

AI TOOLS | Understanding Learning Through Pipeline

Transparency:

A Thinking-Aloud Study of the Decision-Tree Transparency Platform



MOBOTS Lab – Prof. Francesco Mondada

Project Type: Semester project for master students / Master Thesis

Section: IC or STI or Digital Humanities

Assistants: Valentina Ferraioli
(Valentina.ferraioli@epfl.ch)

Context:

The previous semester project developed a web-based platform to study “Pipeline Transparency” in machine learning education, using a decision-tree task (predicting a dinosaur’s diet from field notes) across three conditions: C1 (Glass Box — manual data preparation and manual decision-tree construction), C2 (Gray Box — manual data preparation, automated model), and C3 (Black Box — pre-loaded data, fully automated model). This allows comparing whether uncovering data preparation, model logic, both, or neither, best supports students’ understanding of ML.

Comprehension outcomes can tell us whether a condition helps, but not why, which misconceptions students form along the way, or where in the pipeline their understanding breaks down. The goal of this follow-up project is to investigate this in depth: using Thinking Aloud protocols to observe students working through each transparency condition, and analyzing the resulting data to characterize how they reason about data and model transparency and what mental models they build of the ML pipeline.

Project Description:

This project builds directly on the platform and the three experimental conditions (C1/C2/C3) developed in the previous semester project; no new experimental condition needs to be designed. The student’s contribution is a qualitative study conducted on the existing platform, centered on questions such as: what reasoning strategies do students use when data preparation and/or model construction are manual versus automated? At which pipeline stage do misconceptions arise, and how does this differ across conditions? How do students justify their trust in the model’s predictions after each condition? The core tasks are:

- Protocol design: Design a Thinking Aloud protocol (concurrent verbalization, moderator probes, debrief interview) and extend the platform, where needed, with screen/audio recording synced to interaction logs.
- Data collection: Pilot and run individual sessions with ~12 student participants (4–5 per condition), each combining the platform task with think-aloud verbalization and a short debrief.
- Qualitative analysis: Transcribe sessions and perform thematic analysis (deductive codes by pipeline stage/condition, plus inductive codes) to identify recurring reasoning patterns and misconceptions.
- Synthesis: Compare themes across the three conditions and derive design and pedagogical recommendations for teaching ML transparency.