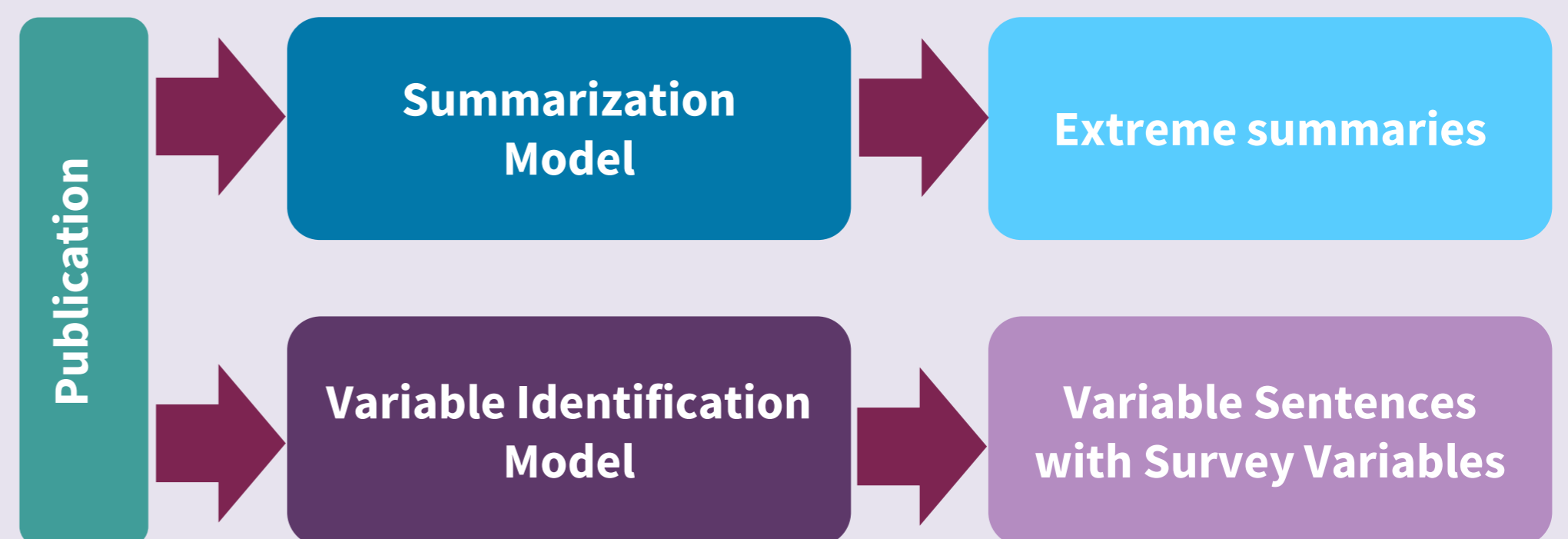


# VADIS – a Variable Detection Interlinking and Summarization System

Yavuz Selim Kartal, Tornike Tsereteli, Sotaro Takeshita,  
Benjamin Zapilko, Andrea Zielinski, Philipp Mayr

- The VADIS system aims to allow for searching and using and help to increase the reproducibility of research survey variables (SV) in context and thereby enhance information access of scholarly publications results.
- Identification of links between publications and the specific variables that are covered in the surveys
- Extreme abstractive and extractive summaries (TLDR) of publications



## Summarization Models

Summarization models are trained on a cross-lingual summarization dataset. They accept English and German documents as inputs and generate summaries in both languages. The models produce two types of summarization:

- Extractive summarization: it selects one salient sentence as a summary in a given document.
- Abstractive summarization: it generates a short summary of a given document.
- Approaches:
  1. English-English: BART-large [Lewis et. al., 2019] fine-tuned on SciTLDR [Cachola et. al., 2020]
  2. English-German: Cross-lingual summarization: mBART-large [Liu et. al. 2020] fine-tuned on X-SciTLDR [Takeshita et. al., 2022]

Model	Rouge-1	Rouge-2	Rouge-L
Eng-Eng	44	22	37
Eng-German	27	9	23

\*ROUGE is a recall of N-grams in human summary in system summary. Higher scores indicate system summary has more N-grams from human summary

## Variable Identification Model

This model proposes an integrated approach for the Variable Identification task by splitting it into the following subtasks:

### Mention Detection

- Identifying whether a sentence contains a variable mention or not. (binary text classification task)

Sentence	Label
In the EU27, support for development aid being very important increased by 10.5%, from 35.9% in 2011 to 39.7% in 2015.	1
To sum up: Less than a third of the polls would predict all parties within their alleged precision.	0

### Approaches

1. Fine tuning Pre-Trained Language Models (PLM)
2. Dataset enlargement with LLM generated samples + Fine tuning PLMs
3. Large Language Models (ongoing work)

Model	P	R	F1
SciBERT	89	37	52
distilBERT	<b>91</b>	38	54
RoBERTa	80	5	21
Llama 2 (Zero Shot)	76	<b>71</b>	<b>73</b>

### Variable Disambiguation

- Identifying which variable from a given vocabulary is mentioned in a sentence.
- This task relies on the sentences labeled as positive (mentioning a variable) from the previous task
- This is viewed as an information retrieval (IR) task, where the goal is to identify all relevant variables for a given sentence

Sentence	Variable Id
In the EU27, support for development aid being very important increased by 10.5%, from 35.9% in 2011 to 39.7% in 2015.	ZA6644_Varqa1 ZA5565_Varqd1
The question is intended to test whether respondents are in favour of a reversal of an arrangement by which men go out to work and women stay at home and care for the children.	ZA2620_VarV13

### Approach:

1. Semantic Textual Similarity methods with kNN

Model	MAP@10	R-Prec
BM25	9.43	9.56
miniLM (sentence-t)	<b>18.93</b>	<b>16.65</b>
T5 (sentence-t)	13.59	12.59

## Sample Publication

### Fear of Robots and Life Satisfaction

Abstract. The use of robots and in particular next-generation robots in the production of goods and services is increasing. What impact robots are having on people's quality of life, including workers, is as yet under-explored. This paper provides initial findings examining whether fear of robots is correlated with one aspect of quality of life: life satisfaction....

There are several questions in the survey that relate to perceptions of robots. **The first is a general question on how the person views robots which has four possible responses, "Generally speaking, do you have a very positive, fairly positive, fairly negative or very negative view of robots?". There are then a series of questions in which people are asked "To what extent they agree or disagree with whether (1) robots are a good thing for society, because they help people (2) robots steal peoples' jobs and (3) robots are necessary as they can do jobs that are too hard or too dangerous for people.** There are four possible responses to these three questions. All people in the survey will answer these questions, but others in the survey are related more to people who are currently employed.

## VADIS Data

### Extreme Summary (abstractive):

We show that fear of robots correlates with lower reported life satisfaction.

**Study:** ZA5933 - Eurobarometer 82.4 (2014)  
**Question text:** Generally speaking, do you have a very positive, fairly positive, fairly negative or very negative view of robots?  
**Answer categories:**  
Value label  
Very positive  
Fairly positive  
...

**Study:** ZA5933 - Eurobarometer 82.4 (2014)  
**Question text:** Here is a list of things that could be done by or with robots. For each of them, please tell me, using a scale from 1 to 10, how you would personally feel about it.  
**Item categories:**  
Having a robot assist you at work (e.g. in manufacturing)  
...

## Outcomes and Future Works

- We processed 607 English and German Open Source Social Science Publications. We present them via a demo UI (demo-vadis.gesis.org) with useful features including document search and direct access.
- We conducted a user study with 6 social scientists and they reported summaries and detected variables facilitate their works.
- We observed leveraging LLMs yield promising results on the preliminary experiments. We further plan to fine tune/train them more in the context of social science to increase reusability of variables.