

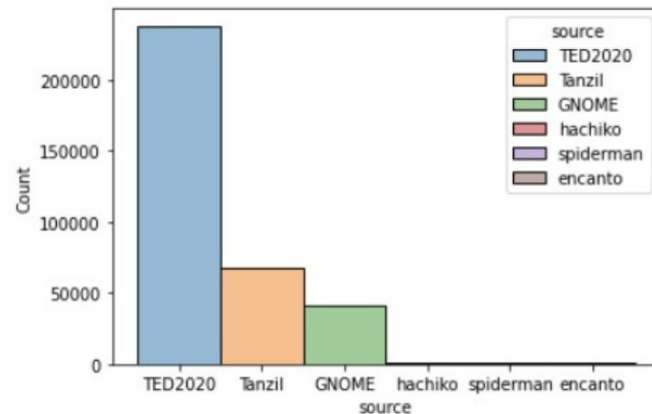
# **Persian - Romanian automated translation**

# Motivation



- Rare language combination
- FAGMA uses pivot strategy (proof in the document)
- Existing results could be better...
- Cloud services compromise your privacy

# Methods



- Helsinki-NLP/mul-to-en pretrained model
- Transfer learning with Hugging Face library
- OpenNMT datasets
- Mining from movie subtitles



# Methods



- Kaggle for computing power
- Azure Cloud for Microsoft translation

# Main Results



- 35.55 BLEU score compared to Microsoft's 12.33
- ... using Microsoft's own model architecture

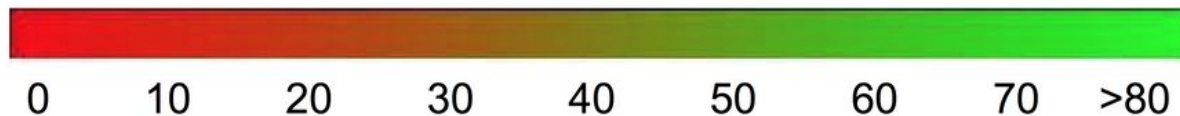


# Results

- good quality translation
- small to no grammar errors

BLEU Score	Interpretation
< 10	Almost useless
10 - 19	Hard to get the gist
20 - 29	The gist is clear, but has significant grammatical errors
30 - 40	Understandable to good translations
40 - 50	High quality translations
50 - 60	Very high quality, adequate, and fluent translations
> 60	Quality often better than human

The following color gradient can be used as a general scale [interpretation of the BLEU score](#):



# Results

Persian text	Our translation	Ground Truth
زرد	Galben	Galben
و در اینجا می بینید، که این با بافت همسان شده است.	Și aici vedeți, este identic cu țesutul.	Așa cum puteți vedea, este reintegrată în țesut.
چه زمانی با رقبا روبرو می‌شویم تا با آنها مبارزه کنیم؟	Când ne confruntăm cu competitorii pentru a lupta cu ei?	Când ne întâlnim competitorii pentru a ne lupta cu ei?
بنابراین تنها یک راه وجود داشت -- ستون را از کامپیوتر خارج کنی.	Așa că exista o singură cale: scoți coloana din calculator.	Prin urmare, exista o singură posibilitate - să scot coloana din computer.
و هر کس از یاد (خدای) رحمان دل بگرداند، برایش شیطانی بر می‌گماریم؛ پس برای وی قرین و همنشین است.	Celui ce se abate de la amintirea Milostivului, Noi îi vom da un diavol ca însoțitor.	Celui ce se abate de la amintirea Milostivului, Noi îi vom da un diavol ca însoțitor.



## Results

Model	Dataset	BLEU score
Helsinki-NLP model with Transfer learning ( <i>ours</i> )	Movie Corpus	1.07
Azure Translator (direct)	Clean corpora	12.33
Azure Translator (with english pivot)	Clean Corpora	12.34
Helsinki-NLP model with Transfer learning ( <i>ours</i> )	Clean Corpora	<b>35.55</b>

+99.44 BLEU score between azure\_pivot and azure\_without\_pivot



# Discussion

- Don't use pivot languages. Use Transfer Learning instead!
- Real World Ready Machine translation is accessible to YOU





# Contributions

Us: Lida Ghadamiyan,  
Smit Oanea  
Bogdan Ouatu

- Real world usable model
- Reproducible
- New directions for practical applications

Model weights here <https://www.kaggle.com/ilikehaskell/ro-en>  
Inference script <https://www.kaggle.com/ilikehaskell/ro-fa-inference>