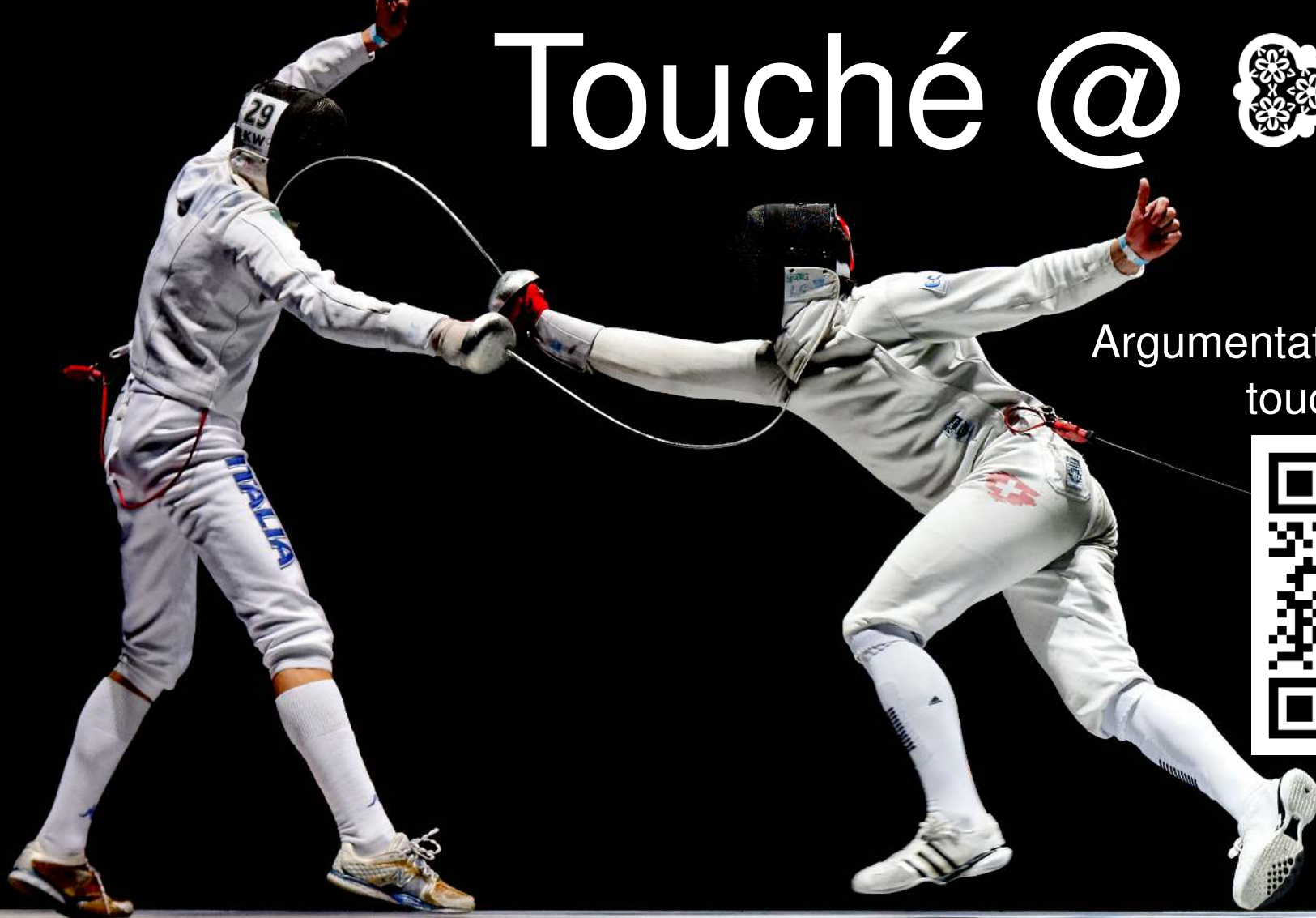


Touché @



5th Edition
Argumentation Systems
touche.webis.de



Monday, September 9, Room 2 (IMAG Amphitheatre)

14:00-15:30 Touché Session 1 (Keynote and Task Overviews)

14:00-15:00 More than Gender-Bias: Understanding the Sociological Imagination of Large Language Models
Keynote
Gilles Bastin

15:00-15:10 Overview of the Human Value Detection (ValueEval) Task [\[paper\]](#)
Johannes Kiesel

15:10-15:20 Overview of the Ideology and Power Identification in Parliamentary Debates Task [\[paper\]](#)
Çağrı Çöltekin (online)

15:20-15:30 Overview of the Image Retrieval/Generation for Arguments Task [\[paper\]](#)
Maximilian Heinrich

15:30-16:30 Poster Session + Coffee break

16:30-18:00 Touché Session 2 (Participant Presentations and Invited Talk)

16:30-16:40 Hierocles of Alexandria at Touché: Multi-task & Multi-head Custom Architecture with Transformer-based Models for Human Value Detection [\[paper\]](#)
Sotirios Legkas, Christina Christodoulou, Matthaios Zidianakis, Dimitrios Koutrintzes, Maria Dagioglou, Georgios Petasis

16:40-16:50 Eric Fromm at Touché: Prompts vs FineTuning for Human Value Detection [\[paper\]](#)
Ranjan Mishra, Meike Morren

16:50-16:55 Philo of Alexandria at Touché: A Cascade Model Approach to Human Value Detection [\[paper\]](#)
Victor Yeste, Mariona Coll-Ardanuy, Paolo Rosso

16:55-17:05 Arthur Schopenhauer at Touché 2024: Multi-Lingual Text Classification Using Ensembles of Large Language Models [\[paper\]](#)
Hamza Yunis (online)

17:00-17:15 SCaLAR NITK at Touché: Comparative Analysis of Machine Learning Models for Human Value Identification [\[paper\]](#)
Praveen K, Darshan R K, Chinta Tejdeep Reddy, Anand Kumar M (online)

17:15-17:25 ValueEval for Politics: Perspectives from the European Commission's JRC
Invited Talk
Mario Scharfbillig (online)

17:25-17:35 Trojan Horses at Touché: Logistic Regression for Classification of Political Debates [\[paper\]](#)
Deepak Chandar S, Diya Seshan, Avaneesh Koushik, P Mirunalini (online)

17:35-17:45 HALE Lab NITK at Touché 2024: A Hybrid Approach for Identifying Political Ideology and Power in Multilingual Parliamentary Speeches [\[paper\]](#)
Sevitha Simhadri, Mauli Mehulkumar Patel, Sowmya Kamath S (online)

17:45-17:55 Pixel Phantom at Touché: Ideology and Power Identification in Parliamentary Debates using Linear SVC [\[paper\]](#)
Janani Hariharakrishnan, Jithu Morrison S, P Mirunalini (online)

17:55-18:00 Closing
Johannes Kiesel



Touché: Argumentation Systems

14:00-15:30 Touché Session 1 (Keynote and Task Overviews)

14:00-15:00 More than Gender-Bias: Understanding the Sociological Imagination of Large Language Models
Keynote
Gilles Bastin

15:00-15:10 Overview of the Human Value Detection (ValueEval) Task [\[paper\]](#)
Johannes Kiesel

15:10-15:20 Overview of the Ideology and Power Identification in Parliamentary Debates Task [\[paper\]](#)
Çağrı Çöltekin (online)

15:20-15:30 Overview of the Image Retrieval/Generation for Arguments Task [\[paper\]](#)
Maximilian Heinrich



Human Value Detection (ValueEval)

Touché'24 Task 1



Johannes
Kiesel



Milad
Alshomary



Nailia
Mirzakhmedova



Nicolas
Handke



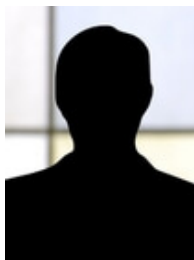
Bertrand
De Longueville



Theresa
Reitis-
Münstermann



Mario
Scharfbillig



Nicolas
Stefanovitch



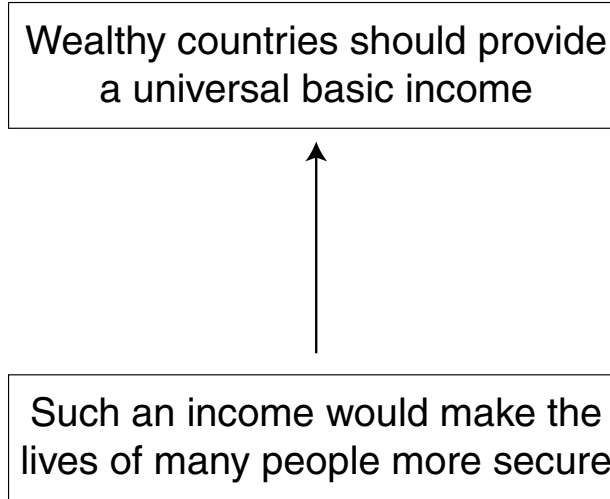
Henning
Wachsmuth



Benno
Stein

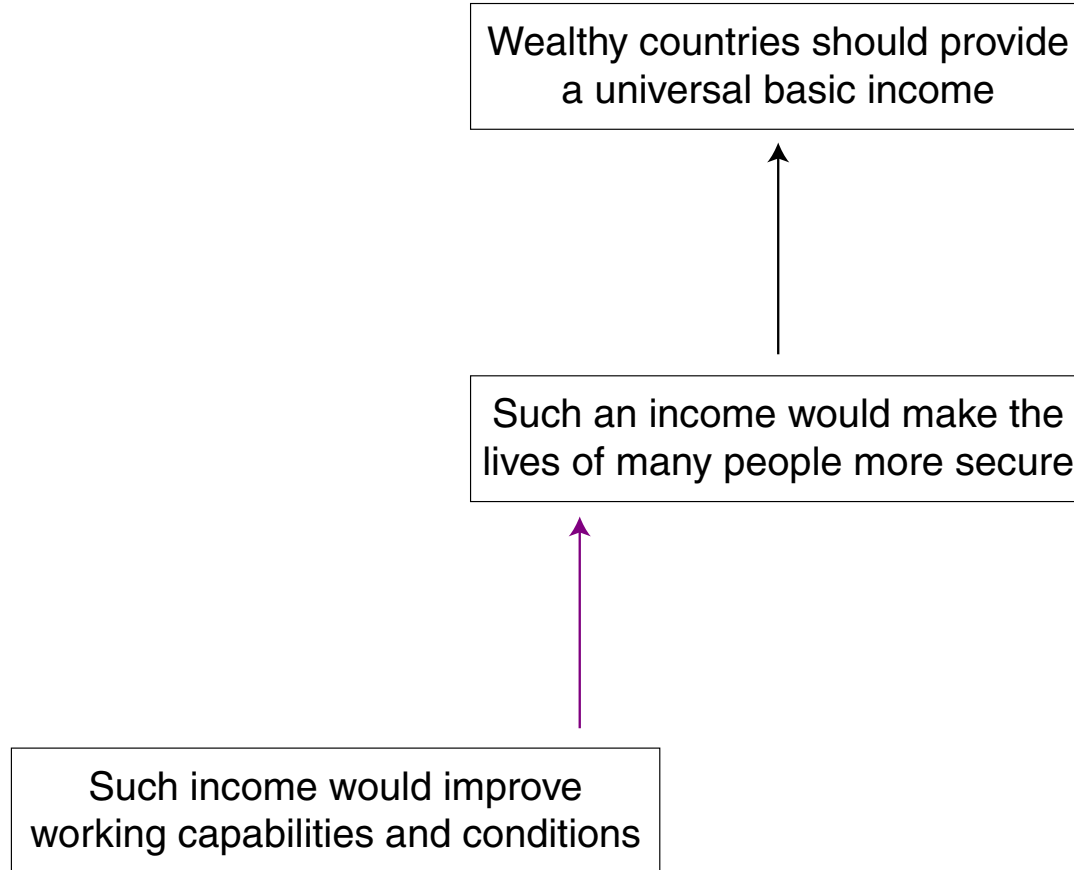
Human Value Detection (ValueEval)

Introduction



Human Value Detection (ValueEval)

Introduction



Human Value Detection (ValueEval)

Introduction

“Epistemological Why”

- Why is this true?

Wealthy countries should provide
a universal basic income



Such an income would make the
lives of many people more secure

epistemological
support



Such income would improve
working capabilities and conditions

Human Value Detection (ValueEval)

Introduction

“Epistemological Why”

- Why is this true?

Wealthy countries should provide
a universal basic income

“Ethical Why”

- Why is this good?

Such an income would make the
lives of many people more secure

epistemological
support

Such income would improve
working capabilities and conditions

ethical
support

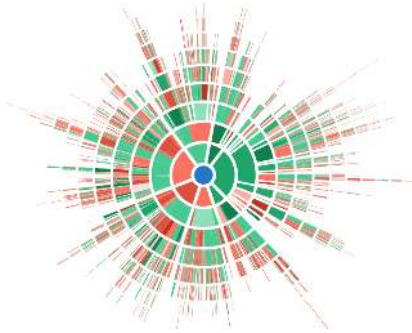
It is good when people have
personal security

Human Value Detection (ValueEval)

Introduction

“Epistemological Why”

- Why is this true?
- Ever branching reasons (infinitely?)



Source: kialo.com

epistemological
support

Such income would improve
working capabilities and conditions

Wealthy countries should provide
a universal basic income

Such an income would make the
lives of many people more secure

ethical
support

It is good when people have
personal security

“Ethical Why”

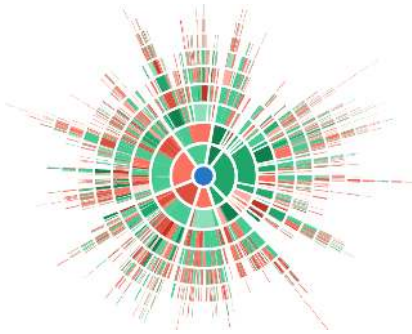
- Why is this good?

Human Value Detection (ValueEval)

Introduction

“Epistemological Why”

- Why is this true?
- Ever branching reasons (infinitely?)



Source: kialo.com

epistemological
support

Such income would improve
working capabilities and conditions

Wealthy countries should provide
a universal basic income

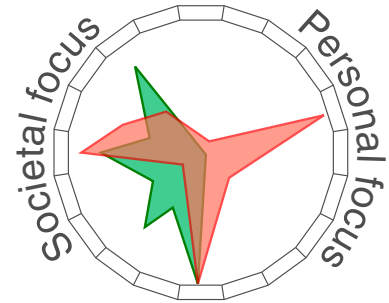
Such an income would make the
lives of many people more secure

ethical
support

It is good when people have
personal security

“Ethical Why”

- Why is this good?
- Leading to *values* (finite set)

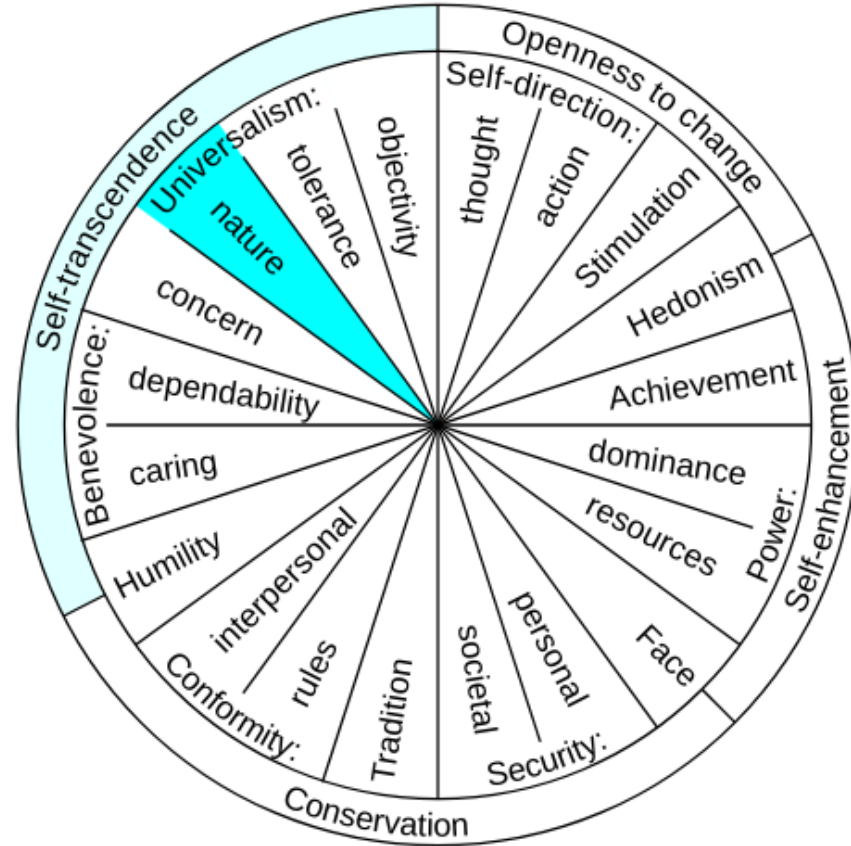


Human Value Detection (ValueEval)

ValueEval'23 Demo

We need to reduce our CO2 emissions to save the environment.

Submit



Human Value Detection (ValueEval)

Task Description

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.

Scenario: Analyzing large quantities of text for social science studies

Task: Given a text, for each sentence, detect

Subtask 1: which human values the sentence refers to (19-label task); and

Subtask 2: whether such reference (partially) attains or (partially) constrains the value

Background: Schwartz' taxonomy of personal human values has been replicated over decades in over 200 samples in 80 countries

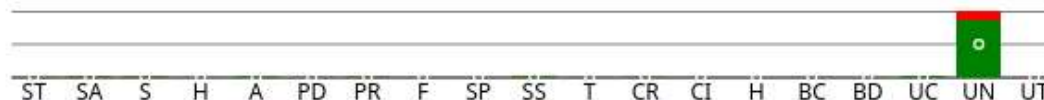
Human Value Detection (ValueEval)

ValueEval'24 Demo

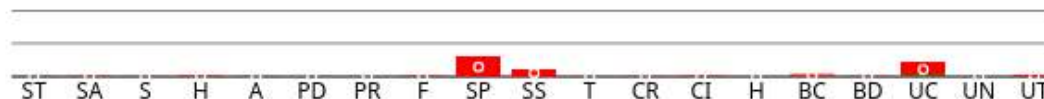
Approach:

We need to reduce our CO2 emissions to save the environment.
If we do not, many people will suffer.

We need to reduce our CO2 emissions to save the environment.



If we do not, many people will suffer.



Human Value Detection (ValueEval)

Dataset

- ❑ 74 231 sentences in 2 648 texts in 9 languages:
Bulgarian, German, Greek, English, French, Hebrew, Italian, Dutch, Turkish
- ❑ News articles and political manifestos (party agendas)
- ❑ Extensive annotation and curation by experts in collaboration with the Joint Research Centre of the European Commission (JRC; short talk from representative later!)
- ❑ Challenge: value distribution is highly skewed
(*Security: societal* behind 8.6% of sentences,
Humility behind 0.2% of sentences)
- ❑ Challenge: difference between annotations of
different language teams



[https://zenodo.org/doi/
10.5281/zenodo.10396293](https://zenodo.org/doi/10.5281/zenodo.10396293)

Human Value Detection (ValueEval)

Dataset



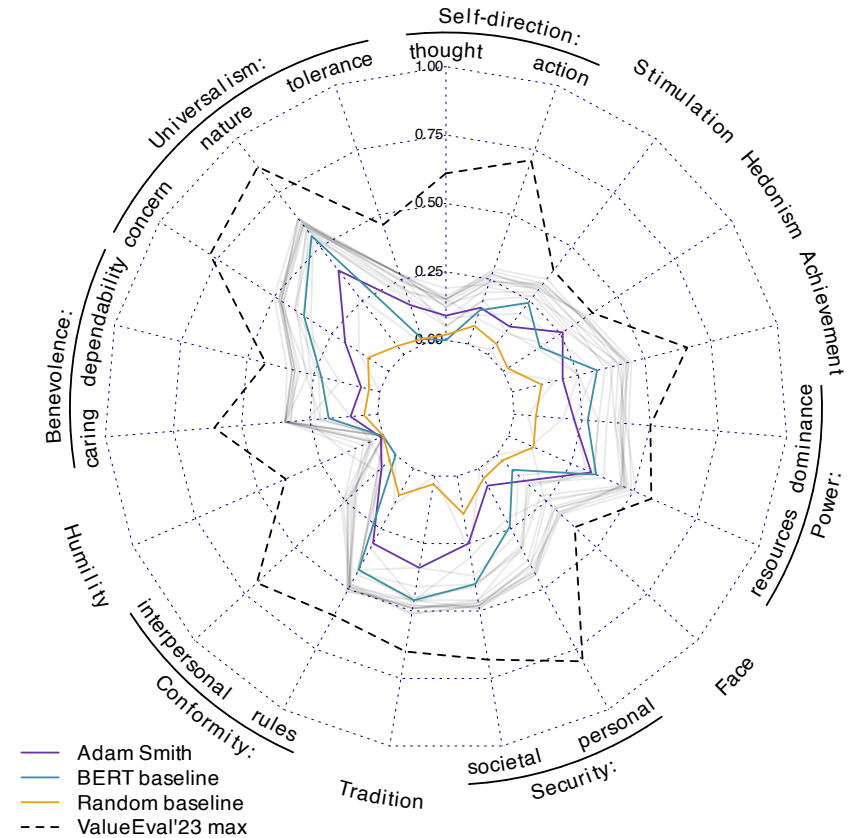
<https://zenodo.org/doi/10.5281/zenodo.10396293>

Human Value Detection (ValueEval)

Results

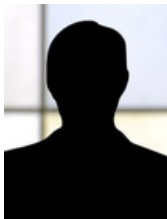
- ❑ Teams largely ignored the attainment subtask
- ❑ Task much harder than last year
- ❑ Multilingual models perform best (top-2)
- ❑ Rarest value (*Humility*) detected best by zero-shot GPT-4o

#	Approach	F ₁ -score
1	Language-specific transformer on sequences (XLM-RoBERTa) Team Hierocles of Alexandria	0.39
2	Multi-lingual transformer ensemble (XLM-RoBERTa) Team Arthur Schopenhauer	0.35
3	Fine-tuned transformer (DeBERTa) Team Philo of Alexandria	0.28
4	Fine-tuned transformer (RoBERTa) Team SCaLAR NITK	0.28
6	GPT-4o zero-shot classification Team Erich Fromm	0.25
8	BERT Baseline	0.24
11	1-Baseline	0.06
12	Random baseline	0.06



Multilingual Ideology and Power Identification in Parliamentary Debates

Touché'24 Task 2



Çağrı
Çöltekin



Katja
Meden



Nikola
Ljubešić



Tomaž
Erjavec



Vaidas
Morkevičius



Matyáš
Kopp

Multilingual Ideology and Power Identification in Parliamentary Debates

Introduction

- Parliamentary debates result in decisions with high societal impact
- Political/parliamentary language is difficult to analyze
 - highly conventionalized
 - strategies like evasion, circumlocution or the use of metaphors are common
- This task is about identifying two fundamental aspects in political discourse
 - *Political orientation*: computational studies becoming popular, including recent shared tasks in IberLEF and EvalITA
 - *Power role*: central in discourse analysis, virtually no computational studies

Multilingual Ideology and Power Identification in Parliamentary Debates

Task Description

Scenario: Identify the political orientation and the power role of the speaker from their speeches in parliamentary debates.

Task: Given a transcribed speech delivered in a parliament

Subtask 1: identify political orientation of the speaker (left–right)

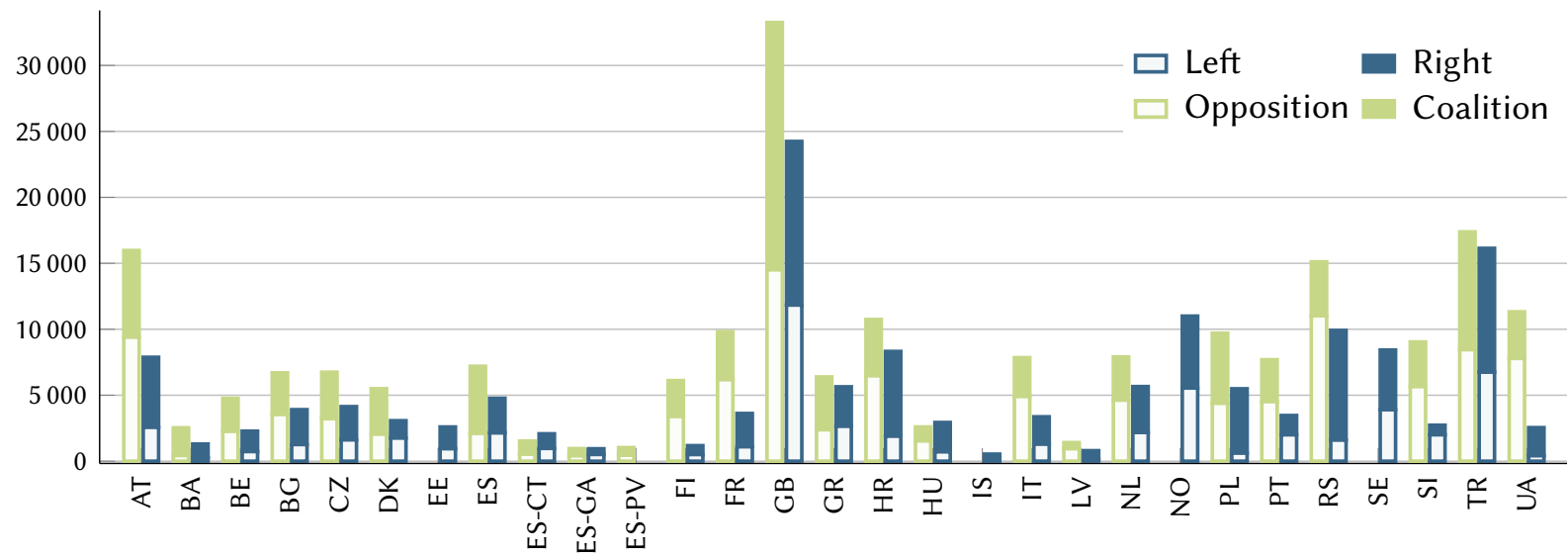
Subtask 2: identify power role of the speaker (coalition–opposition)

Data:

- A subset of the ParlaMint version 4.0
- 29 national and regional parliaments (some available only for one of the tasks)
- 30 languages (also automatic translation to English)
- Date range varies by parliament, but includes at least from 2015 to 2022
- Typically long texts (approx. 600 words on average)

Multilingual Ideology and Power Identification in Parliamentary Debates

A closer look at the data



Task	Speeches		Words	
	Train	Test	Train	Test
Orientation	148943	56257	90M	35M
Power	209241	50000	135M	33M

Multilingual Ideology and Power Identification in Parliamentary Debates

Results: leaderboard

Orientation	
Team	F ₁ -score
Policy Parsing Panthers	0.79
gerber	0.63
HALE Lab	0.61
Pixel Phantoms	0.59
Ssnites	0.59
Trojan Horses	0.59
INSA Passau	0.59
JU_NLP_DID	0.57
Baseline	0.56

Power	
Team	F ₁ -score
Policy Parsing Panthers	0.83
HALE Lab	0.70
Trojan Horses	0.69
gerber	0.68
Vayam Solve Kurmaha	0.68
Pixel Phantoms	0.66
Baseline	0.64
JU_NLP_DID	0.63
INSA Passau	0.62
Ssnites	0.60

Multilingual Ideology and Power Identification in Parliamentary Debates

Results: observations

- ❑ Most teams participated in both tasks, on (almost) all parliaments
- ❑ Participations focusing on a single country/parliament were rare
- ❑ (Fine-tuning) pre-trained models often yielded the best results
- ❑ Many teams also used ‘traditional’ ML methods (SVMs, Logistic Regression, kNN, random forests), and deep learning methods without pre-training (CNNs)
- ❑ The use of both original transcript and English translations was common for most teams
- ❑ Interesting approaches include
 - Ensemble methods
 - Data augmentation (through back-translation, synonym replacement)
 - Adding auxiliary tasks during training (e.g., sentiment scores)
 - The use of domain-specific pre-trained models

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Touché'24 Task 3



Maximilian
Heinrich



Johannes
Kiesel



Martin
Potthast



Benno
Stein

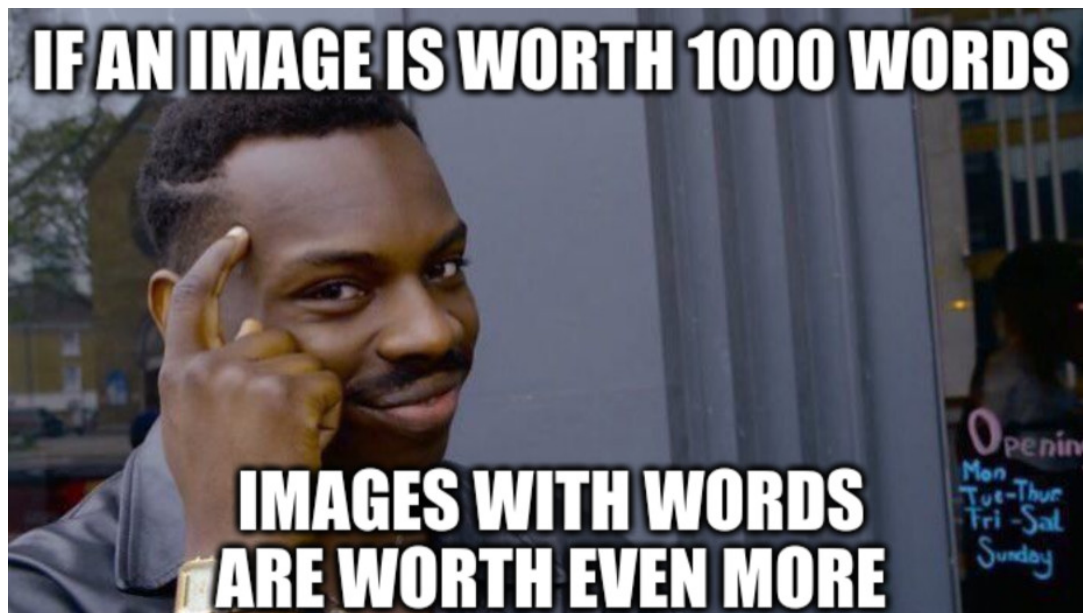


Image Retrieval/Generation for Arguments [Joint Task with Touché]

Task Description

Scenario: Enhance impact of arguments

Task: Given an argument, find images that help to convey the argument's premise.

- Participants can retrieve images from our collection or generate them using a text-to-image model
- Participants can submit an image description (rationale) to explain why the image helps to convey the premise

- Data:
- 106 arguments for 17 topics
 - 9145 crawled images, their webpage, position on that webpage, text extracted from that webpage, webarchive to allow to render the webpage, query used to crawl the image and rank in search engine result page, recognized text in image (OCR), detected objects in image, automated descriptions of image (LLaVA)
 - Access to a Stable Diffusion API

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Example Submission

Topic: Should boxing be banned?

Claim: Boxing poses both physical and psychological threats to participants, hence it should be banned.

Premise: The idea of winning through intentional infliction of pain and harm to another person can nurture a violent and destructive mentality.

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Example Submission

Topic: Should boxing be banned?

Claim: Boxing poses both physical and psychological threats to participants, hence it should be banned.

Premise: The idea of winning through intentional infliction of pain and harm to another person can nurture a violent and destructive mentality.



Image Retrieval/Generation for Arguments [Joint Task with Touché]

Example Submission

Topic: Should boxing be banned?

Claim: Boxing poses both physical and psychological threats to participants, hence it should be banned.

Premise: The idea of winning through intentional infliction of pain and harm to another person can nurture a violent and destructive mentality.



Rationale: The infliction of pain is a central component of boxing.

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Available Data




I<first-2-chars-of-image-url-hash>/	
I<full-24-chars-image-url-hash>/	Directory name == image ID
image.webp	M Image in WebP format
image-vision.json	M Contains an AnnotateImageResponse from the Google Cloud Vision API. Included annotations: face detection, label annotations, localized object annotations, text annotations, full text annotation, safe search annotation, web detection. Documentation: https://web.archive.org/web/20230130192539/https://cloud.google.com/vision/docs/reference/rest/v1/AnnotateImageResponse
image-url.txt	M URL of the image
image-text.txt	M The text recognized by Google Cloud Vision, extracted from image-vision.json
image-caption.txt	M Description of the image, automatically generated with LLaVa
image-phash.txt	M 64bit pHash of the (WebP) image as string: https://www.phash.org/
pages/	
P<full-24-chars-page-url-hash>/	Directory name == page ID
page-url.txt	M URL of the web page (containing the image)
rankings.jsonl	M Each line contains a JSON object describing a query to Google that retrieved the image/page as follows: { "query": "<query text>", "rank": "<image/page rank in result list starting with 1> }
snapshot/	
dom.html	M Snapshot of the HTML DOM
image-xpath.txt	M Each line contains the XPath of a node in the dom.html that references the image (img, picture, or meta in this order)
nodes.jsonl	N Each line contains a JSON object describing a node of dom.html as follows: { "xpath": "<XPath of the node in the dom.html>", "visible": <Boolean whether the node is visible as per https://stackoverflow.com/a/33456469 >, "classes": ["<entry of the class attribute>", ...], "position": [<left border of node pixel position in screenshot.webp starting left with 0>, <top border of node pixel position in screenshot.webp starting top with 0>, <right border of node pixel position in screenshot.webp starting left with 0>, <bottom border of node pixel position in screenshot.webp starting top with 0>], "text": "text content of the node", "css": { "<css-attribute>": "<css-attribute value>", ... } }
screenshot.png	S Screenshot of the page in PNG format
text.txt	M Text content of the dom.html (taken from the first node of the nodes.jsonl)
web-archive.warc.gz	A Web archive file containing all resources requested when taking the snapshot

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Evaluation

- For each of 106 arguments (topic, premise, claim), each submitted image and rationale was judged by one expert (5061 judgments in total)

- Judgment:

0: Image does not convey the premise	80%	
1: Image partially conveys the premise	12%	
2: Image fully conveys the premise	8%	

- Observation: more than half of the images scored 0 are still on-topic
- Systems were evaluated using NDCG@5, NDCG@3, NDCG@1; respective rankings are nearly identical

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Results

Team	Approach	NDCG@5
HTW-DIL	Ada-Summary	0.428
HTW-DIL	Moondream-Text	0.363
HTW-DIL	Moondream-Image-Text-Default	0.293
Baseline	BM25	0.284
Baseline	SBERT	0.232
DS@GT	Generated-Image-CLIP	0.180
HTW-DIL	Moondream-Image-Text-3epochs	0.150
HTW-DIL	Moondream-Image	0.146
DS@GT	Base-CLIP	0.123
HTW-DIL	Moondream-Image-Text-2epochs	0.120

Baselines:

- ❑ **BM25:** Indexed LLaVA captions; Ranked using premise as query
- ❑ **SBERT:** Ranked according to embedding similarity between premise and LLaVA captions

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Results

Team	Approach	NDCG@5
HTW-DIL	Ada-Summary	0.428
HTW-DIL	Moondream-Text	0.363
HTW-DIL	Moondream-Image-Text-Default	0.293
Baseline	BM25	0.284
Baseline	SBERT	0.232
DS@GT	Generated-Image-CLIP	0.180
HTW-DIL	Moondream-Image-Text-3epochs	0.150
HTW-DIL	Moondream-Image	0.146
DS@GT	Base-CLIP	0.123
HTW-DIL	Moondream-Image-Text-2epochs	0.120

DS@GT:

- ❑ **Base-CLIP:** Ranked according to CLIP embedding similarity between arguments and images
- ❑ **Generate-Image-CLIP:** Re-ranked top-40 from Base-CLIP by CLIP average embedding similarity to generated images; images are generated using StableDiffusion from attacking/supporting claims themselves generated using TinyLlama

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Results

Team	Approach	NDCG@5
HTW-DIL	Ada-Summary	0.428
HTW-DIL	Moondream-Text	0.363
HTW-DIL	Moondream-Image-Text-Default	0.293
Baseline	BM25	0.284
Baseline	SBERT	0.232
DS@GT	Generated-Image-CLIP	0.180
HTW-DIL	Moondream-Image-Text-3epochs	0.150
HTW-DIL	Moondream-Image	0.146
DS@GT	Base-CLIP	0.123
HTW-DIL	Moondream-Image-Text-2epochs	0.120

HTW-DIL:

- **Moondream:** Ranked according to Moondream embedding similarity between arguments and embeddings generated from:
 - (1) the image (Moondream-Image),
 - (2) the Bart-summarized webpage and crawl query (Moondream-Text),
 - (3) both (Moondream-Image-Text-Default), or
 - (4) both and after fine-tuning for 2 or 3 epochs to maximize similarity between the images in the dataset and arguments generated using GPT-4 from image and metadata

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Results

Team	Approach	NDCG@5
HTW-DIL	Ada-Summary	0.428
HTW-DIL	Moondream-Text	0.363
HTW-DIL	Moondream-Image-Text-Default	0.293
Baseline	BM25	0.284
Baseline	SBERT	0.232
DS@GT	Generated-Image-CLIP	0.180
HTW-DIL	Moondream-Image-Text-3epochs	0.150
HTW-DIL	Moondream-Image	0.146
DS@GT	Base-CLIP	0.123
HTW-DIL	Moondream-Image-Text-2epochs	0.120

HTW-DIL:

- **Ada-Summary:** Ranked according to ADA embedding similarity between arguments and textual data for each image (Bart-summarized webpage and crawl query)

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Results

Team	Approach	NDCG@5
HTW-DIL	Ada-Summary	0.428
HTW-DIL	Moondream-Text	0.363
HTW-DIL	Moondream-Image-Text-Default	0.293
Baseline	BM25	0.284
Baseline	SBERT	0.232
DS@GT	Generated-Image-CLIP	0.180
HTW-DIL	Moondream-Image-Text-3epochs	0.150
HTW-DIL	Moondream-Image	0.146
DS@GT	Base-CLIP	0.123
HTW-DIL	Moondream-Image-Text-2epochs	0.120

Observations:

- ❑ Top-2 approaches do not use the image itself (though images are from focused crawled via Google image search → image was indirectly used in crawling)
- ❑ No team submitted generated images, though image generation was used for re-ranking
- ❑ No team submitted rationales

Image Retrieval/Generation for Arguments [Joint Task with Touché]

Lessons Learned

Improvements for 2025

- ❑ More focused image crawl for more relevant images in the collection
- ❑ Searching for arguments for a claim instead of topic + claim + premise to avoid confusing both participants and models
- ❑ Providing more baseline implementations to advertise all the different data we collected for each image (OCR, recognized objects, web page, automated captions)

Touché: Argumentation Systems

16:30-18:00 Touché Session 2 (Participant Presentations and Invited Talk)

16:30-16:40 Hierocles of Alexandria at Touché: Multi-task & Multi-head Custom Architecture with Transformer-based Models for Human Value Detection [\[paper\]](#)
Sotirios Legkas, Christina Christodoulou, Matthaïos Zidianakis, Dimitrios Koutrintzes, Maria Dagioglou, Georgios Petasis

16:40-16:50 Eric Fromm at Touché: Prompts vs FineTuning for Human Value Detection [\[paper\]](#)
Ranjan Mishra, Meike Morren

16:50-16:55 Philo of Alexandria at Touché: A Cascade Model Approach to Human Value Detection [\[paper\]](#)
Víctor Yeste, Mariona Coll-Ardanuy, Paolo Rosso

16:55-17:05 Arthur Schopenhauer at Touché 2024: Multi-Lingual Text Classification Using Ensembles of Large Language Models [\[paper\]](#)
Hamza Yunis (online)

17:00-17:15 SCaLAR NITK at Touché: Comparative Analysis of Machine Learning Models for Human Value Identification [\[paper\]](#)
Praveen K, Darshan R K, Chinta Tejdeep Reddy, Anand Kumar M (online)

17:15-17:25 ValueEval for Politics: Perspectives from the European Commission's JRC
Invited Talk
Mario Scharfbillig (online)

17:25-17:35 Trojan Horses at Touché: Logistic Regression for Classification of Political Debates [\[paper\]](#)
Deepak Chandar S, Diya Seshan, Avaneesh Koushik, P Mirunalini (online)

17:35-17:45 HALE Lab NITK at Touché 2024: A Hybrid Approach for Identifying Political Ideology and Power in Multilingual Parliamentary Speeches [\[paper\]](#)
Sevitha Simhadri, Mauli Mehulkumar Patel, Sowmya Kamath S (online)

17:45-17:55 Pixel Phantom at Touché: Ideology and Power Identification in Parliamentary Debates using Linear SVC [\[paper\]](#)
Janani Hariharakrishnan, Jithu Morrison S, P Mirunalini (online)

17:55-18:00 Closing
Johannes Kiesel

Touché: Argumentation Systems

Outlook for Touché 2025

Main Organizing Team

- ❑ Johannes Kiesel (Bauhaus-Universität Weimar)
- ❑ Martin Potthast (University of Kassel)
- ❑ Benno Stein (Bauhaus-Universität Weimar)

Core Team for TIRA Support

- ❑ Maik Fröbe (Friedrich-Schiller-Universität Jena)
- ❑ Tim Hagen (University of Kassel)



Touché: Argumentation Systems

Outlook for Touché 2025

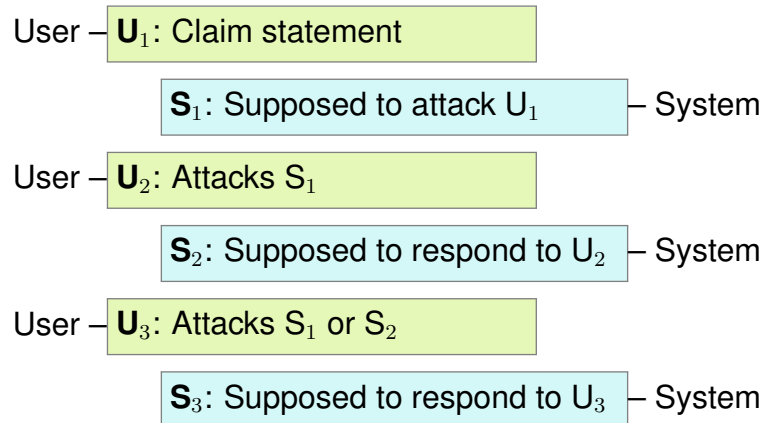


Task 1: Retrieval-Augmented Debating (RAD)

Scenario: Assisting people in forming an opinion on controversial topics and training argumentation skills

Tasks: (1) Retrieve and respond with counterarguments and evidence in simulated debates; (2) Automate the evaluation of such systems

Data: Collection of over 300 000 claims and 100 judged baseline debates



User: (simulated by organizers)
states a claim and attacks the system's responses.

System: (submitted by participants)
counterattacks arguments of user or defends own arguments.

Touché: Argumentation Systems

Outlook for Touché 2025



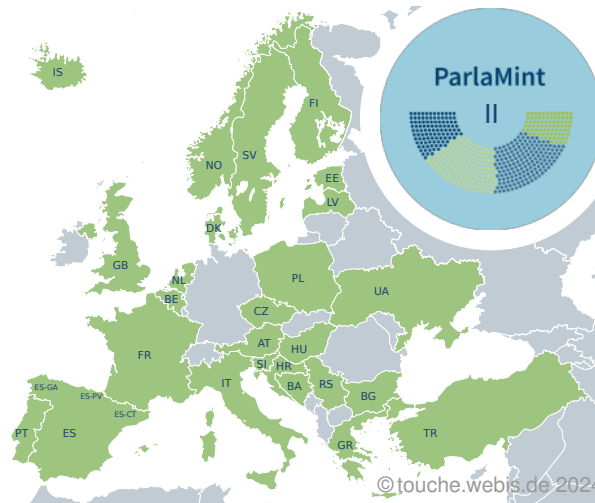
Task 2: Ideology and Power Identification in Parliamentary Debates

Scenario: To better understand how political ideology and the position of the speaker affects parliamentary debates

Tasks: (1) Determine a speaker's political orientation and (2) whether their party is governing or in opposition (multi-lingual)

Data: Speech samples from multiple national/regional parliaments from the ParlaMint project, and their automatic translations to English

- ❑ This task is a re-run of the previous year's task
- ❑ Main differences:
 - Multi-class ideology classification
 - Identifying members of the government



Touché: Argumentation Systems

Outlook for Touché 2025



Task 3: Image Retrieval/Generation for Arguments (ArgImages)

Scenario: Reinforce the impact of arguments with images.

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

Data: Hand-picked claims (similar to topics in TREC), collection of 20,000 images (meta-information: OCR, recognized objects, LLM image descriptions, . . .), text-to-image generation API

Example:

Image retrieved for claim:

“Gambling can be a joyful activity”

Assessment: *good*



Touché: Argumentation Systems

Outlook for Touché 2025



Task 4: Advertisement in Retrieval-Augmented Generation

Scenario: Commercial RAG systems / LLMs may integrate advertisements in their generated answers and users may want to block them

Tasks: (1) Generate relevant responses to queries that advertise a specified brand or product; (2) Detect the advertisements of others

Data: The [Webis Generated Native Ads 2024](#) dataset containing 11k generated responses and 6k inserted advertisements

Example:

Query

spider man remastered

Original Response

Are you looking for information about Marvel's Spider-Man Remastered? It is an action-packed game ...

Product with Qualities to Advertise

PlayStation 5
- 4K graphics
- innovative

Response with Advertisement

Are you looking for information about Marvel's Spider-Man Remastered? **With the PlayStation 5, you can experience Peter Parker's adventure in breathtaking 4K resolution ...**

Touché: Argumentation Systems

Stay Up-to-Date: Register to our New Mailing List!



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](#)

Touché at CLEF 2024

[Overview](#)

[Human Value Detection](#)

[Ideology and Power Identification in Parliamentary Debates](#)

Touché at SemEval 2023

[Human Value Detection](#)