NTCIR-19 2025

FEHU: Fact-based Event-centric Human-value Understanding

Organizers:

Yao Wang[1], Haitao Yu[1], Xin Liu[2], Zhuochen Liu[1], Kyoungsook Kim[2], Adam Jatowt[3]

1. University of Tsukuba
2. National Institute of Advanced Industrial Science and Technology
3. University of Innsbruck

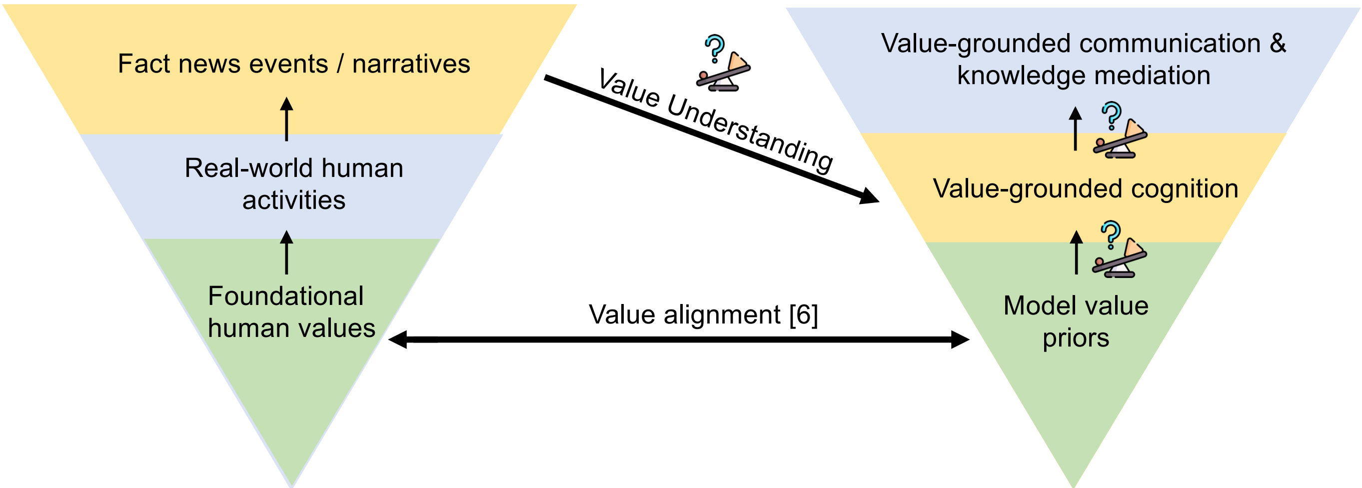
- Sample data schema (Continuously updating)
- Presentation slides
- Task & Contact Information

Background & Motivation

- Human values (HVs) are enduring principles that influence human behavior.
- LLMs exhibit bias in value-grounded cognition and value-consistent expression of Large Language Models (LLMs):
  - Understanding & Expression biases
    - Value hallucination
    - Value distortion
- **Goal**: Event-centric value understanding & value-consistent expression in fact news

News narratives express values in real-world activities

LLM value cognition drives value-consistent expression



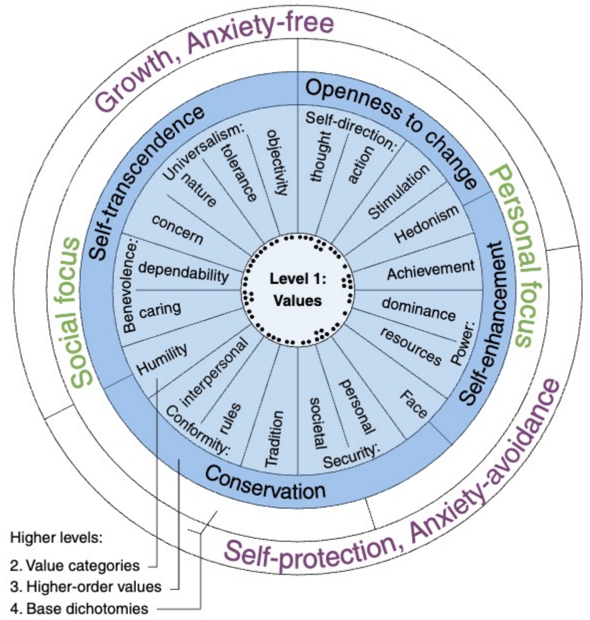
Level 2 Value Category	Level-1 Value
Self-direction: thought	Be creative
	Be curious
	Have freedom of thought
Stimulation	Have an exiting life
	Have a varied life
	Be daring
Achievement	Be ambitious
	Have success
	Be capable
Power: dominance	Have influence
	Have the right to command
	Have a good reputation

Examples of Human Value Categories

Task Overview

- **Main tasks**:
  - **Task-1: Human Value Classification:**
  - **Task-2: Value-aware Text Generation:**
    - **Task-2a: Value-aware text paraphrasing**
    - **Task-2b: Value-aware text summarization**
- **Participation**: Each task can be joined **independently**.
- **Adopted human value taxonomy**: Level-1 54 value categories [1-5] (e.g., Be responsible, Be honest, Have freedom of thought).
- **Evaluation**: organizers (Auto & Human-assisted evaluation)

Human value Taxonomy [1]



Task-1: Human Value Classification

- **Objective**: Identify **human values** expressed by each specific actor in a news article and avoid over-prediction or value distortion.
- **Provided data**:
  - Factual news **article**
  - List of main **actors**
- **Output**:
  - For each actor, predict a set of values with direction:
    - ***Aligned with human values***
    - ***Contradictory to human values***
- **Error definition**:
  - **Omission error:** missed gold labels
  - **Hallucination error :** predicted values not in gold labels
  - **Reversal error :** predicted opposite direction (aligned ↔ contradictory)

Essex lorry deaths: Man charged with immigration offence

A man has been charged with an immigration offence in connection with the deaths of 39 people in a lorry in Essex. Gazmir Nuzi was arrested in London on Thursday, according to Essex Police. The 42-year-old has been charged with conspiracy to facilitate the commission of a breach of UK immigration law by a non-EU person……

Actors:
UK Law Enforcement and Judicial System
Suspected Human Smugglers

Provided data

Output

Be just
Be responsible

Aligned human values

Actor: UK Law Enforcement and Judicial System

Have good health
Be just
Be responsible

Contradictory human values

Actor: Suspected Human Smugglers

Task-2a: Value-aware Text Paraphrasing

- **Objective**: Perform **sentence-level paraphrasing** while keeping original **semantics and human values of entire article** unchanged
- **Provided data**:
  - News article
  - List of main Actors
  - Specified paraphrasing type (each with ~100 samples):
    - Negation Switching | Entailment Extension | Opposite Polarity | Ellipsis / Compression
  - Sentence-level applicable transformations
- **Output**:
  - Rewritten article
- **Output Requirements**:
  - Faithfulness to original news & Natural and fluent news style
  - Must **preserve human value expressions**

Essex lorry deaths: Man charged with immigration offence

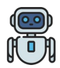
{[A man has been charged with an immigration offence in connection with the deaths of 39 people in a lorry in Essex.] [Gazmir Nuzi was arrested in London on Thursday, according to Essex Police.]}

Actors:
UK Law Enforcement and Judicial System
Suspected Human Smugglers

Target paraphrasing type:
Ellipsis / Compression

Sentence-level applicable transformations:
[“sentence_0”: [“Ellipsis / Compression”], …]

Provided data



Output

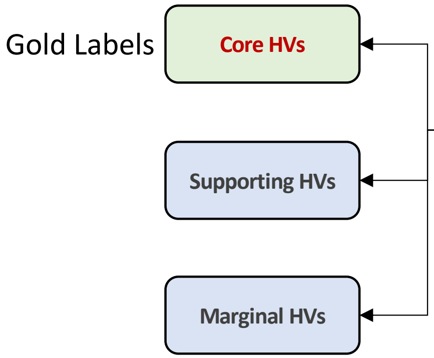
Essex lorry deaths: Man charged with immigration offence

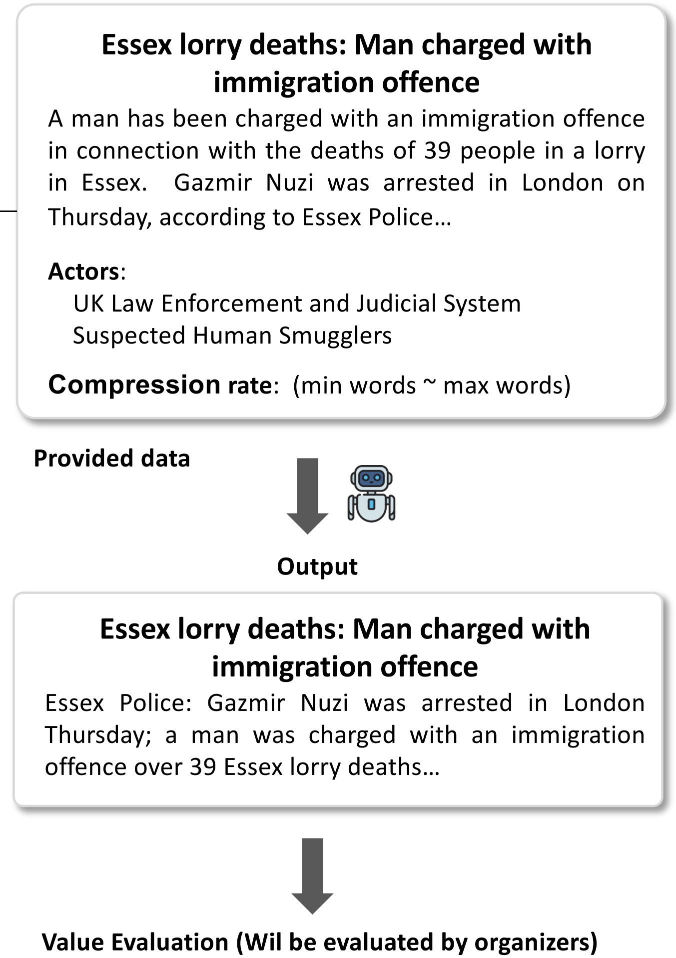
{["A man was charged with an immigration offence over 39 deaths in an Essex lorry.] [Essex Police said Gazmir Nuzi was arrested in London on Thursday.]}

Document-level value evaluation
(Wil be evaluated by organizers)

Task-2b: Value-aware Text Summarization

- **Objective**: Generate a concise news summary while preserving **core human values**.
- **Provided data**:
  - Full news article (document-level)
  - List of main Actors
  - Compression rate: (min_words ~ max_words)
- **Output**:
  - Generated summary
- **Output Requirements**:
  - Faithfulness to original news
  - Readable and fluent summary style
  - **Core value must be retained; omission will be penalized heavily.**
  - **Omit supporting/marginal values is accepted**, but their inclusion provides **small bonus rewards**.
  - No hallucination or value distortion





Evaluation-(1/3): Task-1 | Human Value Classification

- **Metric: Weighted F1**

$$Precision = \frac{Correct\ predicted\ values}{Predicted\ values + \lambda_2 \times Hallucination + \lambda_2 \times Reversal}$$

$$Recall = \frac{Correct\ predicted\ values}{Gold\ values} \qquad F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- **Error penalties: Hallucination** (λ_1 = 1.0) & **Reversal** (λ_2 = 2.0)

Direction error: ***aligned*** ↔ ***contradictory***	Reversal

Gold labels	Predicted labels (Not in gold labels)	Error Type
Only aligned HVs	Aligned labels	Hallucination
	Contradictory labels	Reversal
Only contradictory HVs	Contradictory labels	Hallucination
	Aligned labels	Reversal
Both	Either	Hallucination
Empty	Either	Hallucination

Distinguishing Hallucination VS Reversal Error

Evaluation-(2/3): Task-2a | Value-aware Paraphrasing

- **Main Evaluation:**
  - **Value Consistency Score** (VCS): Weighted F1 scores (same as Task 1)
    - Preserve **original human values** (avoid hallucination / reversal)
  - **Style-aware Paraphrasing Score** (SPS):

$$SPS = (\lambda_1 ER + \lambda_2 LRR + \lambda_3 SD) \times TS$$

    - **Type satisfaction** (TS): Whether the generated paraphrased document correctly achieves the required transformation type *t*.
    - **Degree of Paraphrasing:**
      - **Edit Rate** (ER): proportion of edit operations (insert, delete, substitute) between source and paraphrase.
      - **Lexical Rewrite Rate** (LRR): proportion of changed content words between source and paraphrase.
      - **Syntactic difference** (SD): proportion of differences in dependency/syntactic relations between source and paraphrase.

- **Supporting Evaluation:**
  - **Semantic Consistency** (BERTScore or QA-based checks)
  - **Fluency** (GPT-as-judge / perplexity)

Evaluation-(3/3): Task-2b | Value-aware Summarization

- **Main Evaluation:**
  - **Value Consistency Score**
    - Core Value**:** Weighted F1 score (same as Task-1)
    - Supporting Coverage

$$Coverage_{sup} = \frac{Hits_{sup}}{Gold_{sup}}$$

    - Marginal Coverage

$$Coverage_{mar} = \frac{Hits_{mar}}{Gold_{mar}}$$

    - Final Score

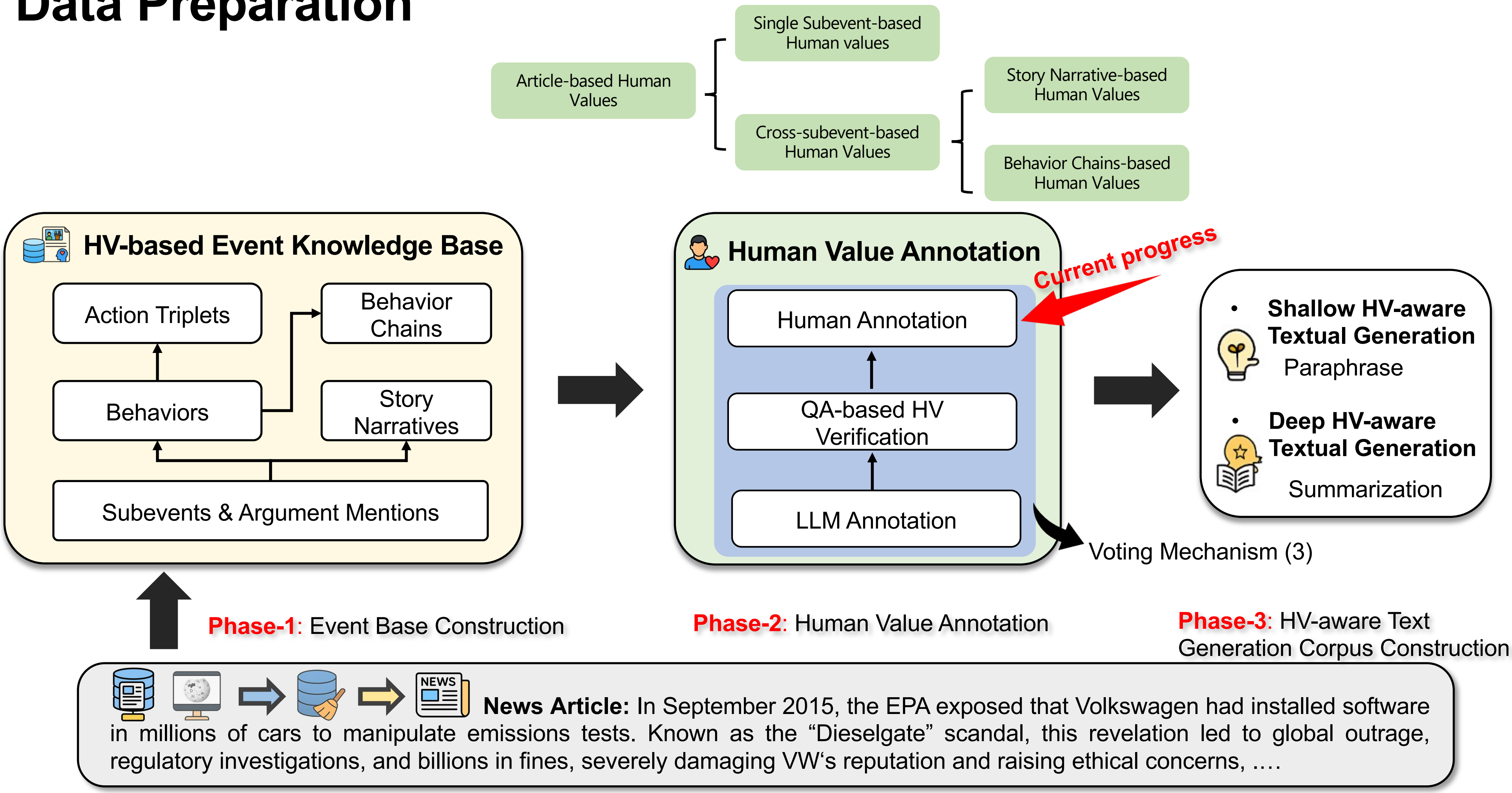
$$Final\ score = F1_{core} + \eta \cdot Coverage_{sup} + \mu \cdot Coverage_{mar}$$

Value Type	If retained	Penalty	Credit
Core	Yes	**No**	**Yes++**
	No	**Yes**	**No**
Supporting	Yes	**No**	**Yes+**
	No	**No**	**No**
Marginal	Yes	**No**	**Yes**
	No	**No**	**No**

Penalty and credit for different matches

- **Credit weight:** supporting (η) > marginal (μ)
- **Supporting Evaluation**
  - Semantic / Factual Consistency**:** BERTScore, QA-based checks
  - Fluency (GPT-as-judge / perplexity)
- **Text Similarity with Gold Reference (Optional):** ROUGE, BERTScore

Data Preparation



Expected Impact & Closing

- **Benchmark Contribution**
  - Build a more complete benchmark dataset for human value understanding through participants' contribution.
- **Evaluation & Application**
  - Encourage the community to propose **automatic value evaluation approaches** for LLM generation.
  - Integrate human values into model pre-training to enhance **value-sensitive content generation**, especially in news and media.

- **Closing**
  - Contribute to NTCIR-19's mission of evaluating **quality & social trustworthiness** in AI-generated content.

Important Dates

- **2025**
  - September: Partial Task Overview Release
  - November: Trial Data Release
- **2026**
  - January: Final Data Release
  - January: Dry run
  - March-July: Formal run
  - August 1: Evaluation results returned
  - August 1: Task Overview Release
  - September: Paper submission due
  - November: Camera-ready submission due
  - December: NTCIR-19 Conference

- **Sample data schema (Continuously updating)**
- **Presentation slides**
- **Task & Contact Information**

Coming soon：The detailed data schema will be released on the task website.

Thank you!

Reference

[1] Kiesel, J., Alshomary, M., Handke, N., Cai, X., Wachsmuth, H. & Stein,B. Identifying the human values behind arguments. Proceedings Of The 60th Annual Meeting Of The Association For Computational Linguistics (Volume 1: Long Papers) . pp. 4459-4471 (2022)

[2] Shalom H Schwartz, Jan Cieciuch, Michele Vecchione, Eldad Davidov, Ronald Fischer, Constanze Beierlein, Alice Ramos, Markku Verkasalo, Jan-Erik Lönnqvist, Kursad Demirutku, et al. 2012. Refining the theory of basic individual values. Journal of personality and social psychology, 103(4).

[3] Milton Rokeach. 1973. The nature of human values. New York, Free Press.

[4] Duane Brown and R. Kelly Crace. 2002. Life values inventory facilitator's guide.

[5] C. Haerpfer, R. Inglehart, A. Moreno, C. Welzel, K. Kizilova, Diez-Medrano J., M. Lagos, P. Nor- ris, E. Ponarin, and B. Puranen. 2020. World values survey: Round seven-country-pooled datafile.

[6] [5] Yao, J., Yi, X., Wang, X., Wang, J. & Xie, X. From Instructions to Intrinsic Human Values–A Survey of Alignment Goals for Big Models. ArXiv Preprint ArXiv:2308.12014 . (2023)