Towards Reducing Misinformation and Toxic Content Using Cross-Lingual Text Summarization

ROMCIR 2023

The 3rd Workshop on Reducing Online Misinformation through Credible Information Retrieval – April 2nd, Dublin, Ireland

Hoai Nam Tran and **Udo Kruschwitz**
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Universität Regensburg

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Introduction

- Fake News and Hate Speech spread toxicity multilingually
- Users get persuaded and influenced by social media posts
- Excessive usage of social media can lead to mental health issues
  - Desire for validation
  - Fear of rejection

- Limited input size in Transformer models (e.g. 512 tokens for BERT)
- Which language(s) can really be considered "resource-rich" except English?

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Occurrence of languages present over 1,836 tasks

1. English (>1000 tasks)
2. Spanish (~60 tasks)
3. Japanese (~60 tasks)

...

10. German (~30 tasks)

Can German be considered as a "resource-rich" language?

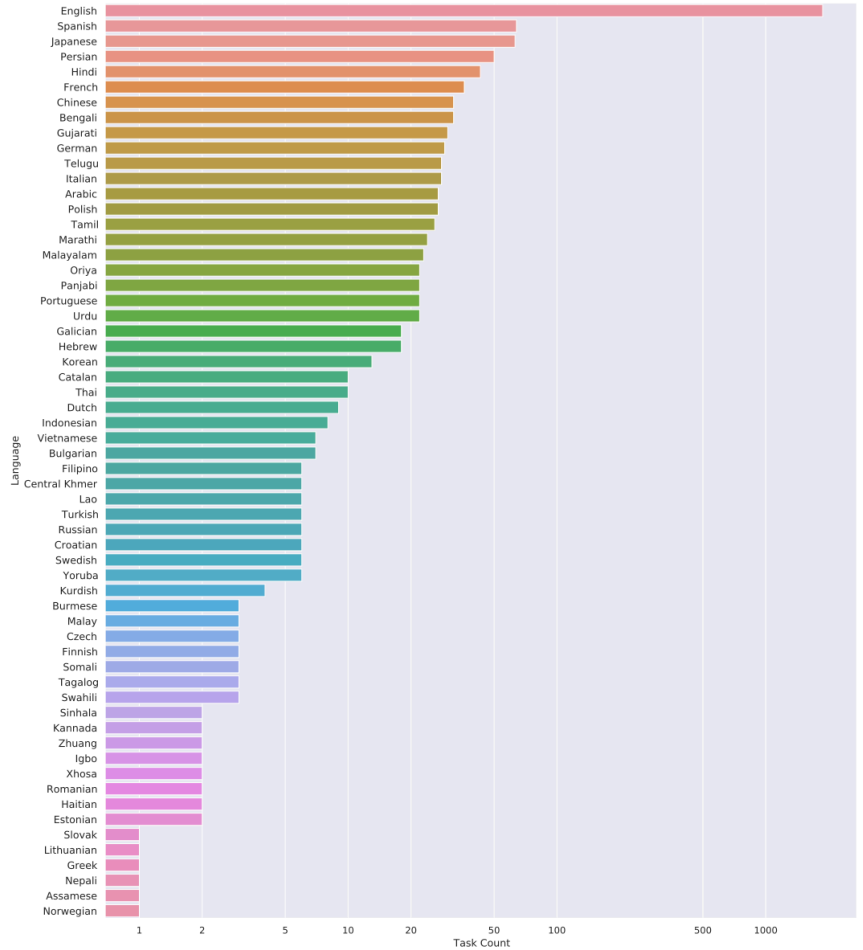


Figure 18: The **languages represented across tasks** in our full multi-task finetuning data suite.

Figure from: Chung, H. W., Hou, L., Longpre, S., Zoph, B., Tay, Y., Fedus, W., ... & Wei, J. (2022). Scaling instruction-finetuned language models. *arXiv preprint arXiv:2210.11416*.

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Datasets

- Binary Classification Tasks:
  - GermEval 2018 Subtask 1
  - GermEval 2019 Task 2 Subtask 1
  - GermEval 2021 Subtask 1-3

Short text (Tweets, Comments)

- Multi-Class Classification Task:
  - CLEF 2022 CheckThat! Lab Task 3

Long text (News articles, Blog posts)
max. 100,000 characters

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Dataset Sizes

Table 1
GermEval Dataset Sizes

		GermEval 2021			GermEval 2018	GermEval 2019 Task 2
Dataset	**Label**	**Subtask 1**	**Subtask 2**	**Subtask 3**	**Subtask 1**	**Subtask 1**
Training	True	1122	865	1103	1688	1287
	False	2122	2379	2141	3321	2707
Test	True	350	253	314	1202	970
	False	594	691	630	2330	2061

Table 2
CLEF CheckThat! 2022 Dataset Sizes

	CLEF CheckThat! 2022			
	Training Set	Development Set	Test Set	
Label	**Subtask 3**	**Subtask 3**	**Subtask 3A**	**Subtask 3B**
True	142	69	210	243
False	465	113	315	191
Partially False	217	141	56	97
Other	76	41	31	55

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Macro F1 and Macro F1 (Opitz and Burst, 2021)

Averaged F1: arithmetic mean over harmonic means

$$\mathcal{F}_1 = \frac{1}{n}\sum_x \mathrm{F1}_x = \frac{1}{n}\sum_x \frac{2P_xR_x}{P_x + R_x}$$

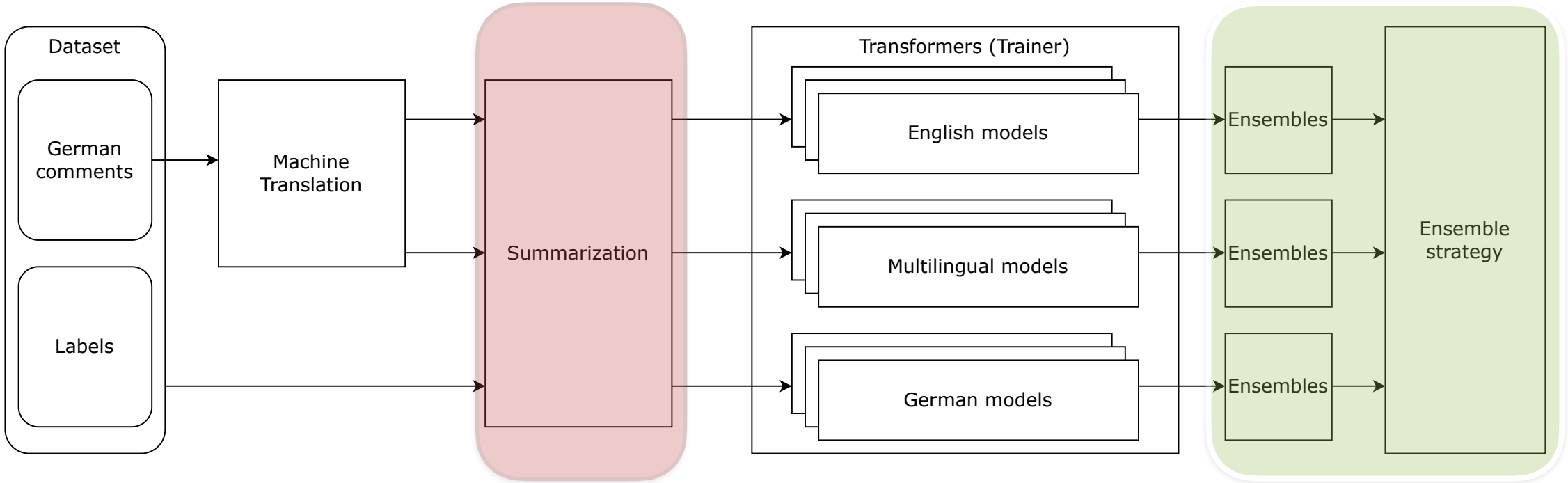
F1 of averages: harmonic mean over arithmetic means

$$\mathbb{F}_1 = H(\bar{P}, \bar{R}) = \frac{2\bar{P}\bar{R}}{\bar{P} + \bar{R}} = 2\frac{(\frac{1}{n}\sum_x P_x)(\frac{1}{n}\sum_x R_x)}{\frac{1}{n}\sum_x P_x + \frac{1}{n}\sum_x R_x}$$

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

System Architecture



Our experiments:

Multi-Class Classification

Binary Classification

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Automatic Machine Translation

Table 5
Machine Translation Performance

Translation Service	Run 1	Run 2	Run 3	Run 4	Run 5	Hard	Soft
Google Translate	69.48	67.08	67.67	68.74	68.28	68.39	68.42
DeepL Translator	70.01	70.22	69.24	68.26	67.67	70.13	70.09

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Summarization

Input:

$$title + .\_ + text$$

- **Extractive** Summarization with DistilBART-CNN-12-6
- **Abstractive** Summarization with T5-3B

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Example: CNN Praises Taliban For Wearing Masks During Attack

KABUL—Approximately twelve minutes after U.S. troops withdrew from Afghanistan, Taliban fighters have completely taken over the entire country. "Woah, that's a bummer," said the Biden Administration's foreign policy team. "We didn't see that one coming." As the Taliban began its campaign of shooting and killing, as is their time-honored tradition, CNN anchors gushed with praise after noticing all the Taliban fighters were responsibly wearing masks to protect themselves and others from COVID. "Wow! In the midst of the battle and bloodshed, these noble desert knights of Islamic superiority are wearing masks! Bravo!" said Brian Stelter. TV anchor and world-renown polemicist Don Lemon was also quick to weigh in. "All things considered, we ought to be praising the COVID-safe masks these majestic mujahideen warriors are wearing," he said. "They are showing all of us the proper way to behave during a pandemic—something those horrible idiot Trump supporters don't seem to get." Inspired by their example, the Biden Administration has invited the Taliban to the White House to record TikTok videos in hopes of convincing Trump supporters to get vaccinated. (**1.161 chars, without title**)

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Extractive Summarization with DistilBART-CNN-12-6

CNN Praises Taliban For Wearing Masks During Attack. As the Taliban began its campaign of shooting and killing, as is their time-honored tradition, CNN anchors gushed with praise after noticing all the Taliban fighters were responsibly wearing masks to protect themselves and others from COVID. " All things considered, we ought to be praising the COVID- safe masks these majestic mujahideen warriors are wearing," he said. (**424 chars**)

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Abstractive Summarization with T5-3B

taliban fighters in afghanistan wearing masks to protect themselves from covid. cnn's brian stelter: "wow! in the midst of the battle and bloodshed, these noble desert knights of islamic superiority are wearing... bravo!" tv anchor and polemicist don lemon: 'they are showing all of us the proper way to behave' (**311 chars**)

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Hyperparameter Search with Optuna

➡ 100 "random" trials with different parameters in certain ranges

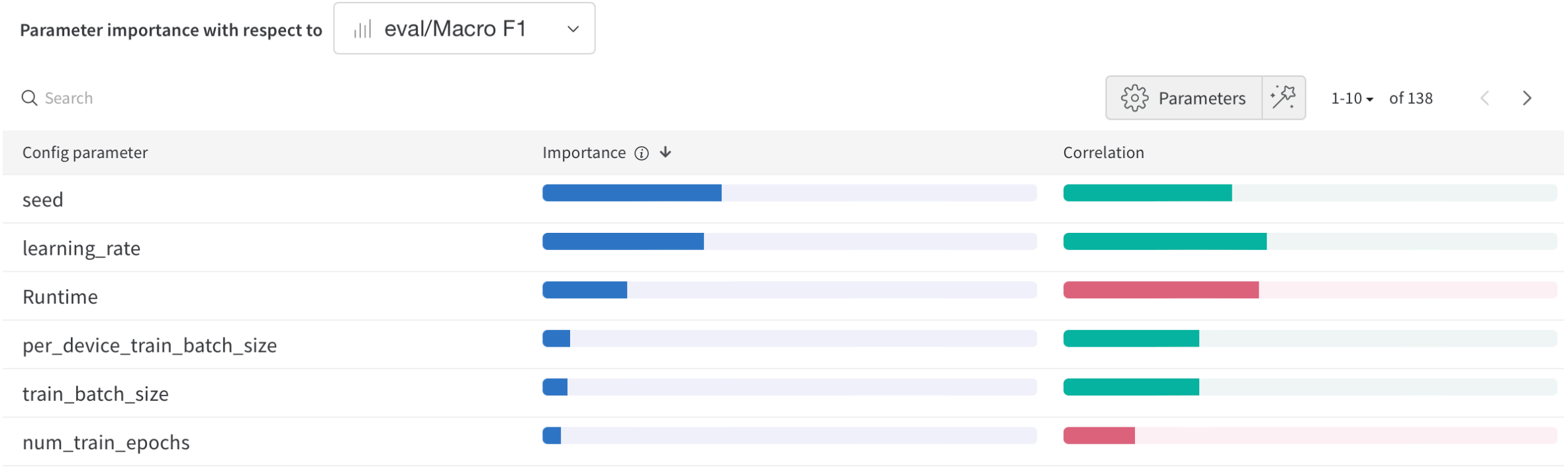
Table 6
Random Hyperparameter Tuning with Optuna

Run	Dataset	With Hyperparameter Tuning	Without Hyperparameter Tuning
1	Development Set	70.49	70.57
	Test Set	67.61	67.67
2	Development Set	70.44	70.43
	Test Set	68.15	70.22
3	Development Set	69.81	69.81
	Test Set	67.25	68.26

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Parameter importance of Optuna trials with W&B



Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Splitting Methods

Table 7
Splitting Strategy

Splitting Strategy	Run 1	Run 2	Run 3	Run 4	Run 5	Hard	Soft
Stratified K-Fold Cross-Validation	68.78	67.79	68.21	69.62	67.29	68.99	69.59
Random Seed	70.01	70.22	69.24	68.26	67.67	70.13	70.09

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Results

GermEval '18 ST1: **+0.78%**
GermEval '19 T2 ST1: **+5.41%**
GermEval '21 ST1: **+4.48%**
GermEval '21 ST2: **+0.38%**
GermEval '21 ST3: **+1.56%**

Preferred Ensembling Strategy:
Majority Voting

Table 8
Binary Classification on GermEval datasets (new best performance in **bold**)

	GermEval '18 Subtask 1		GermEval '19 T2 Subtask 1		GermEval '21 Subtask 1		GermEval '21 Subtask 2		GermEval '21 Subtask 3	
Model	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**
$GBERT_{base}$	76.28	75.91	76.50	76.64	68.11	67.84	68.22	68.30	74.23	74.78
$GELECTRA_{base}$	75.45	75.37	75.15	74.92	69.38	69.68	67.96	67.60	76.52	77.11
$BERTweet_{base}$	78.02	78.05	77.23	77.44	70.13	70.09	68.23	68.84	75.51	75.47
$BERT_{base}$	77.23	77.17	76.63	76.55	64.68	64.71	68.89	69.39	73.36	72.71
$XLM\text{-}R_{base}$ (de)	75.71	76.00	75.51	75.17	67.37	67.21	68.49	67.90	73.84	74.26
$XLM\text{-}R_{base}$ (en)	76.67	77.04	77.35	77.11	68.24	68.20	69.11	69.72	74.35	74.61
$GBERT_{large}$	80.74	80.63	80.06	80.23	72.09	72.69	69.45	68.89	75.77	76.10
$GELECTRA_{large}$	80.06	79.85	80.80	80.79	71.62	71.72	70.16	70.24	75.06	74.26
$BERTweet_{large}$	79.97	79.86	79.56	79.86	73.60	72.24	69.82	**70.36**	75.14	75.48
$BERT_{large}$	78.34	78.32	77.79	77.79	67.00	65.26	69.86	69.47	74.58	75.07
$XLM\text{-}R_{large}$ (de)	-	-	-	-	69.04	69.12	69.51	68.60	76.36	76.82
$XLM\text{-}R_{large}$ (en)	-	-	-	-	71.71	71.48	68.77	69.99	76.54	77.44
Ensemble	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**	**Hard**	**Soft**
Gradient Boosting	79.97	80.95	80.28	81.77	**76.23**	74.03	68.25	69.47	75.82	76.12
Logistic Regression	79.97	80.91	81.14	81.52	74.11	75.09	68.56	69.65	75.61	74.03
Majority Voting	80.99	**81.48**	82.06	**82.36**	75.22	74.72	69.22	70.09	**77.82**	76.89
SOTA	80.70 [40]		76.95 [49]		71.75 [50]		69.98 [51]		76.26 [50]	

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Results

Test 3A: **+5.63%**
Test 3B: **+1.07%**

Table 9
Multi-Class Classification on CheckThat! 2022 dataset (new SOTA in **bold**)

Summarization Model	Classification Model	Run Nr.	Dev	Dev-Test	Test 3A	Test 3B
DistilBART-CNN-12-6 (extractive)	$BERT_{large}$	1	52.40	52.18	28.33	28.99
		2	46.43	39.96	26.87	19.46
		3	48.77	52.78	30.70	28.69
		4	49.21	48.44	32.31	25.32
		5	53.25	51.85	30.19	20.46
	$XLM\text{-}R_{large}$	1	50.53	41.04	30.42	27.40
		2	50.93	44.54	33.11	28.01
		3	49.08	48.56	30.82	26.09
		4	50.80	43.99	28.23	21.94
		5	50.95	40.29	32.47	23.34
	T5-3B	1	48.05	46.52	**39.54**	29.58
T5-3B (abstractive)	$BERT_{large}$	1	56.33	51.15	28.89	21.34
		2	45.85	37.87	32.88	23.43
		3	55.08	46.80	35.24	28.33
		4	52.15	47.08	36.48	27.01
		5	51.32	46.91	30.56	21.77
	$XLM\text{-}R_{large}$	1	51.54	44.81	31.66	28.99
		2	49.36	42.84	35.63	**30.06**
		3	49.73	44.91	35.67	27.82
		4	50.59	44.79	36.01	26.86
		5	51.78	40.25	35.29	28.09
	T5-3B	1	52.08	43.82	29.72	23.72
SOTA					33.91 [52]	28.99 [53]

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

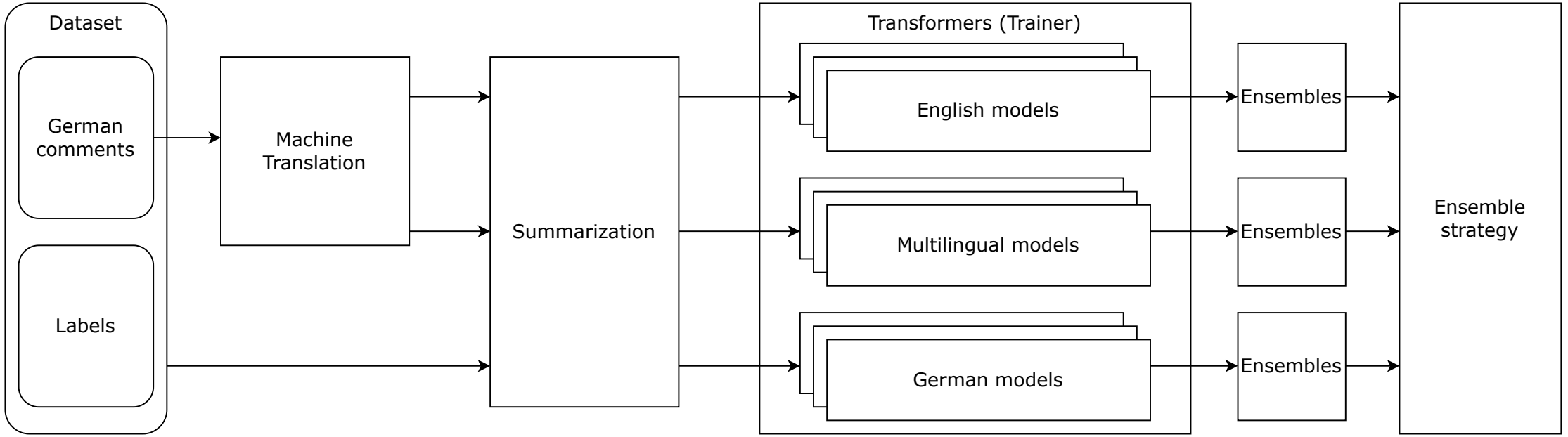
Limitations and Future Work

- Usage of full system architecture for all datasets
- Performance difference between translation services on all datasets
- Dataset quality (Inter-rater reliability)
- Other datasets, other languages
- Other (L)LMs (ChatGPT, GPT-4, Longformer, etc.)

Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Conclusion / System Architecture



Universität Regensburg

Hoai Nam Tran
Lehrstuhl für Informationswissenschaft
FAKULTÄT FÜR INFORMATIK UND DATA SCIENCE

Thank you for your attention!