
Current Challenges

Philipp Koehn

31 October 2024






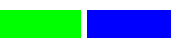

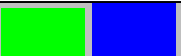
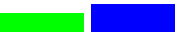




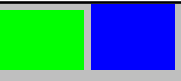








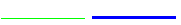
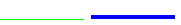








WMT 2024: LLMs Arrive



English Japanese	English Spanish	English Hindi	English Icelandic	English Czech	English German	English Chinese	Czech Ukrainian
Human	Human	TranssionMT	Human	Human	GPT-4	Human	Human
Online-B	Dubformer	Tower-70B	Dubformer	Tower-70B	Dubformer	Online-B	Claude 3.5
Command-R+	GPT-4	Claude-3.5	Claude-3.5	Claude-3.5	Tower-70B	IOL-Research	Gemini-1.5-Pro
GPT-4	IOL-Research	Online-B	Tower-70B	Online-W	Online-B	Tower70B	Tower70B
Claude-3.5	Mistral-Large	Gemini-1.5-Pro	AMI	CUNI-MH	TranssionMT	GPT-4	IOL-Research
Gemini-1.5-Pro	Tower-70B	GPT-4	IKUN	Gemini-1.5-Pro	Human-B	Gemini-1.5-Pro	Online-W
Tower70B	Claude-3.5	Human	Online-B	GPT-4	Mistral-Large	Claude-3.5	Command-R+
IOL-Research	Gemini-1.5-Pro	IOL-Research	GPT-4	Command-R+	Command-R+	Command-R+	GPT-4
Aya23	Command-R+	Llama3-70B	IKUN-C	IOL-Research	Online-W	Llama3-70B	IKUN
NTTSU	Llama3-70B	Aya23	IOL-Research	SCIR-MT	Claude-3.5	HW-TSC	GPT-4
Team-J	Online-B	IKUN-C	Llama3-70B	CUNI-DocT	Human-A	Aya23	CUNI-Transf
Llama3-70B	IKUN			Aya23	IOL-Research	IKUN	IKUN-C
IKUN-C	IKUN-C			CUNI-GA	Gemini-1.5-Pro	IKUN-C	
	MSLC			IKUN	Aya-23		
				Llama3-70B	Online-A		
				IKUN-C	Llama3-70B		
					IKUN		
					IKUN-C		

NMT
LLM
Research System

Domain Mismatch

System ↓	Law	Medical	IT	Koran	Subtitles
All Data	 30.5 32.8	 45.1 42.2	 35.3 44.7	 17.9 17.9	 26.4 20.8
Law	 31.1 34.4	 12.1 18.2	 3.5 6.9	 1.3 2.2	 2.8 6.0
Medical	 3.9 10.2	 39.4 43.5	 2.0 8.5	 0.6 2.0	 1.4 5.8
IT	 1.9 3.7	 6.5 5.3	 42.1 39.8	 1.8 1.6	 3.9 4.7
Koran	 0.4 1.8	 0.0 2.1	 0.0 2.3	 15.9 18.8	 1.0 5.5
Subtitles	 7.0 9.9	 9.3 17.8	 9.2 13.6	 9.0 8.4	 25.9 22.1

Robustness to User Generated Content

English ↔ German

daily content of #scaramouche from genshin impact #原神 ★ mute #mouchecc for no cc tweets ! not leak free ★ <http://dailymouche.carrd.co> ×

täglicher Inhalt von #scaramouche von genshin impact #原神 ★ stumm #mouchecc für keine CC-Tweets! nicht auslaufsicher ★ <http://dailymouche.carrd.co>

Challenges



- Jargon and acronyms
- Misspellings (sometimes intended for effect)
- Mangled grammar
- Special symbols (emojis, etc.)
- Hashtags, URLs, ...
- Use of dialectical languages
- Use of non-standard writing systems (e.g., Latin script due to lack of keyboard)

Some Methods

- Special handling of non-words like emojis, hashtags, URLs
- Creating synthetic noisy training data
- Adversarial training
- Resources
 - Machine translation of noisy text data set (MTNT)
 - WMT 2020 Shared Task on Machine Translation Robustness