improvement for LIME?

LIME Model Work Discussion

» Were the explanations helpful to you personally?

» Are you more confident in your knowledge about the model?

» How does LIME compare to SHAP in terms of interpretability?

» What is the *gist* of the model? How would you explain it to a non-expert?

» Have you tested other models and datasets? What are your insights?

Human-Centered Computing
Research Group

Freie Universität Berlin

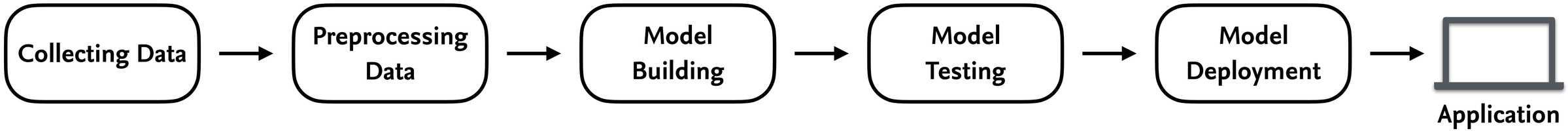
«Human-Centered Data Science»

Assignment 6

Designing an Explanation Interface

https://github.com/FUB-HCC/hcds-summer-2022/wiki/11_exercise_A6

Scope of Human-Centered Data Science



Human and the machine are equally important actors in carrying out data science.

A Participatory Process for Interpretability Techniques

WHAT to explain?

Expert Mental Model

What happens to the best of our knowledge?

What can be explained?

What does an expert mental model of the system look like?

User Mental Model

How do users currently make sense of the system?

What is the user mental model of the system based on its current UI?

How does it differ from the expert mental model?

Target Mental Model

Which key components of the algorithm do users want to be made transparent in the UI?

To what extent are users actually interested in the rationale behind the algorithms?

HOW to explain?

User Mental Model ⟷ Target Mental Model

How can the target mental model be reached through UI design?

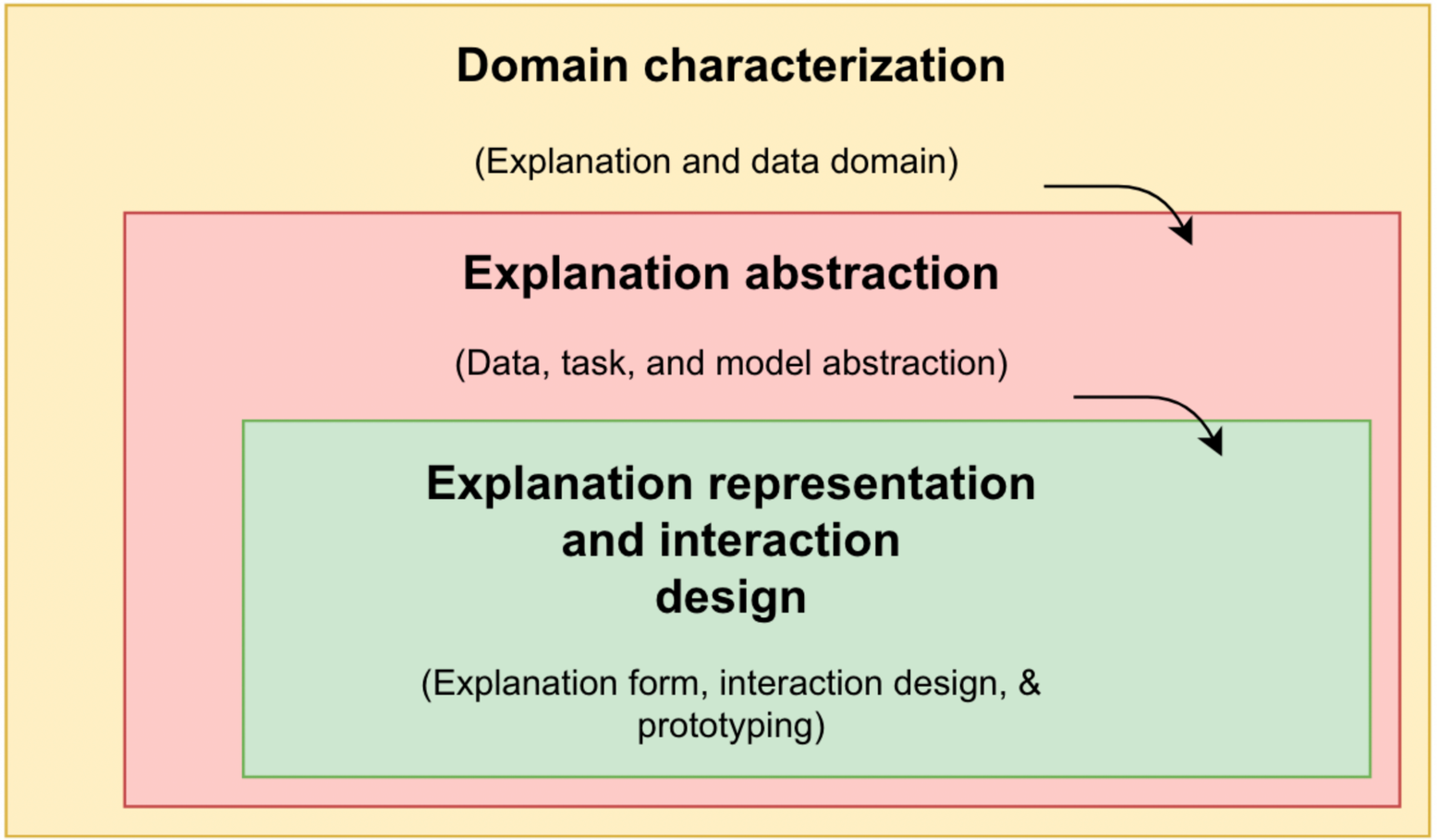
How and where can transparency be integrated into the UI of the system?

How has the users mental model developed?

Has the target mental model been reached?

Malin Eiband, Hanna Schneider, Mark Bilandzic, Julian Fazekas-Con, Mareike Haug, and Heinrich Hussmann. 2018. Bringing Transparency Design into Practice. In Conf. on Intelligent User Interfaces, 211–223.

Designing Explanation Systems



O. Anuyah, W. Fine, and R. Metoyer, "Design Decision Framework for AI Explanations," Mensch und Computer, p. 7, 2021.

Domain Characterization

Understanding the explanation system's domain and the needs of the target users.

Two questions should be answered:

I. Who are the target users?

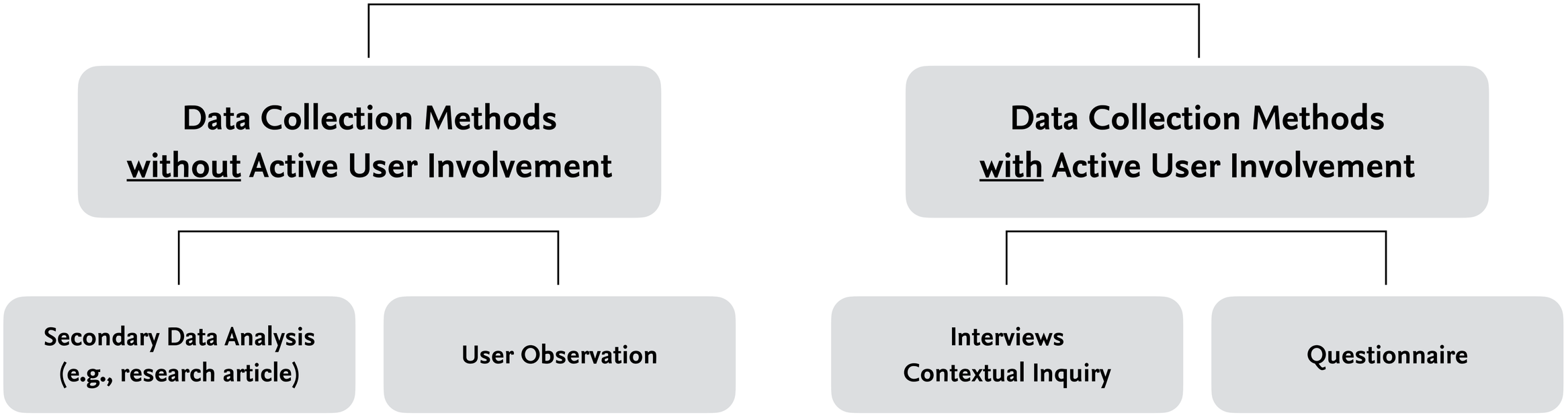
II. What questions do these users need answered based on not just the system, but also the general problem space the system is developed to address?

Users generally ask questions about:

- The **Data Domain** (data that is relevant to the problem space itself)
- The **Explanation Domain** (data pertaining to the system's decision or output)

O. Anuyah, W. Fine, and R. Metoyer, "Design Decision Framework for AI Explanations," Mensch und Computer, p. 7, 2021.

Eliciting Explanation Needs



Domain Characterization - Example

Domain: Trace link generation; **Target Users:** Domain experts in software traceability

User questions	Domain
What are the types of the two linked artifacts?	Exp
What concepts do the two artifacts have in common?	Exp
What concepts does the system consider in determining the trace links?	Exp
What does [concept X] mean?	Data
What is the semantic relationship between [concept X] and [concept Y]?	Exp
How confident is the intelligent system in the link prediction? [Predicted score]	Exp

O. Anuyah, W. Fine, and R. Metoyer, "Design Decision Framework for AI Explanations," Mensch und Computer, p. 7, 2021.

Explanation abstractions

Determine the intermediate generic/abstract tasks that the users should be able to perform with the explanation through:

Task Abstraction: Mapping the set of domain-specific questions obtained from users, to more abstract or generic tasks in the vocabulary of AI / ML.

Data Abstraction: Transform raw data for answering user questions into data types that useful for representing the explanations.

O. Anuyah, W. Fine, and R. Metoyer, “Design Decision Framework for AI Explanations,” Mensch und Computer, p. 7, 2021.

Domain Characterization - Example

Domain: Trace link generation; **Target Users:** Domain experts in software traceability

User questions	Domain	Abstract tasks
What are the types of the two linked artifacts?	Exp	**Identify** artifact types
What concepts do the two artifacts have in common?	Exp	**Compare** artifact content
What concepts does the system consider in determining the trace links?	Exp	**Identify** attention features
What does [concept X] mean?	Data	**Identify** data definitions
What is the semantic relationship between [concept X] and [concept Y]?	Exp	**Relate** features
How confident is the intelligent system in the link prediction? [Predicted score]	Exp	**Determine** relevance

O. Anuyah, W. Fine, and R. Metoyer, "Design Decision Framework for AI Explanations," Mensch und Computer, p. 7, 2021.

Domain Characterization - Example ctd.

Domain: Trace link generation; **Target Users:** Domain experts in software traceability

Data	**Abstract data type**
Artifact content	Qualitative (source code, structured text, descriptive text)
Concept definitions	Qualitative (text)
Concept relationships (**derived**)	Qualitative (semantic relationships) **[determined from ontology]**
Similarity score between the source and target artifacts	Quantitative
Concept importance	Quantitative
Link confidence score	Quantitative

O. Anuyah, W. Fine, and R. Metoyer, "Design Decision Framework for AI Explanations," Mensch und Computer, p. 7, 2021.

Explanation Representation and Interaction Design

Generating **explanation representations** and creating the **interaction design**.

» Identify explanation representations that are specially suited to identified data and tasks.

» Design the interaction between all explanation components in a way that is meaningful to the users and provides a good experience for them.

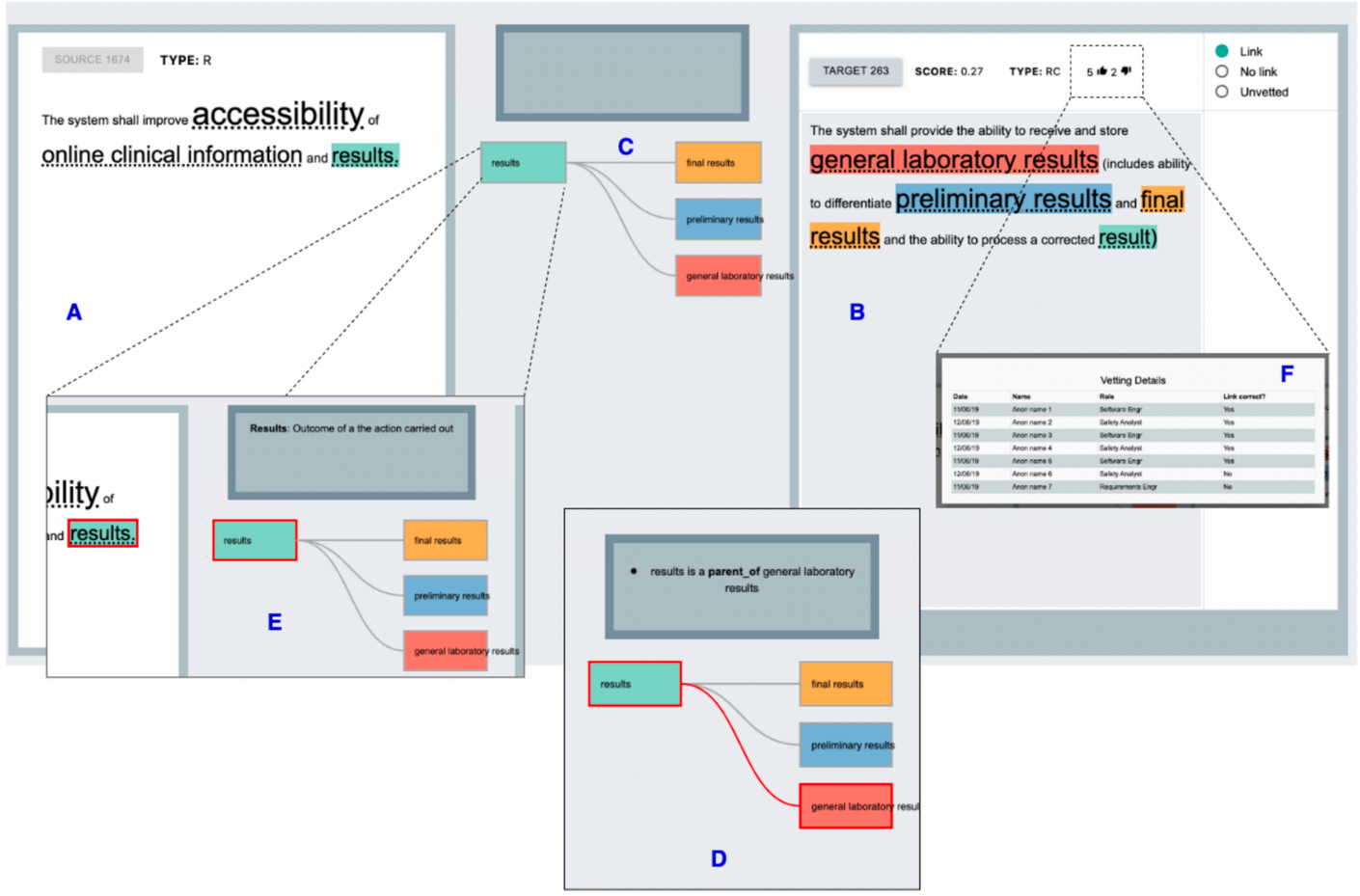
Ensure that the explanation representations are:

I. Comprehensible or understandable by the target users.

II. Answers their questions and/or supports their tasks.

O. Anuyah, W. Fine, and R. Metoyer, “Design Decision Framework for AI Explanations,” Mensch und Computer, p. 7, 2021.

Explanation Repr. and Interaction Design - Example



O. Anuyah, W. Fine, and R. Metoyer, "Design Decision Framework for AI Explanations," Mensch und Computer, p. 7, 2021.

Principles of Good Design

Aim of a good design is to minimize the gulfs of execution and evaluation.

In order to do this the design should

- Help the user build the correct conceptual model of the system
- Make the right parts visible
- Provide memory aids to the user
- Provide good feedback
- Accommodate errors

Norman, D. (2013). The design of everyday things: Revised and expanded edition. Basic books.

Next Time

you will have …

1. actively participated in the lecture
2. worked on the sixth programming assignment

Have fun!