

# NEURAL TRANSLATION FOR EUROPEAN UNION (NTEU) PROJECT

M. García-Martínez\*, M. Herranz\*, A. Estela\*, M. Melero‡, L. Bié\*, A. Cerdà-i-Cucó\*, H. Degroote\*, A. Kohan\*, T. O'Dowd‡, S. O'Gorman‡, M. Pinnist†, R. Rozist†, R. Superbo‡, A. Vasiļevskist†  
Pangeanic\*, Barcelona Supercomputing Cencer‡, KantanMT‡ and Tilde†



## NTEU Project Description

- NTEU will provide a capacity service to eTranslation by building a near-human-professional-quality neural engine farm which includes all the 24 EU language combinations (around 550 translation engines in total), without the necessity to use a high-resourced language as a pivot.

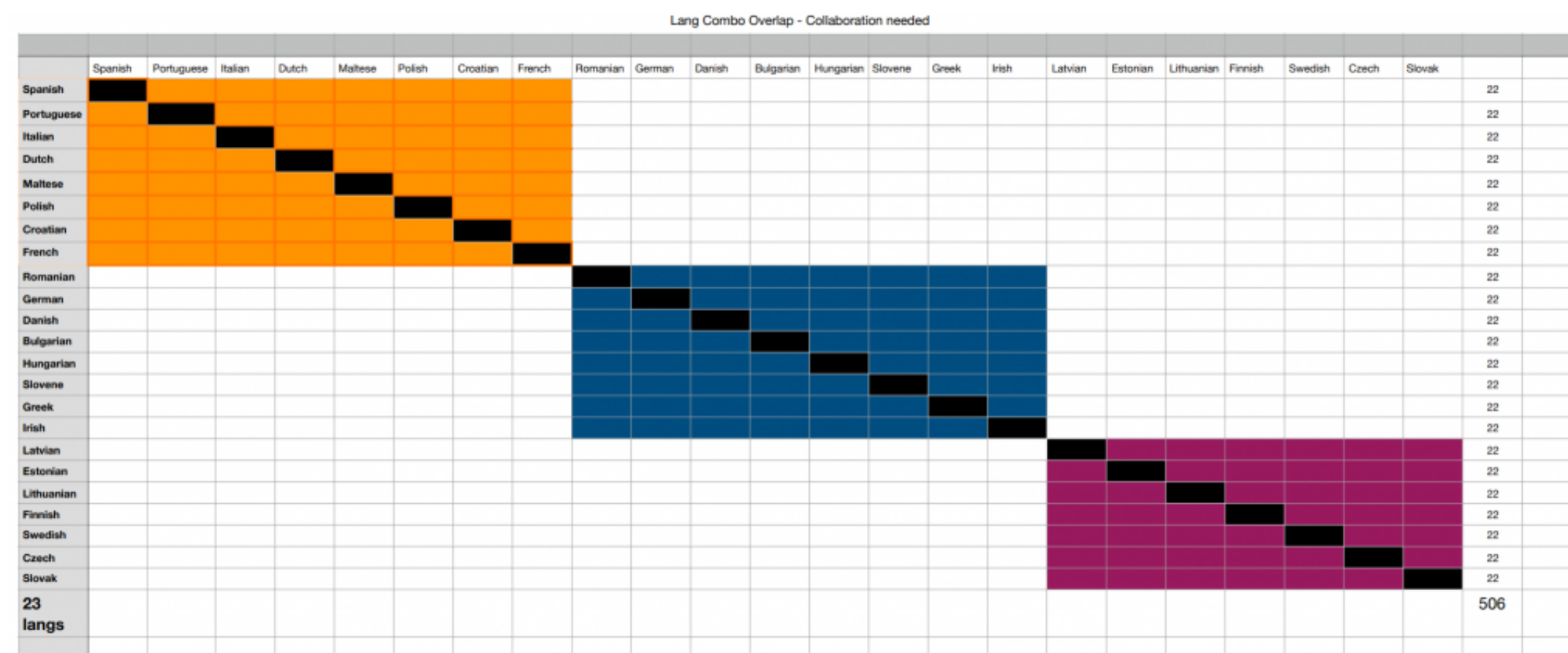


Fig. 1: NTEU language combinations.

- Lower-resourced languages will be a challenge, and more effort will be required to obtain well-performing engines for them.
- Fair automatic and human evaluation.
- Deposit dockerised engines and collected data for easy use by PA's under ELRC repository.
- NTEU started in September 2019 and will run until August 2021.

## NMT models

- State-of-the-art NMT models.
- Generation of synthetic data by doing back-translation
- Transfer learning.
- Pre-trained models such as BERT.

## Data collection

- Clean training data collection from all language combinations so that the engines can be retrained with other future technologies:
  - 15M segments 1-1 resourced langs
  - 10-12M segments under-resourced langs
  - 10M ultra-under-resourced (Maltese, Irish)
- ELRC-SHARE repository: [elrc-share.eu](http://elrc-share.eu)

## Evaluation

- DGT test set excluded from training.
- Automatic evaluation BLEU (around 40-50), TER (around 40-60).
- Human evaluation:
  - Direct Assignment following WMT.
  - Open-source platform to score (0-100) and compare translations to other systems. <https://mtet.pangeamt.com>

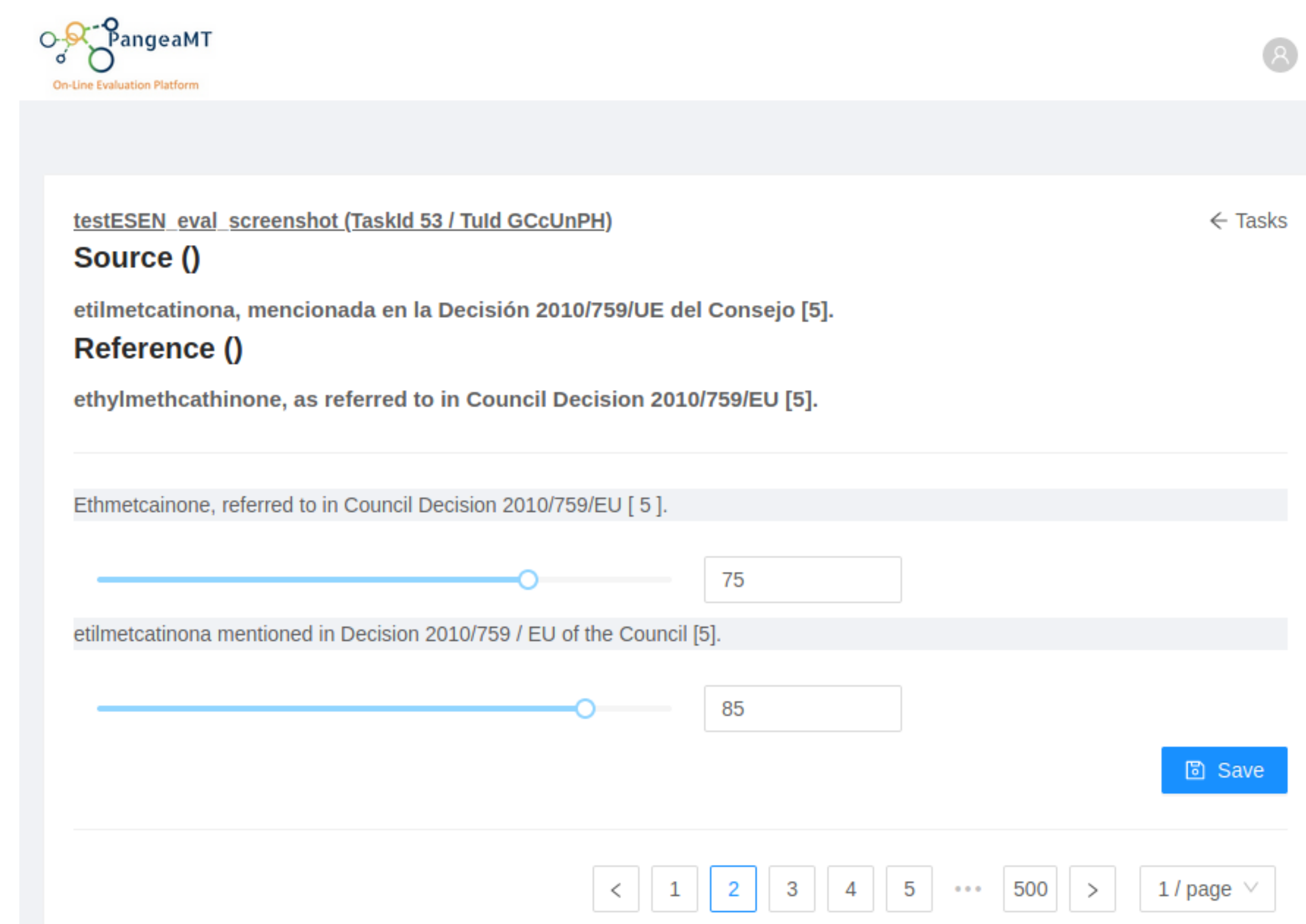


Fig. 2: MTET evaluation tool.

- Feedback will be used to improve translations.
- SEDIA as supervisor (AI secretary of the ministry of Spain).

## Benefits to the community

- Facilitating a Single Digital Market
- Flexibility of Machine Translation across member states.
- Significant Savings in Translation Costs
- Provision of highly secure translation service

## Website

<https://nteu.eu>

## Partners



## Acknowledgements

The work reported in this paper was conducted during the NTEU project, which was funded by Innovation and Network Executive Agency (INEA) through grant N o INEA/CEF/ICT/A2018/1816500 as part of the EU's CEF Telecommunications Program.

