ETH zürich

Analyzing the Global Energy Discourse with Machine Learning

Malte Toetzke, Benedict Probst, Yasin Tatar, Stefan Feuerriegel, Volker Hoffmann

ETH zürich SusTec

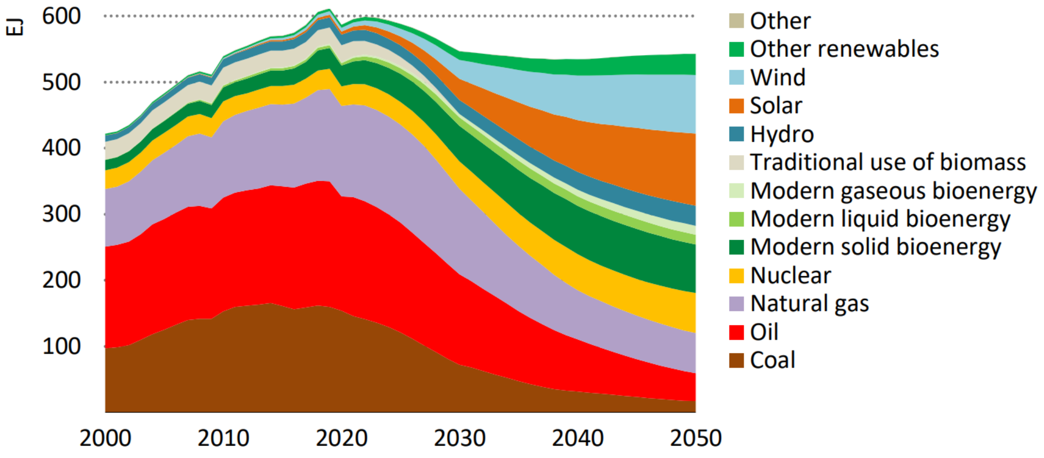
LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

LMU MUNICH SCHOOL OF MANAGEMENT

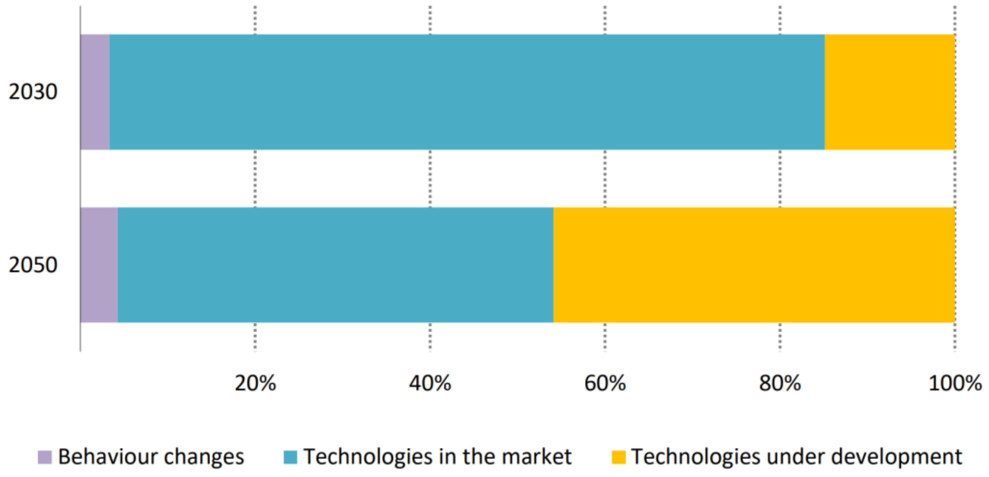
NeurIPS 2022 Workshop on Tackling Climate Change with Machine Learning

Rapid innovation and **diffusion** of clean energy technologies is crucial to replace fossil energy and capture unavoidable emissions

Total energy supply in the
Net-Zero Emissions by 2050 Scenario (IEA)



Annual CO2 emissions savings in the net zero pathway, relative to 2020



Source: IEA (https://www.iea.org/reports/net-zero-by-2050)

The media plays an **important role** in the energy transformation with the potential to both **accelerate** and **regress** industrial development

Displaying the energy transformation:

- Reflecting **public** opinions
- Describing the **political** discourse
- Featuring **technological** and **economic** trends
- …

Influencing the energy transformation:

- Attracting **demand** (e.g., for EVs, solar)
- Instigating **resistance** (e.g., against wind parks, nuclear plants)
- Putting **pressure** decision makers in industry, finance, and politics
- …

https://www.pexels.com/

Existing literature analyzing the media discourse around the energy transformation has limitations

https://www.pexels.com/

- **Qualitative** assessment
  □ **No quantitative conclusions** about associations between real-world outcomes and the media discourse possible

- **Exploratory** approaches
  □ **Lack of detail** from analyses with unsupervised models due to noisy data

- **Single country** or **single media outlet** analysis
  □ Difficult to **generalize** or **contrast** observations between countries and media outlets

We create a **comprehensive dataset** covering global newspaper articles on energy technologies

13 Energy technologies:
Biofuels, biomass, carbon capture, coal, gas, heat pumps, hydrogen, EVs, nuclear, oil, solar, synthetic fuels, wind

7 countries:
China, USA, India, Japan, Canada, Germany, United Kingdom

~22 years:
2000—2022

14 newspaper outlets:
2 major outlets of each country by circulation

~ 5 million articles

~ 40 million paragraphs

We **collect** and **process** the data in 4 steps

1. Keyword definition

Set of generic keywords for each technology

- Automated keyword extraction from samples
- Manual evaluation & extension
- Translation & validation with native speakers and domain experts

Goal: **Low false negative rate** (get all potentially relevant articles)

2. Article collection

Collecting article URLs

- Via newspaper search engines
- Based on keywords

Crawling articles for each URL

- On paragraph level
- Including metadata (e.g., date, matched technology, etc.)

3. Translation to English

Via LibreTranslate

- Self-hosted

4. Machine learning pipeline

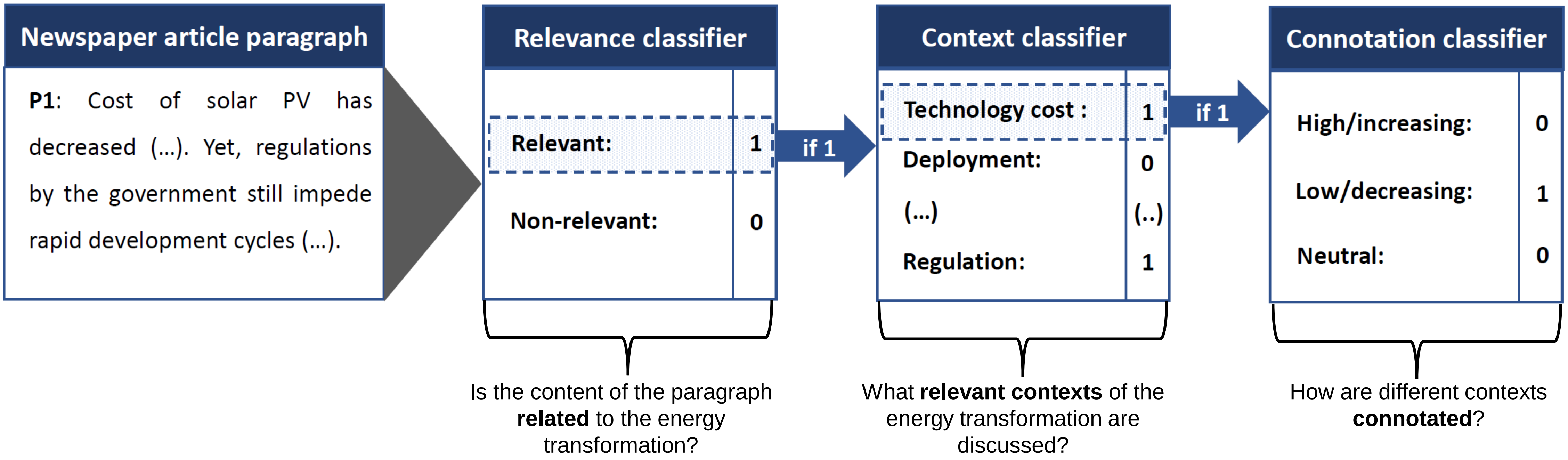
Multi-stage classification based on ClimateBERT

- Relevance classifier
- Context classifier
- Connotation classifier

Online SQL Database

The machine learning pipeline: we classify newspaper articles on paragraph level regarding relevance, context, and related connotation

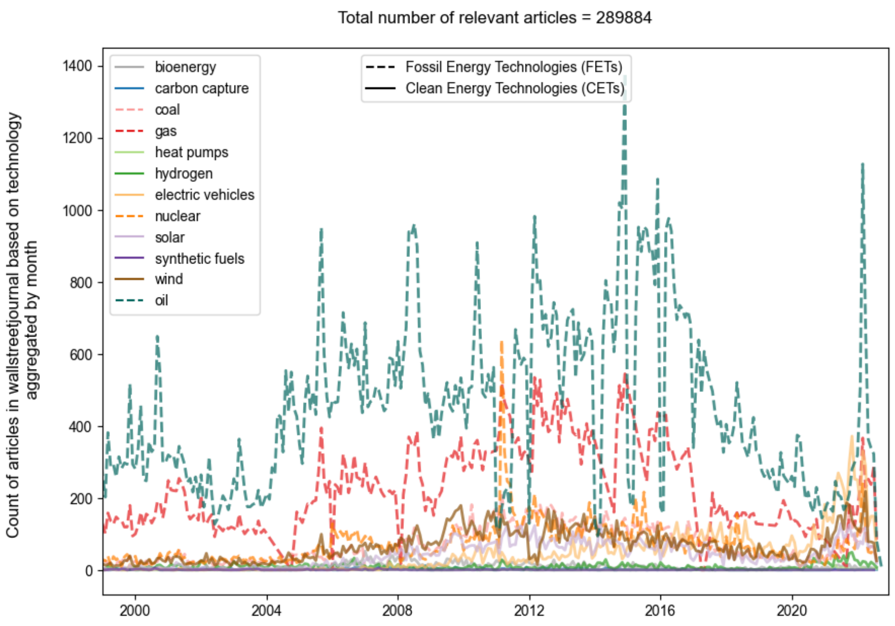
Example:



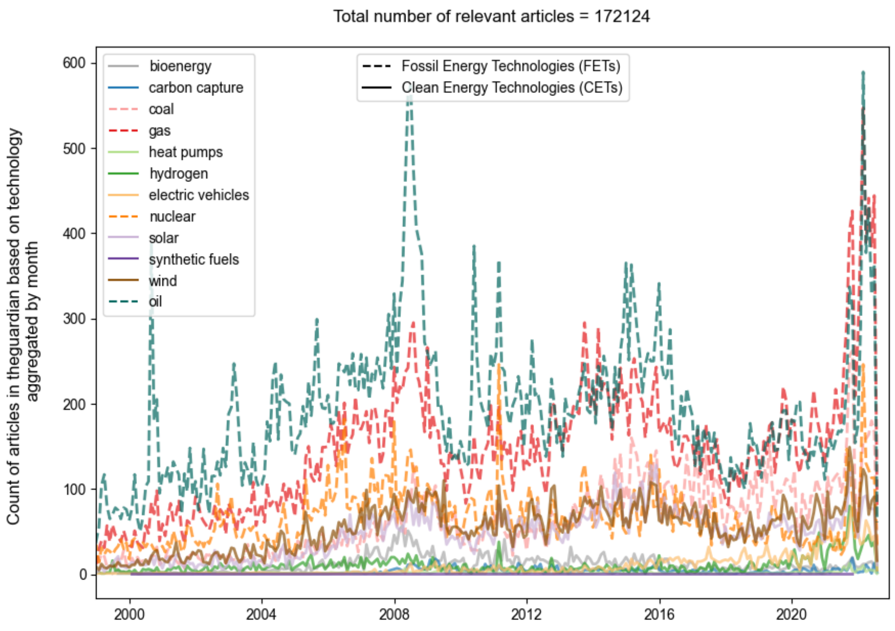
Context labels based on literature and an expert user study

Industry	Politics	Society	Technology	Climate
Manufacturing & production	Political targets & agendas	Social impact	Technology cost & price	Ghg emissions
Deployment	Climate & energy policies	Energy security	Technology efficiency	Climate change
Private investment	Regulation	Employment	Technology characteristics	
Companies	Trade	Public perception	Infrastructure	
Revenues & firm values	Science & public research	Disasters & health risks	Raw materials	
Partnerships	Public finance & investment			
Competition & markets				
Research & development				
Industry transformation				

Number of **article occurrences** per technology after relevance classification



Wallstreetjournal



The Guardian

What are these articles discussing?

Current **project status** and **next steps**

Article Collection and preprocessing

- Defining the keywords
- Crawling the articles
- Translating paragraphs

Machine learning pipeline

- Training the relevance classifier
- Training the context classifier
- Training the connotation classifier

Data analysis:

- Comparing patterns across countries and technologies
- Forecasting technology investment and deployment based on news data