# Data extraction with LLMs to visualize human value models from deliberative transcripts

TU Delft

## Motivation

**Value-Focused Thinking (VFT)** is a decision-making approach that starts by identifying what people care about, i.e. their values, rather than first focusing on the available alternatives.


VFT is applied in several domains, including water management[1] and climate action[2].

Prior research uses LLMs to make participants reflect on their values in a public safety context, assisting in VFT.

We try to see if we can further assist this reflection process by creating visualizations to summarize each conversation.

To generate these visualizations, data needs to be extracted, so we also compare (locally run) LLMs in their ability to extract this data.

## Method

Two human coders first analyzed each transcript *independently* to code:

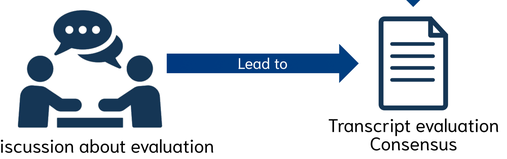
- Values identified in the transcript.
- Ordinal rank & 1-5 importance value of these values.
- Justifications for rank and importance assigned to each value.
- A summary of the value's meaning in the context of the transcript.

After this, the human coders discussed the differences and reached consensus on all identified items.

Then, 3 open-source locally runnable LLMs were used to extract the same data: **Qwen3.6:35b**, **Phi4-reasoning:14b**, and **Gemma4:e4b**.

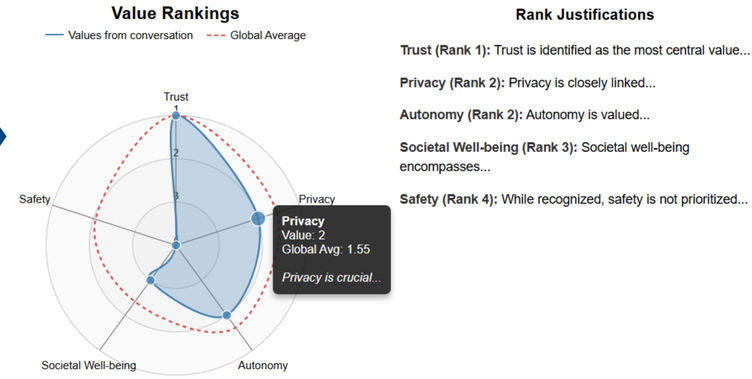
Gemma 4 Phi-4 Qwen3.6

Data extracted by the LLMs was compared to human evaluations of the transcripts, and to other LLM evaluations.




## Proposed visualizations

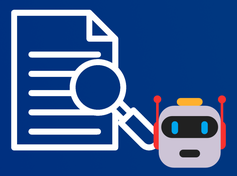
### 1. Radar charts




## 1. Chat


Participant and LLM have a conversation about the participants' values in a public safety decision-making context.

## 2. Analyze



Another LLM analyzes the conversation to extract data related to the human values present in the conversation.

## 3. Visualize

Using the extracted data, visualizations of the participants' "value model" can be created.

## Proposed visualizations (cont'd)

### 2. Value cards


## Results

**1. (Very) high agreement with human codes on which values are included in each transcript.** (Cohen's $\kappa \geq 0.808$)

Gemma 4 Phi-4 Qwen3.6

$\kappa$ = **0.848** $\kappa$ = 0.831 $\kappa$ = 0.808

**2. Moderate agreement with human codes on rank and importance.** ($0.536 \leq$ Weighted Cohen's $\kappa \leq 0.570$).

$\kappa$ = 0.553 $\kappa$ = **0.570** $\kappa$ = 0.536

**3. Textual Justifications are generally similar to those provided by humans**

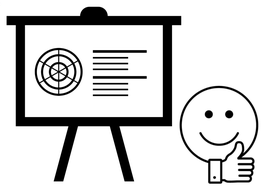
**4. Summarized value meaning is generally roughly similar. Less so compared to those provided by humans**

| Coder A | QP | QG | QH | PG | PH | GH |
|---|---|---|---|---|---|---|
| % 4 & 5 | 0.61842 | 0.6875 | 0.42105 | 0.7027 | **0.46479** | 0.4 |
| % 3-5 | 0.94737 | 0.925 | 0.71053 | 0.95946 | 0.66197 | **0.7125** |
| Coder B | QP | QG | QH | PG | PH | GH |
| % 4 & 5 | 0.56579 | 0.675 | 0.38158 | 0.71622 | **0.40845** | 0.4 |
| % 3-5 | 0.97368 | 0.9625 | **0.73684** | 0.94595 | 0.73239 | 0.725 |

Table 1: Share of similarity scores assigned by coder A and B to value summaries, excluding automatically assigned scores. Coder pairs who provided the summaries are shortened to Q, P, G and H. These refer to Qwen3.6, Phi4-reasoning, Gemma4, and Human coders respectively.

## Conclusion

**Based on literature, the proposed visualizations exhibit properties that should make them effective.** However, human evaluation should be performed to determine their effectiveness in practice.



**Local LLMs can reliably be used to identify human values present in text, and justify why these values are important or not**.
However, **they cannot fully replace human coders.** If an LLM were to output incorrect ranks or meanings, the visualizations will be showing the wrong data. This could potentially lead to a participant being misinformed about their own value model.

**(Locally run) LLMs can *assist* humans in coding transcripts**, with oversight.
For now, the task of interpreting *human* values remains a fundamentally *human* endeavor.

**Author:** Maximiliaan van der Veek (m.r.h.vanderveek@student.tudelft.nl) **Responsible Professor:** Willem-Paul Brinkman **Supervisor:** Michaël Grauwde

**Poster-specific references:**

**Research paper:**