



LABEL
YOUR
DATA

Case study

Academia

LINGUISTIC ANNOTATION FOR TECHNOLOGICAL UNIVERSITY DUBLIN

Overview

Kyle Hamilton from Technological University Dublin collaborated with Label Your Data for her research on rhetorical devices of propaganda used in news articles.

357

sentences

22

categories

3

annotators

Client

Kyle Hamilton is a PhD Researcher at TU Dublin who focuses on applying neuro-symbolic AI to detect propaganda in news feeds.

Challenges

Kyle needed a high-quality text dataset containing classified propaganda-related sentences to compare humans and ChatGPT in detecting propaganda in news.

Solutions

Label Your Data provided skilled annotators with linguistic background to classify and label 357 sentences using the Client's platform.

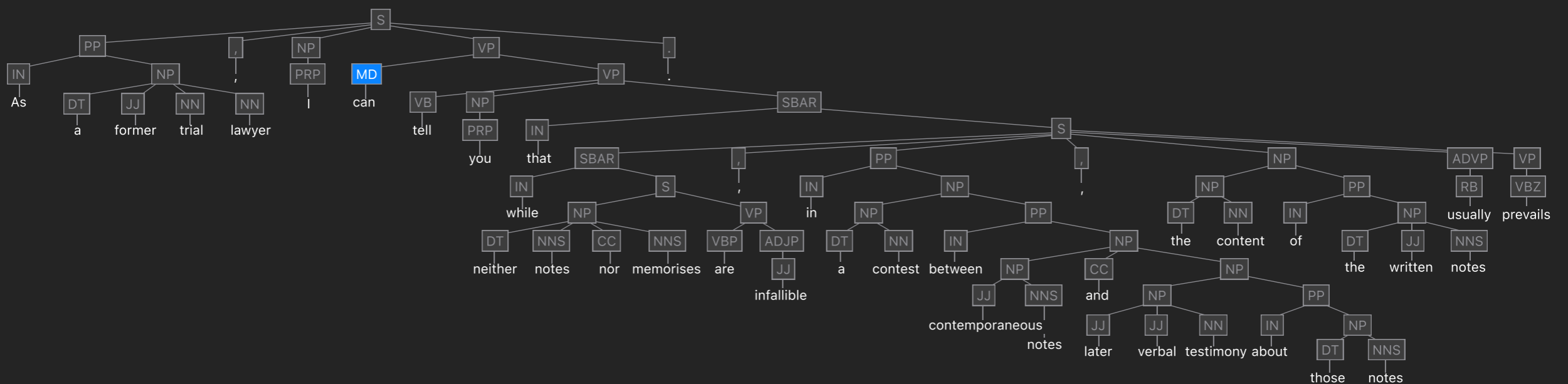
Results

Despite inconsistencies in human annotations, they prove more reliable in propaganda analysis due to ChatGPT's lack of real-world knowledge.

Client

Kyle Hamilton is a PhD Researcher at Technological University Dublin. Her research work is centered around neuro-symbolic AI for detecting propaganda in news articles. Holding a Master's in Information and Data Science and a Bachelor of Fine Art, Kyle brings a unique interdisciplinary approach to her work.

Rhetoric Annotation Tool



Challenges

Most efforts to automate the detection of propaganda and misinformation are mainly focused on using natural language processing (NLP). Following the same idea, Kyle aimed to create a tool that helps identify propaganda in news. But first, she wanted to compare ChatGPT and human linguists in performing the same task.

Kyle Hamilton needed help with:



Exploring the possibilities of using ChatGPT for automated propaganda detection.



Hiring an expert annotation team with domain experts to compare human annotations to ChatGPT's responses.



Getting high-quality annotated text dataset for the research.



Dealing with the subjective nature of the sentence classification task.



Solution

	Ratio of partial agreement among all three annotators for each feature	Chat GPT Agreement among itself when prompted 3 times
Verb choices	227	202
Tropes	253	220
Tense	308	297
Subject choices	312	306
Series	40	34
Sentence architecture	292	281
Prosody and punctuation	4	-
Predication	308	311
Phrases built on verbs	147	147
Phrases built on nouns	70	60
Parallelism	106	43
New words and changing uses	252	250
Mood	316	308
Modifying phrases	307	310
Modifying clauses	226	212
Lexical and semantic fields	316	320
Language varieties	316	315
Language of origin	316	299
Figures of word choice	213	169
Figures of argument	125	106
Emphasis	316	299
Aspect	309	294

Label Your Data helped with:



Hiring 3 dedicated data annotators with linguistic background for classification of 357 sentences.



Working in a flexible mode to seamlessly integrate into Kyle's annotation platform.



Conducting a cross-reference QA to address the project's subjectivity.



Delivering high-quality annotated text corpus.



Results

This initial phase aimed at supporting Kyle Hamilton's research and securing the grant. While findings are initial, Label Your Data expects a larger dataset to annotate.

Though human experts' annotations vary, their real-world knowledge surpasses AI models like ChatGPT, which solely relies on internet-derived data.

1.

Analyzed the discrepancies in agreement between annotators and ChatGPT.

2.

Identified the most challenging areas to achieve consensus:

- Series
- Prosody and punctuation
- Phrases built on verbs
- Parallelism
- Figures of word choice
- Figures of argument

3.

Despite inconsistencies, defined human annotators as more reliable for propaganda analysis.

4.

Facilitated research on automated propaganda detection.





“

Working with Label Your Data has been a truly positive experience. The annotators were genuinely interested in the success of my project, asked good questions, and were flexible in working in my proprietary software environment.

”

Kyle Hamilton
PhD Researcher at TU Dublin



Run a free pilot and see for yourself!



Security-compliant

NLP often implies working with sensitive data, thus it's important to cooperate with secure annotation service providers.



Domain Expertise

Text annotation requires linguistic knowledge to produce quality datasets for NLP model training. We can assemble SMEs with the necessary skills needed for your project.



Tool-agnostic

NLP labeling demands the use of specialized tooling. We offer our in-house solution but also can seamlessly work within your own labeling software.



No commitment

To ensure that every point mentioned above holds true, you can send your data to us and receive the first batch annotated for free.