

Touché 2024: Argumentation Systems

Touché is a series of scientific events and shared tasks on computational argumentation and causality

SHARED TASKS MAILING LIST

Touché at CLEF 2025
Overview
Retrieval-Augmented Debating
Ideology and Power Identification in Parliamentary Debates
Image Retrieval/Generation for Arguments
Advertisement in Retrieval-Augmented Generation
JOIN THE MAILING LIST

Touché at CLEF 2024
Overview
Human Value Detection
Ideology and Power Identification in Parliamentary Debates
Image Retrieval/Generation for Arguments
PROCEEDINGS

Touché at SemEval 2023
Human Value Detection
PROCEEDINGS

Touché continues in 2025 with four tasks:

1. Retrieval-Augmented Debating (new)
2. Ideology and Power Identification in Parliamentary Debates
3. Image Retrieval/Generation for Arguments
4. Advertisement in Retrieval-Augmented Generation (new)

Get notified: Join our new mailing list!



touche.webis.de

Task 1: Human Value Detection (ValueEval)

Johannes Kiesel Milad Alshomary Nailia Mirzakhmedova Nicolas Handke Nicolas Stefanovitch Bertrand De Longueville Mario Scharfbillig Henning Wachsmuth Benno Stein

Values play a pivotal role in shaping perspectives on policies and events. This task aims to facilitate large-scale analyses of values expressed in argumentative texts.

Table: F₁-score of the best submission per team (measured by overall F₁-score) on the test dataset, and whether the submission used the original multilingual dataset or the automatic translation to English (EN).

Detect for each sentence of a text to which of 19 human values (Schwartz' system) it refers to (multi-label).

Data: 2648 news articles and political manifestos (74 231 sentences) in nine languages: BE, DE, EL, EN, FR, HE, IT, NL, TR.

Team	Lang.	F ₁ -score																			
		Overall	Self-direction: thought	Self-direction: action	Stimulation	Hedonism	Achievement	Power: dominance	Power: resources	Face	Security: personal	Security: societal	Tradition	Conformity: rules	Conformity: interpersonal	Humility	Benevolence: caring	Benevolence: dependability	Universalism: concern	Universalism: nature	Universalism: tolerance
Hierocles of Alexandria	multil.	39	15	27	30	37	45	42	49	31	42	49	46	51	24	00	34	33	47	63	27
Arthur Schopenhauer	multil.	35	12	24	33	35	40	37	47	24	38	46	49	50	19	00	32	31	46	60	27
Philo of Alexandria	EN	28	08	22	27	31	35	31	34	17	33	40	47	42	09	00	21	28	40	57	21
SCaLAR NITK	EN	28	05	17	27	27	38	34	38	15	34	40	41	43	09	00	23	28	37	56	16
Edward Said	EN	28	05	17	11	15	25	31	34	16	32	41	45	44	06	05	10	23	41	57	27
Erich Fromm	EN	25	15	10	10	18	25	18	09	24	21	30	46	33	09	15	26	15	41	55	20
Lawrence Kohlberg	EN	25	08	11	19	23	31	22	31	11	28	37	34	42	09	00	21	23	34	54	18
Aristotle (BERT baseline)	EN	24	00	13	24	16	32	27	35	08	24	40	46	42	00	00	18	22	37	55	02
Adam Smith (best ValueEval/23)	EN	20	09	14	13	26	19	22	33	14	07	25	34	31	07	01	10	07	19	39	15
John Shelby Spong	EN	07	00	00	02	00	16	05	11	00	01	28	00	15	00	00	00	00	13	27	00
Alain Badiou	EN	07	00	00	02	00	16	05	11	00	01	28	00	15	00	00	00	00	13	27	00
Aristotle (random baseline)	EN	06	02	07	05	02	11	08	10	03	04	14	03	11	03	00	05	04	09	04	02

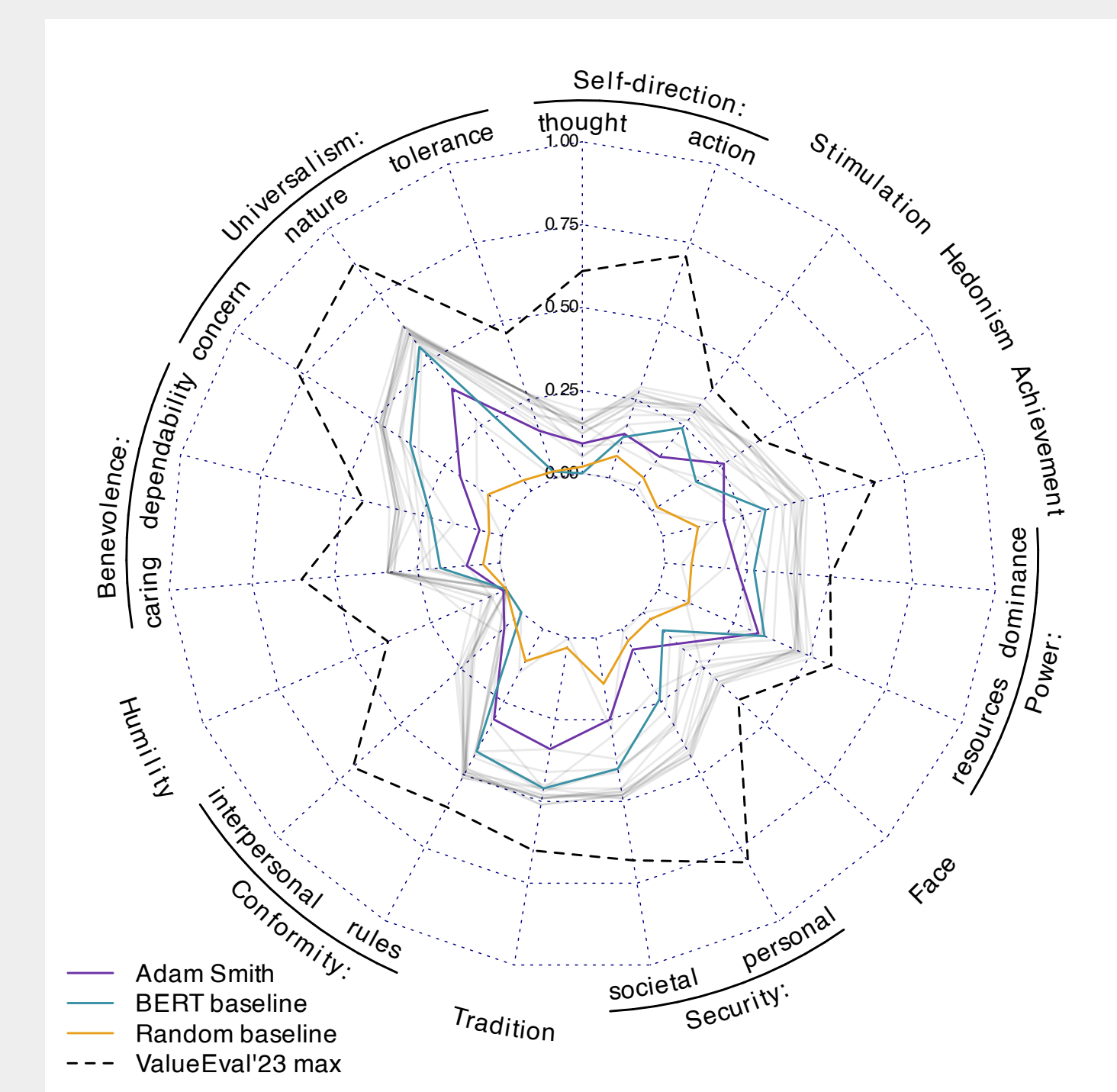


Figure: F₁-score for each value of each submission, with lines corresponding to one submission each. The farther a line is from the center, the better the value prediction.

Task 2: Multilingual Ideology and Power Identification in Parliamentary Debates

Çağrı Çöltekin Nikola Ljubešić Katja Meden Tomaž Erjavec Vaidas Morkevičius Matyáš Kopp

Parliamentary speeches are the channel between society and their elected representatives.

This (multi-lingual) task aims to investigate two important aspects of parliamentary debates.

Determine (1) a speaker's political orientation and (2) whether their party is governing or in opposition.

Data for (1): 205 200 speeches
Data for (2): 259 241 speeches

Table: F₁-scores of all submissions in orientation detection (left-wing vs. right-wing) for each parliament.

Team	Overall	F ₁ -score																											
		AT	BA	BE	BG	CZ	DK	EE	ES	ES-GA	FI	FR	GB	GR	HR	HU	IS	IT	LV	NL	NO	PL	PT	RS	SE	SI	TR	UA	
Policy Parsing Panthers	79	77	51	71	77	63	84	64	94	80	98	77	75	92	89	65	87	71	77	67	71	82	88	95	79	95	78	93	83
gerber	63	80	45	54	62	52	56	00	77	66	76	54	58	76	72	51	69	00	80	49	53	00	72	69	64	00	58	84	73
HALE Lab	61	56	44	59	60	52	56	52	76	69	84	52	48	74	71	43	67	57	60	49	53	61	62	67	55	77	49	83	60
Pixel Phantoms	59	58	49	56	56	47	56	54	72	64	75	59	58	72	71	55	68	57	54	60	54	59	54	51	61	47	78	56	
Ssnites	59	50	53	55	53	50	61	52	61	58	64	55	56	64	59	53	60	58	53	51	56	66	71	64	64	75	58	79	53
Trojan Horses	59	61	25	57	61	51	60	57	72	67	00	33	60	73	74	53	71	55	86	00	60	61	88	63	00	74	00	80	68
INSA Passau	59	60	53	54	61	47	57	53	63	61	66	34	58	69	59	56	68	56	56	54	58	58	69	55	61	66	51	80	62
JU NLP DID	57	53	42	42	55	51	60	57	69	57	70	00	50	71	63	43	60	55	61	47	56	59	51	67	48	73	46	77	57
Baseline	56	62	42	45	53	52	56	47	72	65	67	54	43	74	74	43	57	39	56	45	51	62	46	63	53	75	39	84	58

Table: F₁-scores of the best submissions per team in power identification (governing vs. opposition) for each parliament.

Team	Overall	F ₁ -score																											
		AT	BA	BE	BG	CZ	DK	EE	ES	ES-GA	ES-PA	FI	FR	GB	GR	HR	HU	IT	LV	NL	NO	PL	PT	RS	SE	SI	TR	UA	
Policy Parsing Panthers	83	88	56	74	81	78	87	88	91	98	90	80	82	83	95	75	97	78	75	74	90	85	84	81	94	65	81	94	65
HALE Lab	70	69	46	61	68	69	70	65	85	88	78	65	67	75	82	68	88	69	62	64	78	65	69	61	84	49	61	84	49
Trojan Horses	69	72	57	63	67	63	68	69	82	85	74	39	66	72	83	67	86	72	64	64	74	65	75	62	83	56	62	83	56
gerber	68	68	51	60	66	64	63	72	80	86	74	60	71	72	68	63	87	52	63	64	77	66	73	58	84	48	64	48	48
Vayam Solve Kurmaha	68	48	48	65	69	68	69	72	83	87	76	35	66	47	85	67	88	72	62	68	75	67	75	63	85	48	64	48	48
Pixel Phantoms	66	70	50	59	63	65	69	65	64	77	69	61	64	73	72	57	80	69	58	62	70	66	69	60	80	52	60	52	52
Baseline	64	66	45	61	68	64	56	65	78	83	71	56	66	71	63	60	86	43	51	62	76	62	65	53	83	46	66	46	46
JU NLP DID	63	68	47	55	58	57	67	60	78	55	72	00	59	00	77	65	83	71	47	63	70	63	54	56	78	43	67	43	43
INSA Passau	62	67	45	60	66	65	54	65	00	00	00	56	66	72	56	61	85	45	52	64	77	62	63	54	84	47	67	43	43
Ssnites	60	66	45	58	60	61	61	62	58	62	60	60	65	60	69	65	79	62	54	57	62	58	60	57	61	46	67	43	43

Task 3: Image Retrieval/Generation for Arguments (joint task with ImageCLEF)

Maximilian Heinrich Johannes Kiesel Martin Pothast Benno Stein

Images can strengthen an argumentative position through visual communication, but traditional image search is not well suited for finding images that support a particular point of view. This task aims to close this gap.

Given a claim, find (retrieve or generate) images that convey that claim

Data: 9145 images from the web and a Stable Diffusion API for 136 arguments.

Image metadata:

- URL and content of the page that contained the image
- Position of the image on that web page as XPath
- Google Cloud Vision: OCR (recognized text), recognized objects
- Automaticall generated description using LLaVA

Table: NDCG values of all submissions for the top 5, top 3 and top most relevant images.

Rank	Team	Approach	NDCG@5	NDCG@3	NDCG@1
1	HTW-DIL	Ada-Summary	0.428	0.409	0.404
2	HTW-DIL	Moondream-Text	0.363	0.355	0.356
3	HTW-DIL	Moondream-Default-Image-Text	0.293	0.302	0.317
4	Baseline	BM25	0.284	0.273	0.293
5	Baseline	SBERT	0.232	0.225	0.221
6	DS@GT	Generated-Image-CLIP	0.180	0.178	0.197
7	HTW-DIL	Moondream-Image-Text-EP3	0.150	0.163	0.183
8	HTW-DIL	Moondream-Image	0.146	0.155	0.178
9	DS@GT	Base-CLIP	0.123	0.111	0.106
10	HTW-DIL	Moondream-Image-Text-EP2	0.120	0.140	0.178