

## Dimension Identification



Personality



Values



Self-Efficacy



Theory of Mind



Emotional Intelligence

### Principles for Dimension Identification



**Appropriateness:** identified by psychology literature as appropriate for depicting behaviors



**Meaningfulness:** has meaningful interpretation for large language models

## Dataset Design

### Data Sources



Psychometrics Test

- Big Five Inventory
- Cultural Orientation
- ...



Established Dataset

- Emotion Understanding
- False Belief Task
- ...



Scenario Curation

- Human-Centered Values
- AI Self-Efficacy Test
- ...



### Item Design



- ☐ Alternative-Choice
- ☐ Multiple-Choice
- ☐ Rating-Scale
- ☐ Open-Ended

### Prompt Design

**PROMPT**

#### [Instruction]:

You are a helpful assistant.  
You will be given ...

#### [Answer Rule]:

- Your Answer should be ...
- You are not allowed to ...

## Assessment & Validation



Internal Consistency



Parallel Forms Reliability



Inter-Rater Reliability



Option Position Robustness



Adversarial Attack Robustness

### Metrics:

- Standard deviation  $\sigma$
- Cohen's Kappa coefficient  $\kappa$
- Match rate (MR)
- Agreement rate (AR)