SemEval-2023 Task 4

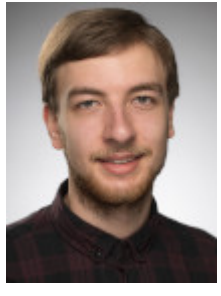
ValueEval: Identification of Human Values Behind Arguments

Johannes Kiesel[1]

Milad Alshomary[2]

Nailia Mirzakhmedova[1]

Maximilian Heinrich[1]



Nicolas Handke[3]

Henning Wachsmuth[2]

Benno Stein[1]

Bauhaus-Universität Weimar [1]

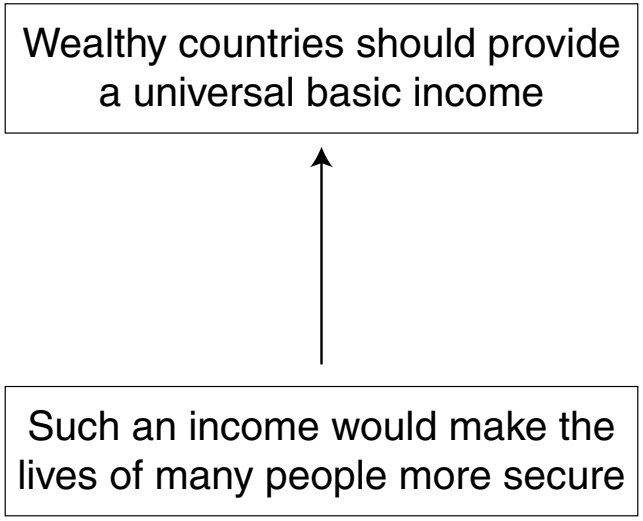
Leibniz Universität Hannover [2]

UNIVERSITÄT LEIPZIG [3]

Webis [1,2,3]

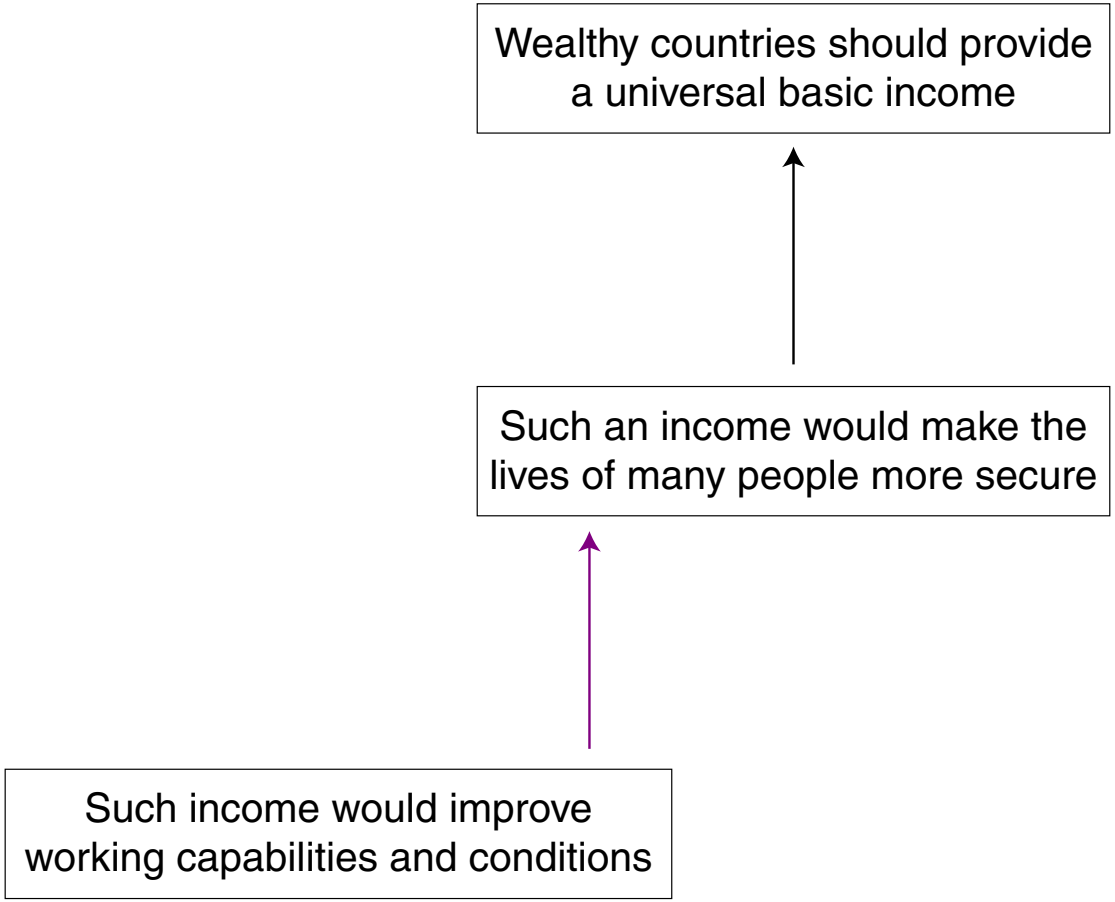
Task 4. ValueEval: Identification of Human Values Behind Arguments

Background



Task 4. ValueEval: Identification of Human Values Behind Arguments

Background

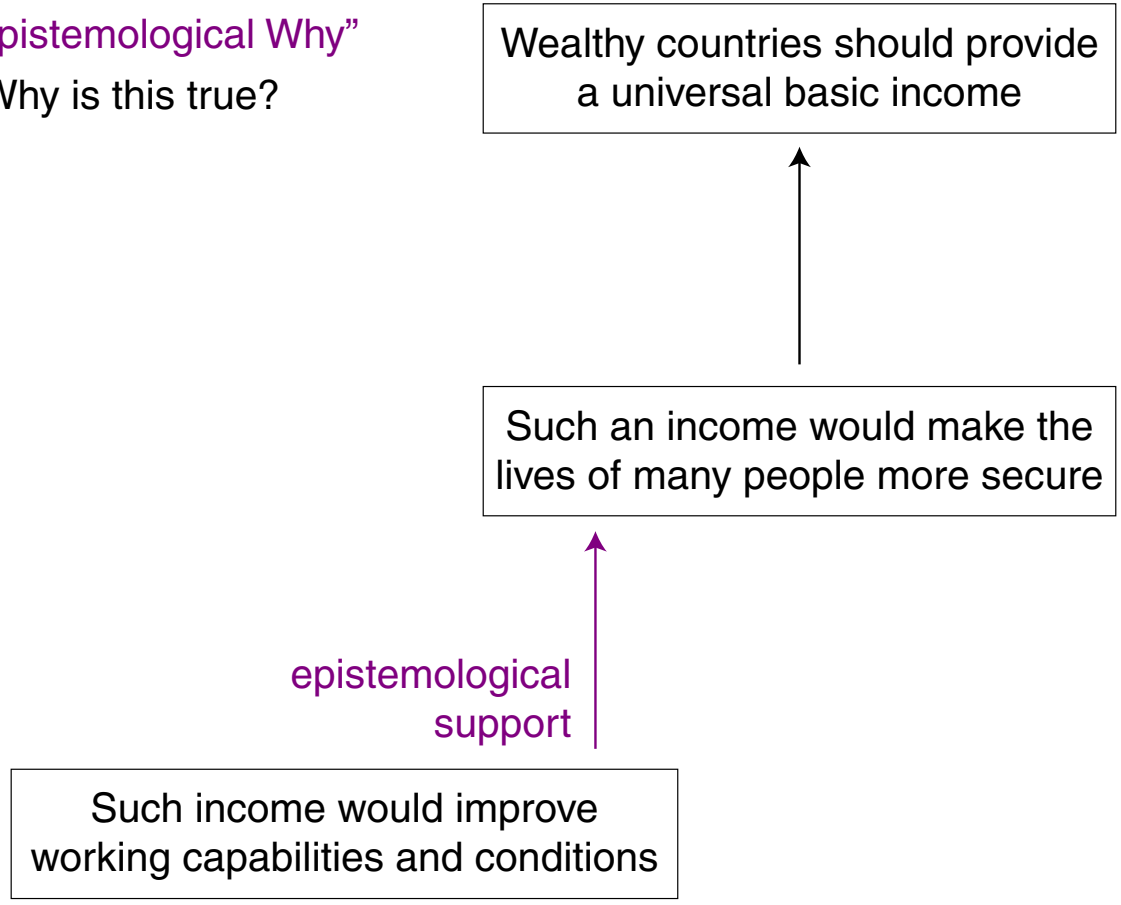


Task 4. ValueEval: Identification of Human Values Behind Arguments

Background

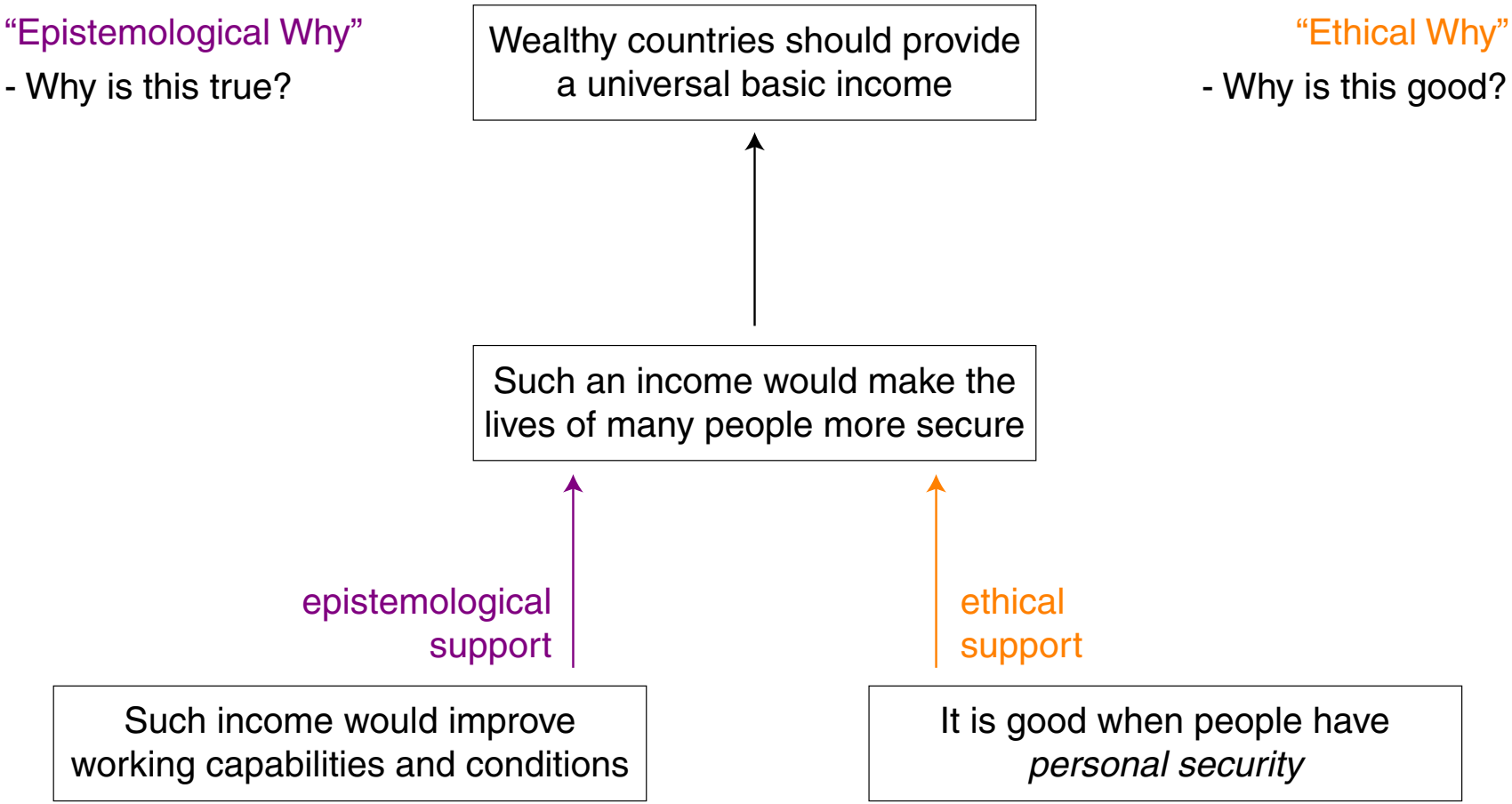
"Epistemological Why"

- Why is this true?



Task 4. ValueEval: Identification of Human Values Behind Arguments

Background



Task 4. ValueEval: Identification of Human Values Behind Arguments

Task Description

Task: Given a textual argument and a human value, classify whether the argument resorts to that value or not.

- Set of 20 values (categories)
- Derived from social science research, especially Schwartz' Taxonomy
- Multi-label classification

Task 4. ValueEval: Identification of Human Values Behind Arguments

The Touchè23-ValueEval Dataset

Argument source	Year	Arguments			
		Train	Validation	Test	$\sum$
Main dataset					
IBM-ArgQ-Rank-30kArgs (crowdsourced free-text)	2019–20	4576	1526	1266	7368
Conf. on the Future of Europe (democracy platform)	2021–22	591	280	227	1098
Group Discussion Ideas (debate points collection)	2021–22	226	90	83	399
$\sum$ (main)		5393	1896	1576	**8865**
Supplementary dataset					
Zhihu (QA forum posts)	2021	-	100	-	100
Nahj al-Balagha (religious text)	900–1000	-	-	279	279
The New York Times (newspaper editorials)	2020–21	-	-	80	80
$\sum$ (supplementary)		-	100	359	**459**
$\sum$ (complete)		5393	1996	1935	**9324**

Arguments annotated for values by 3 crowdworkers each ($\alpha = 0.49$)

Task 4. ValueEval: Identification of Human Values Behind Arguments

Submissions

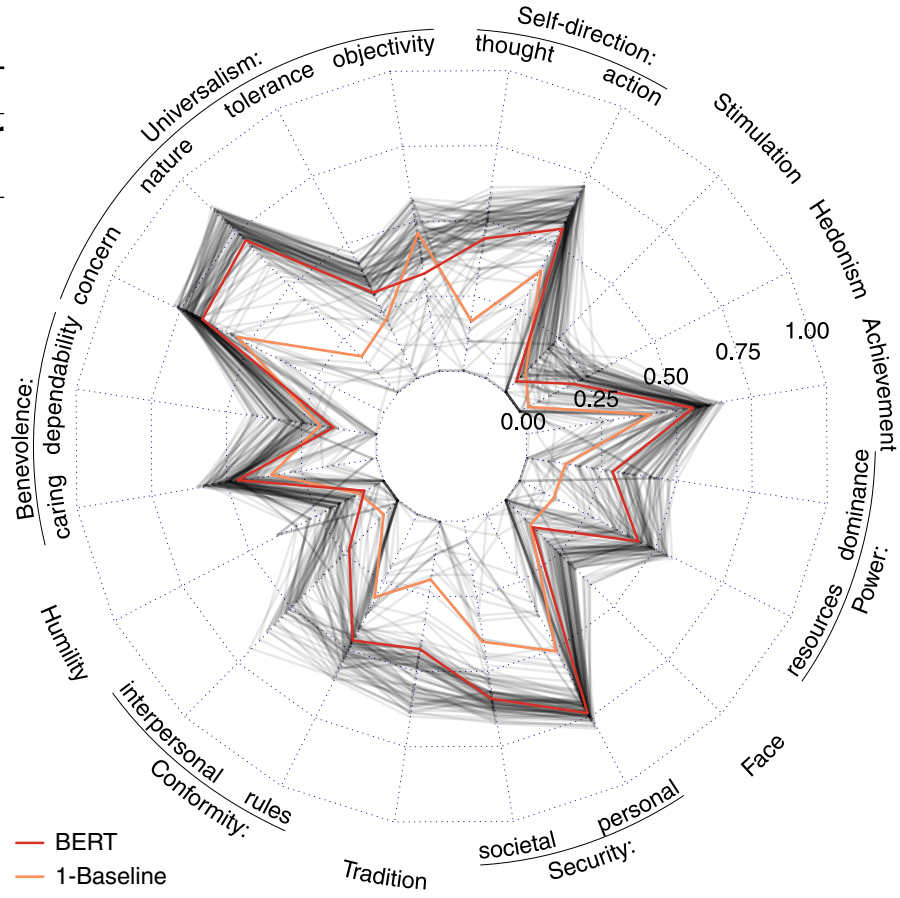
39 teams submitted runs (sent description: 37, published paper: 29, code: 23, Docker: 8)

- Mostly Transformer-based approaches (35 teams)
- Few natural language inference approaches (QA, entailment; 4 teams)
- Few uses of external data (value descriptions; 5 teams)
- Few perform pre-training (Full IBM data, ValueNet; 4 teams)
- Few perform contrastive learning (2 teams)

Task 4. ValueEval: Identification of Human Values Behind Arguments

Results

#	Approach	Best F_1-score		
		Main	Religious text	Newspaper editorials
	Best per value	0.59	0.48	0.47
1	Pretrained transformer ensemble Team Adam Smith	**0.56**	**0.40**	0.27
2	Fine-tuned transformer (DeBERTa) Team John Arthur	0.55	-	-
3	Transformer ensemble (weighted) Team PAI (Theodor Zwinger)	0.54	-	-
	...			
7	Question-answering (BART+T5) Team Hitachi (R. M. Hare)	0.51	0.34	**0.34**
	...			
30	BERT Baseline	0.42	0.28	0.24
	...			
38	1-Baseline	0.26	0.13	0.15
	...			

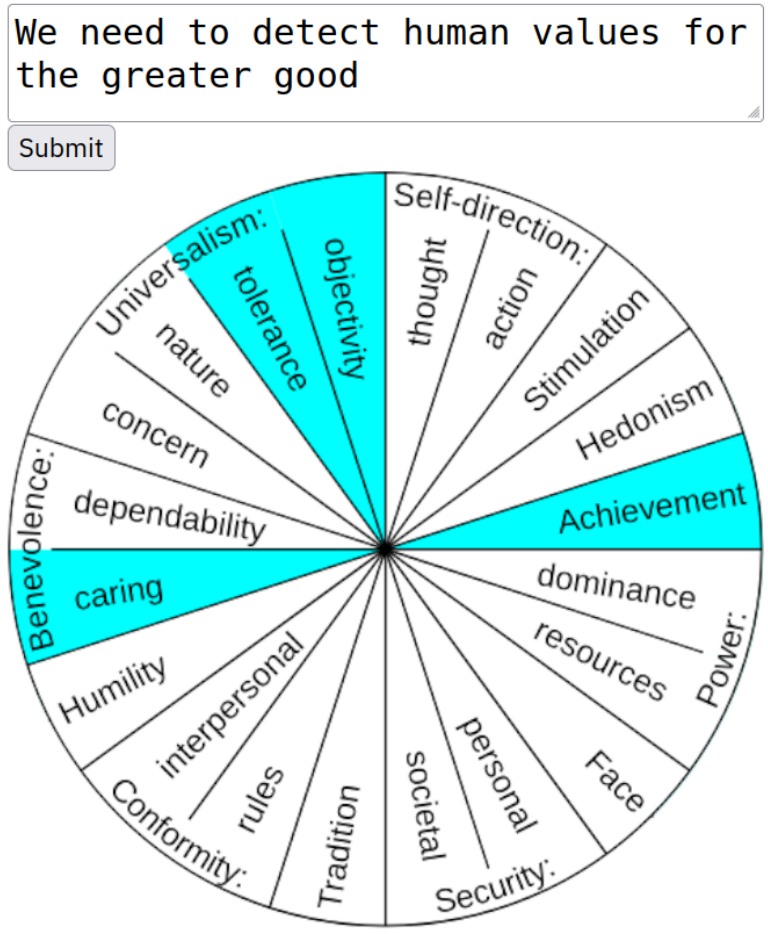


Task 4. ValueEval: Identification of Human Values Behind Arguments

Result Dissemination

Web page: https://valueeval.webis.de

- Dataset of 9324 arguments from 6 sources
- 29 participant papers
- 23 code repositories
- 8 Dockerized submissions
- Online demo of best approach: https://values.args.me



ValueEval'24

Multi-lingual Human Value Detection in Texts

Organized by Webis and the European Commission's Joint Research Centre.
More information coming soon: subscribe to our mailing list.

ValueEva

ValueEval continues for 2024!

https://valueeval.webis.de

Stay tuned: subscribe to our mailing list