# On Stance Detection in Image Retrieval for Argumentation

**Speaker: Miriam Louise Carnot**

Authors:
M. L. Carnot, L. Heinemann,
J. Braker, T. Schreieder,
J. Kiesel, M. Fröbe,
M. Potthast, B. Stein

SIGIR TAIPEI | TAIWAN | 2023
Student Travel Grant

ScaDS.AI DRESDEN LEIPZIG
MARTIN-LUTHER-UNIVERSITÄT HALLE-WITTENBERG
UNIVERSITÄT LEIPZIG
Bauhaus-Universität Weimar

# What is Image Retrieval for Argumentation?

**Query: "Should the penny stay in circulation?"**

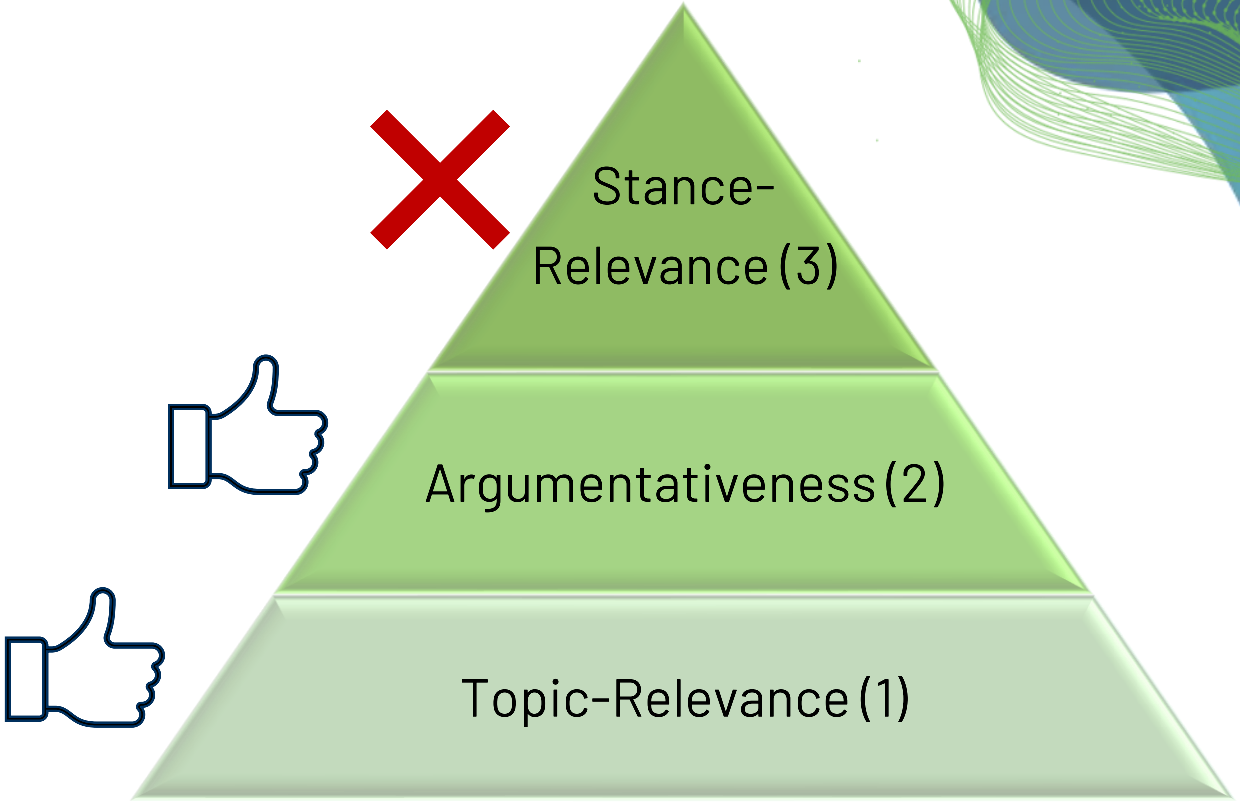
PRO


CON

- → The user enters a controversial topic
- → Search for argumentative images
- → Division of images into pro and con

# Work Done So Far

- →First shared task at the Touché lab of the CLEF conference in 2022
- →Three-stage evaluation
- →Very good results for (1) and (2)
- →Unsatisfactory results for (3)



# Re-Using the Touché’22 Dataset

Freely accessible

**50** controversial topics as queries

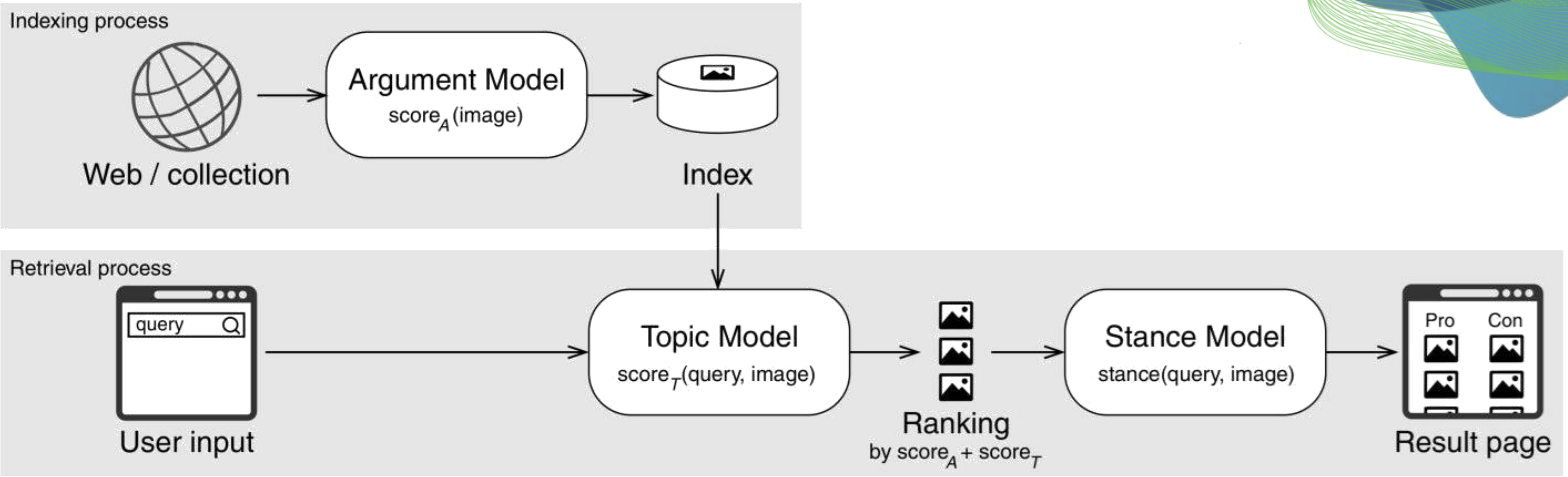
**23,841** images

Included for each image:

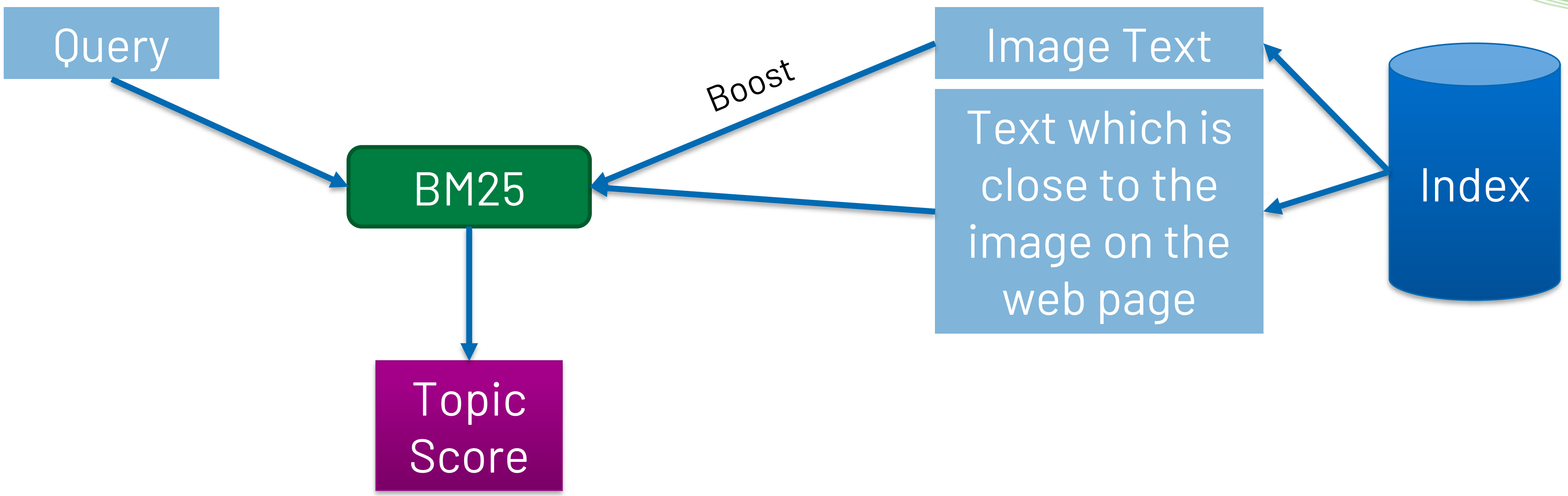
- Image pixel values
- Web page screenshot
- Web page text and HTML source code
- Etc.

Relevance ratings for **6,607** images

# Our unified image retrieval system for arguments

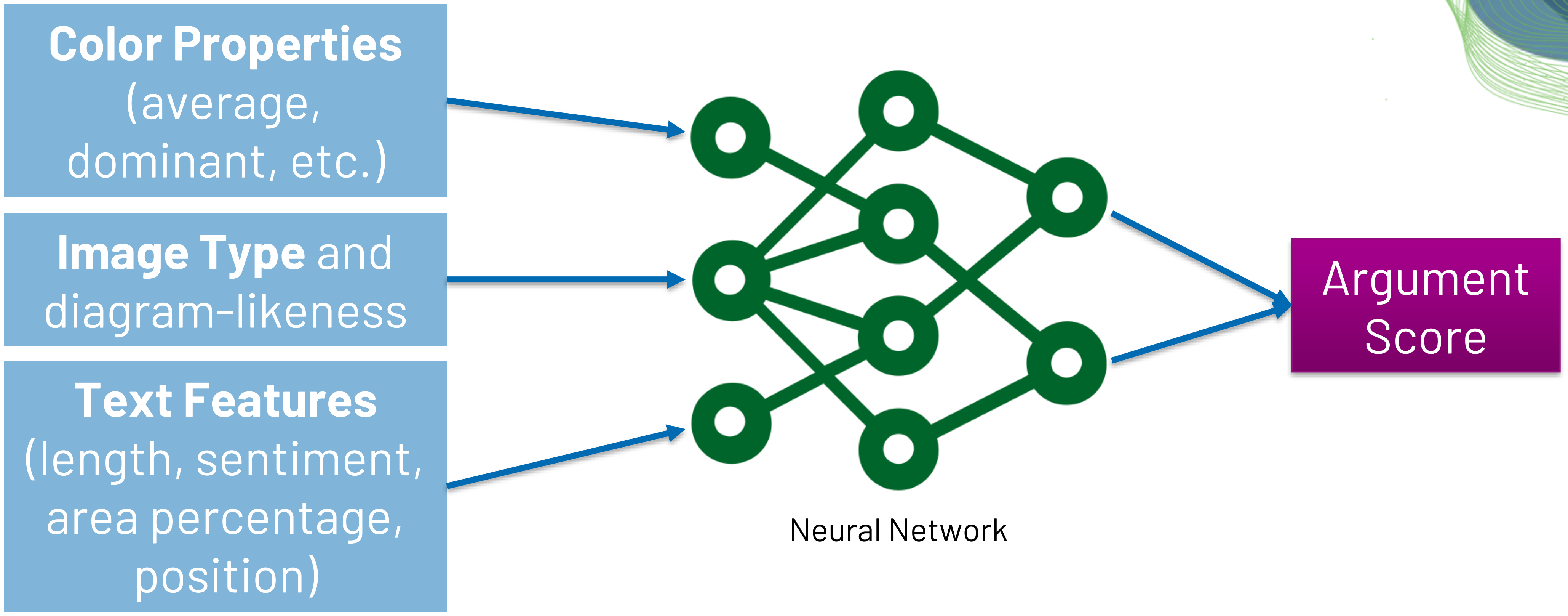


# Topic Model




ScaDS.AI DRESDEN LEIPZIG

# Argument Model



# Stance Models (1/2)

| | |
|---|---|
| **Oracle** | upper limit using the ground-truth stance labels |
| **Both-sides baseline** | each image in pro *and* con |
| **Random baseline** | each image in pro *or* con with equal probability |
| **Crawl query stance** | labels each image based on which result list it was originally found while crawling (query was extended with pro/anti) |
| **CLIP query stance** | uses CLIP to compute the image's similarity to the query extended with "good" for pro / "anti" for con |
| **BERT title sentiment** | uses a BERT-model to classify the sentiment of the web page's title |
| **AFINN text sentiment** | sums up the AFINN sentiment score for each word of the web page's text |

>0 → pro
<0 → con

# Stance Models (2/2)

| | |
|---|---|
| **Aramis Formula** | uses a heuristic formula that is based on 13 features (developed by Team Aramis) |
| **Aramis Neural** | uses the same features as *Aramis Formula* as input for a NN, classifies images into pro/con/neutral |
| **Neural text+image 3class** | combines a BERT model with a ResNet50V2 extended by some dropout layers; uses the image, the query, and the OCR text as input; 3 output neurons |
| **Neural text+image 2x2class** | same as *Neural text+image 3class* but with a single output neuron, trained twice (for pro and for con independantly) |
| **Neural text 3class** | same as *Neural text+image 3class* but with the title of the web page instead of the image |
| **Neural text+page 3class** | same as *Neural text 3class* but additionally uses HTML text in the window around the image as input |

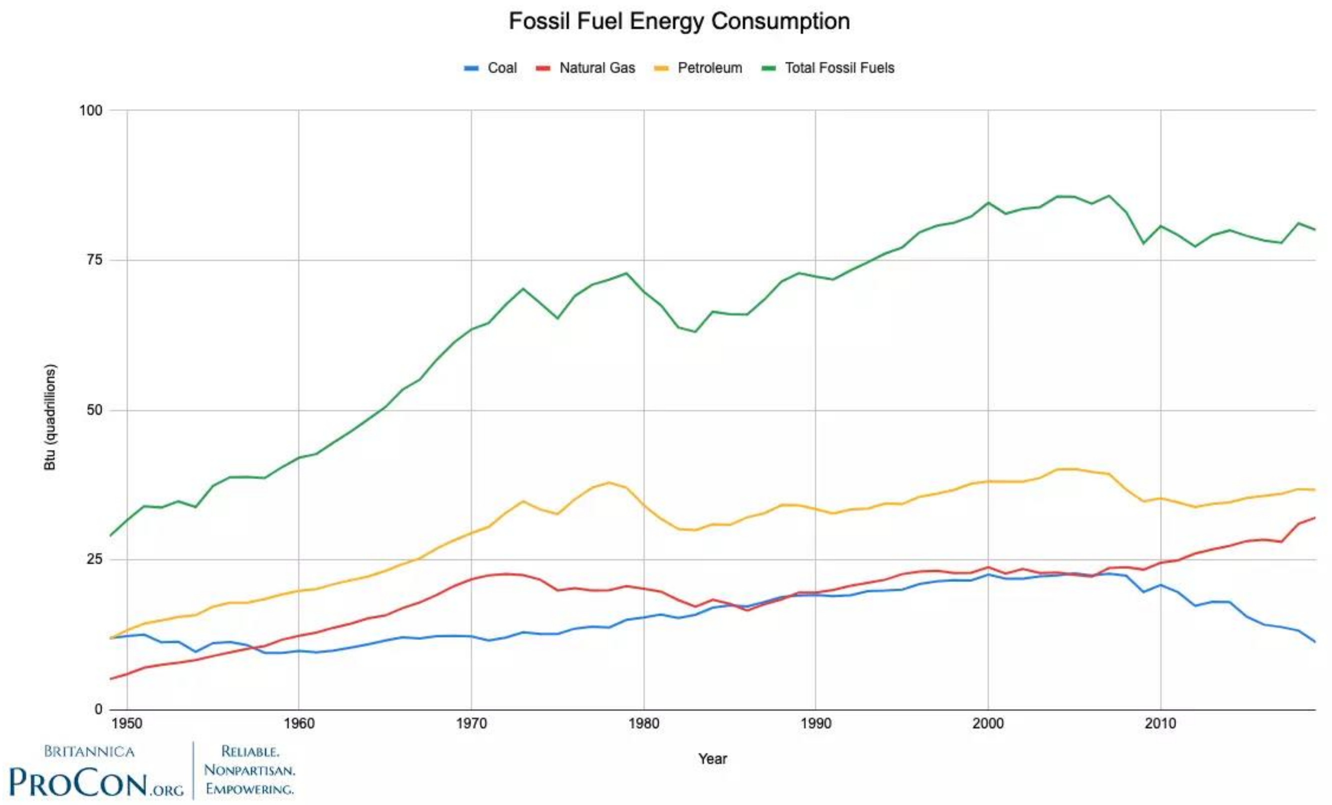
# Performance
## (precision@10)

| | Topic-relevance | Argumenta-Tiveness | Stance-relevance | -Pro | -Con |
|---|---|---|---|---|---|
| Oracle | 1.000 | 1.000 | 0.901 | 1.000 | 0.802 |
| Neural text+image 2x2class | 0.873 | 0.798 | **0.485** | 0.660 | **0.310** |
| BERT title sentiment | 0.882 | 0.804 | 0.462 | 0.674 | 0.250 |
| CLIP query stance | **0.932** | 0.830 | 0.459 | 0.662 | 0.256 |
| Aramis Formula | 0.867 | 0.790 | 0.453 | **0.690** | 0.216 |
| Both-sides baseline | 0.926 | **0.832** | 0.447 | 0.662 | 0.232 |
| Neural text+image 3class | 0.895 | 0.815 | 0.443 | 0.660 | 0.226 |
| Random baseline | 0.891 | 0.814 | 0.443 | 0.664 | 0.222 |
| Aramis Neural | 0.685 | 0.654 | 0.433 | 0.588 | 0.278 |
| Best of Touché'22 (Boromir) | 0.878 | 0.768 | 0.425 | 0.594 | 0.256 |
| Crawl query stance | 0.779 | 0.719 | 0.412 | 0.610 | 0.214 |
| AFINN text sentiment | 0.837 | 0.761 | 0.393 | 0.564 | 0.222 |
| Neural text+page 3class | 0.630 | 0.579 | 0.329 | 0.504 | 0.154 |
| Neural text 3class | 0.668 | 0.602 | 0.324 | 0.458 | 0.190 |

# Why is Stance Detection so hard?

# 1. Semantic Gap for Diagrams

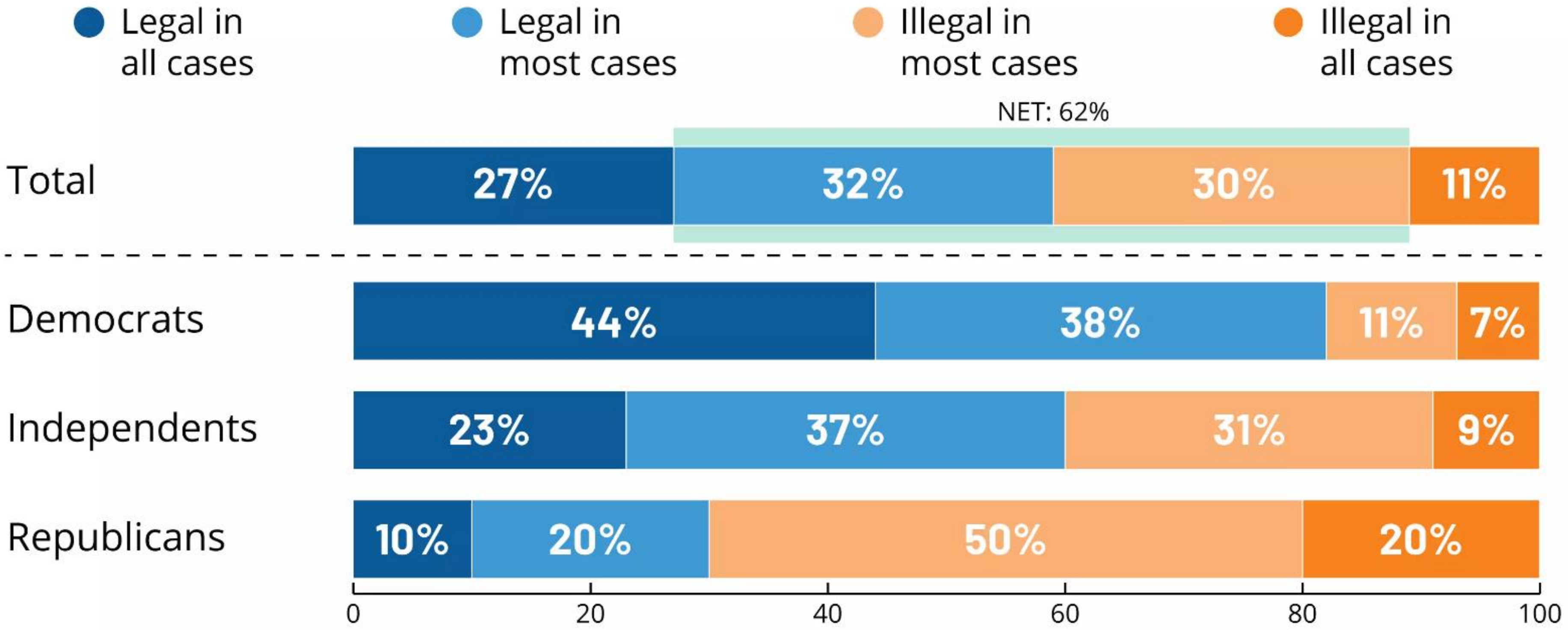
**Query: "Can alternative energy effectively replace fossil fuels?"**




ScaDS.AI DRESDEN LEIPZIG

# 2. Different valuations cause stance ambiguity

**Query: "Should abortion be legal?"**

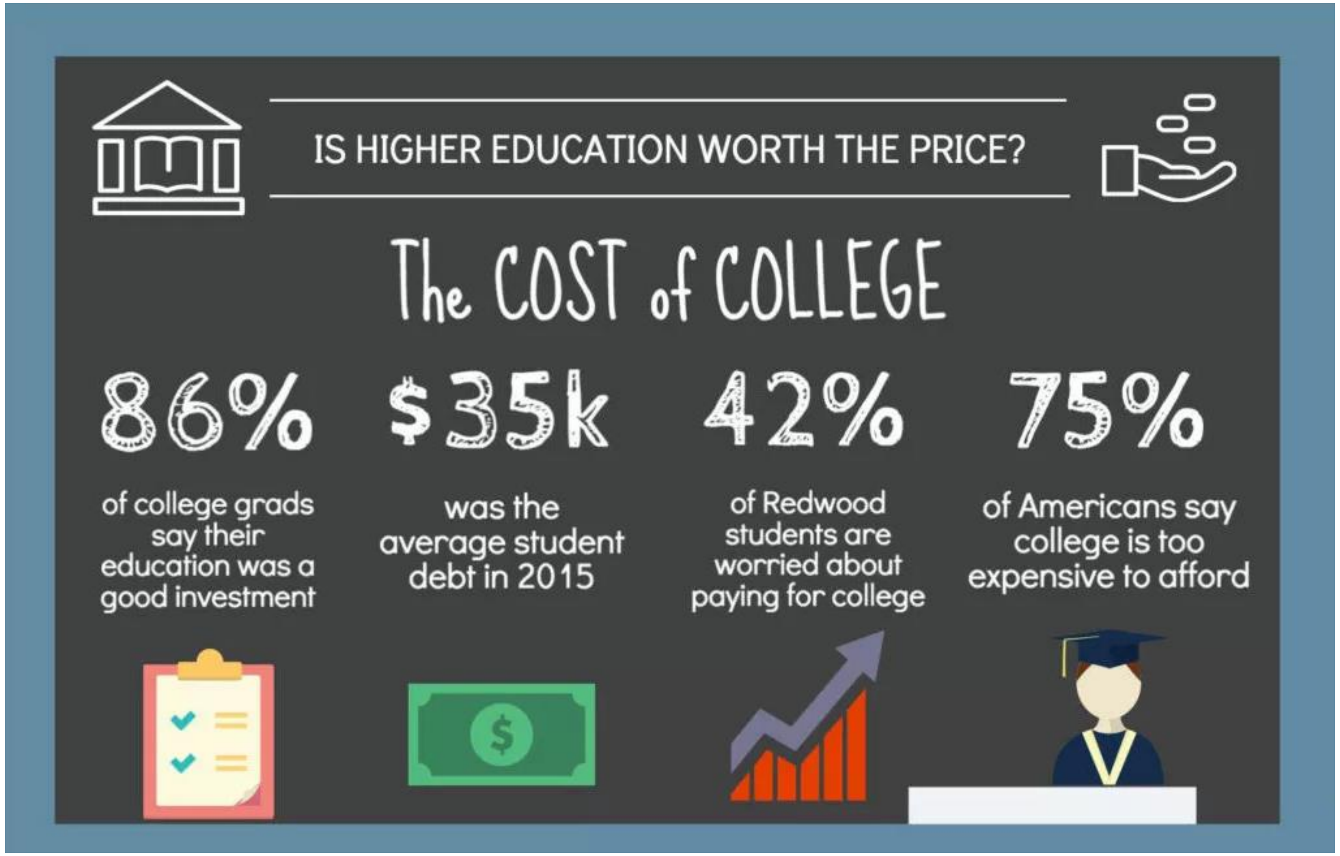


# 3. Image understanding depends on background knowledge

**Query: "Is human activity primarily responsible for global climate change?"**

# 4. Regional images

**Query: "Is a college education worth it?"**



# 5. Unbalanced image stance distribution

**Query: "Should bottled water be banned?"**

PRO



CON

?

# 6. Both stances in one image

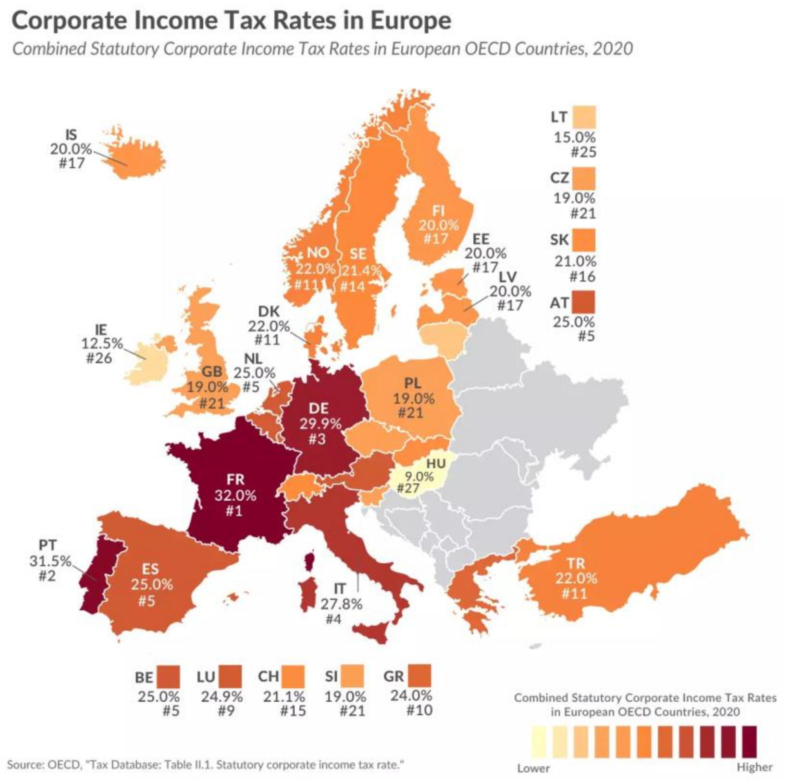
**Query: "Should adults be allowed to carry a concealed handgun?"**

"Should adults have the right to carry a concealed handgun?"

| YES | NO |
|---|---|
| 1.) Criminals less likely to attack someone that they believe might be armed. | 1.) Concealed handguns are not an effective form of self-defense. Someone carrying a gun for self-defense is 4.5 times more likely to be shot during an assault than a victim without a gun. |
| 2.) Concealed-carry laws reduce murders by 8.5%, aggravated assaults by 7%, rapes by 5%, and robberies by 3% | 2.) Concealed-carry laws lead to increases in rates of rape, robbery, and violent crime. |
| 3.) The right to carry concealed handguns is guaranteed by the Second Amendment ("Right to Bear Arms") | 3.) Ability to carry a concealed handgun NOT guaranteed by the Constitution. Second Amendment for military and militia purposes, not personal carry. |
| 4.) "Guns don't shoot people; People shoot people." | 4.) Guns are a primary tool used by people to kill people. |

# 7. Neutral images

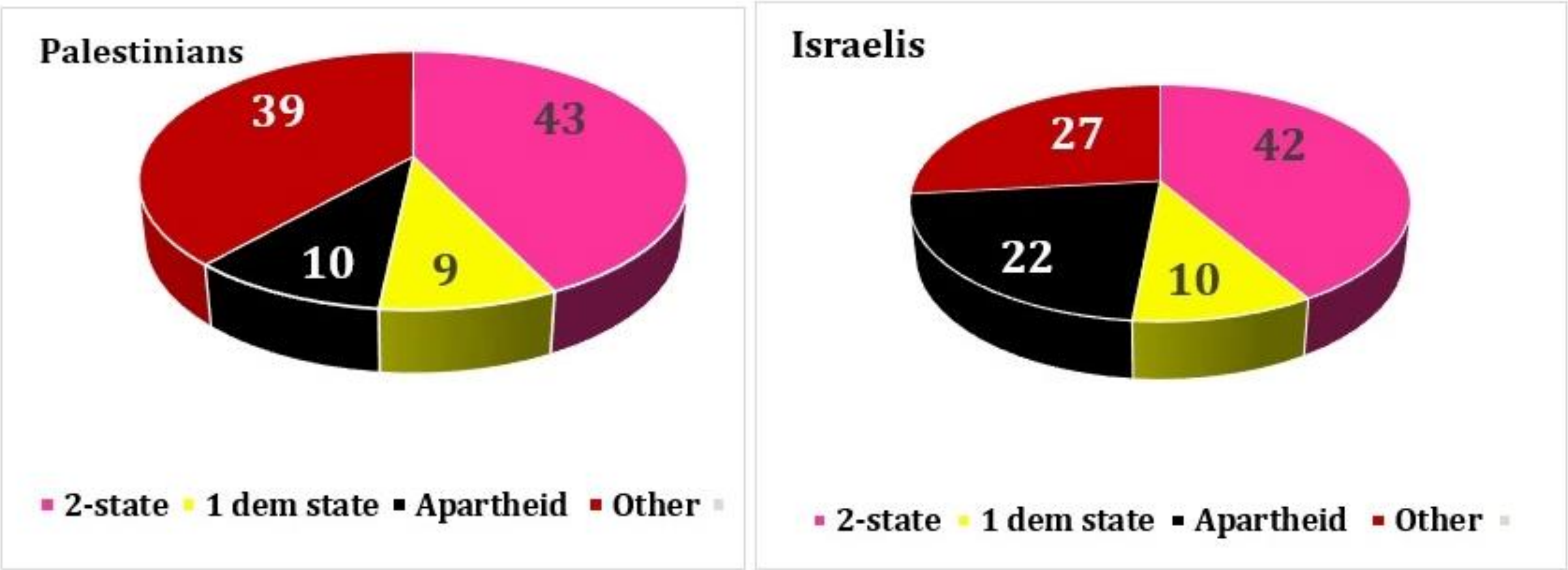
**<u>Query</u>: "Does lowering the federal corporate income tax create jobs?"**



# 8. More than two stances

**Query: "Is a two-state solution an acceptable solution to the Israeli-Palestinian conflict?"**

**Support for the two-state solution and two alternative options among Palestinians and Israeli Jews, 2020**



# 9. Irony and Jokes

**<u>Query</u>: "Do violent video games contribute to youth violence?"**

violence is introduced to
humanity for the first time
(1978)


ScaDS.AI
DRESDEN LEIPZIG

# Lessons learned

→A modular image retrieval system works very well for finding topic-relevant and argumentative images (new state-of-the-art)

→None of the 14 reproduced or new approaches can significantly beat a random baseline at stance detection

→Stance detection of images is an unsolved problem

→The task provides many different challenges