

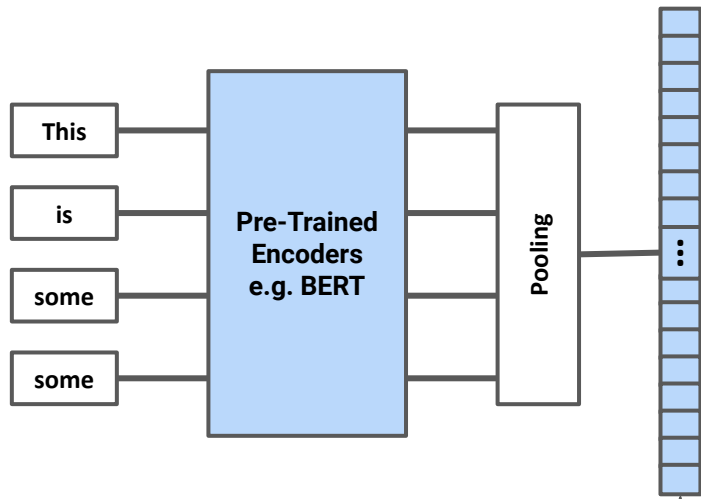
Focus Webinar

Release of TAUS QE v2.0

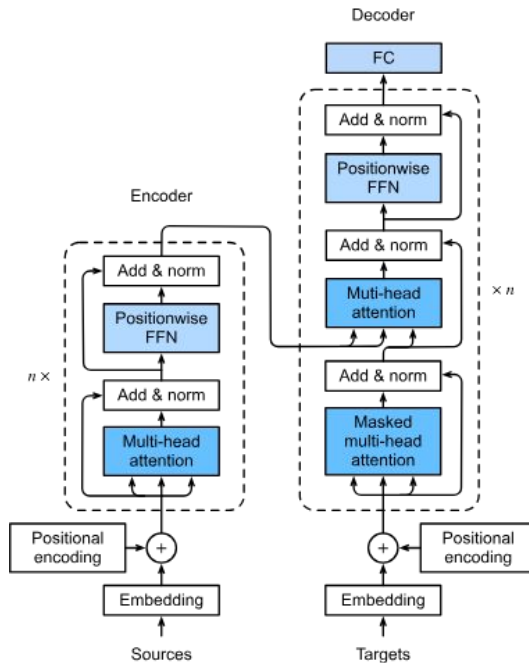
28 March 2024

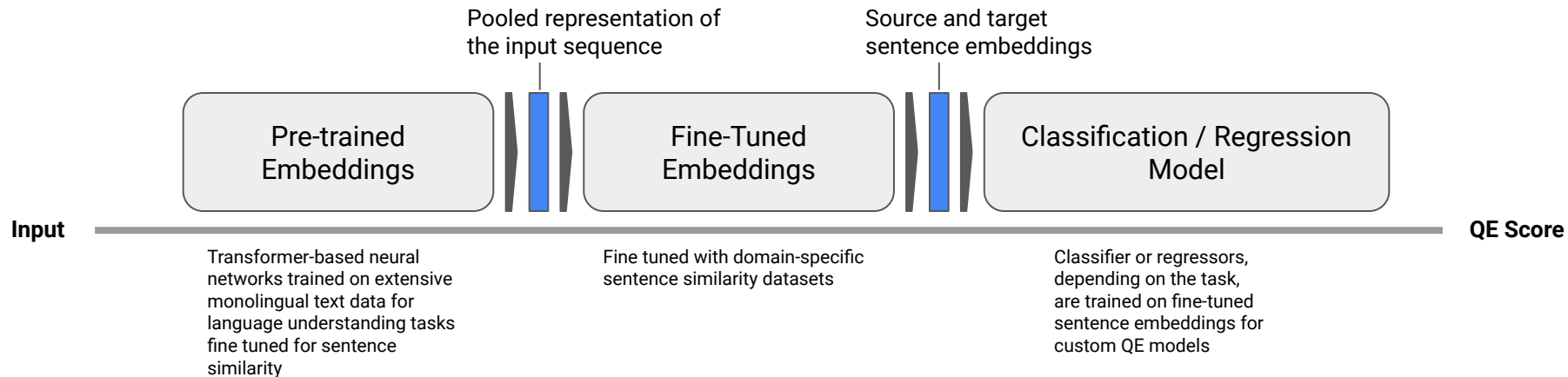
- What's new in TAUS QE v.2.0
- New subscription model
- Q&A

Release of TAUS QE v2.0



Mathematical representation of text such that similar meaning have a similar representation, known as **Embeddings**

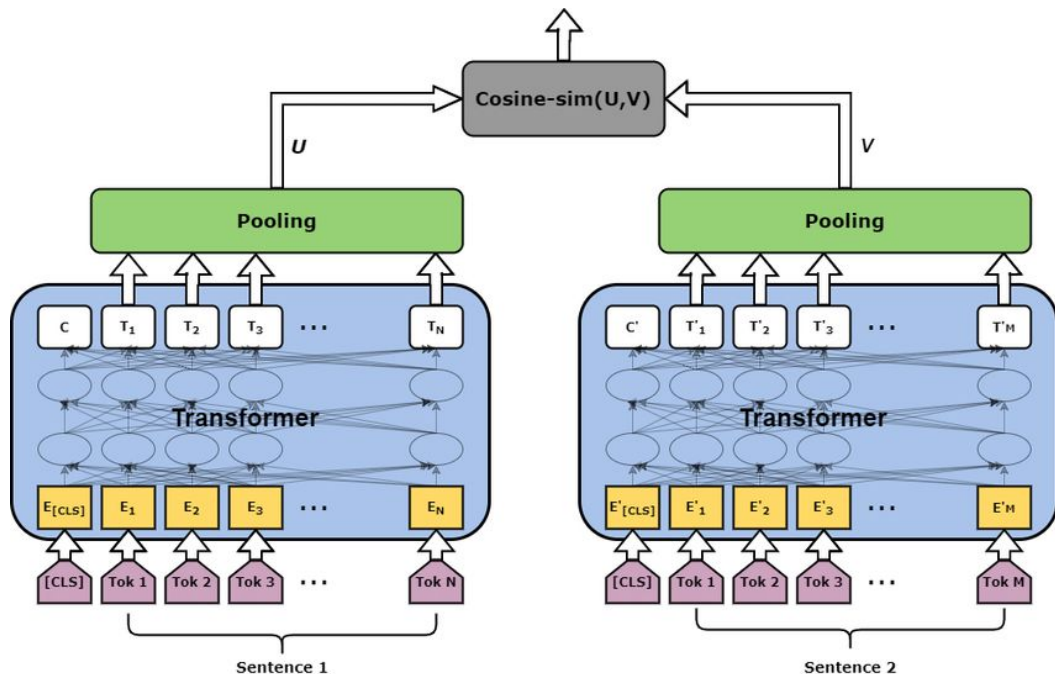




Pre-trained language models can be fine-tuned for tasks related to sentence similarity, and additional classifiers can be incorporated to predict translation quality. This approach leverages the capabilities of pre-trained models to enhance the performance of translation quality assessment models

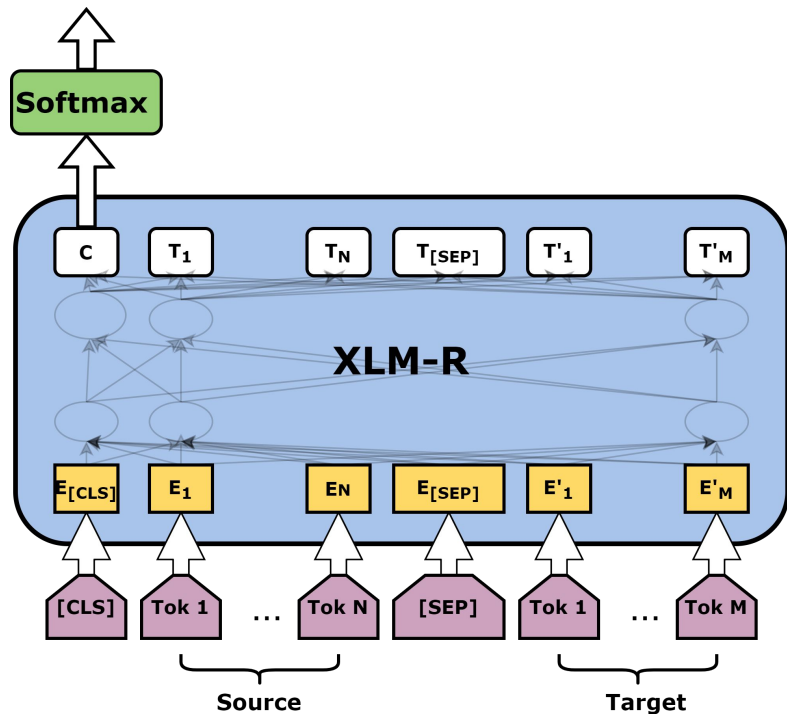
Based on the customization of **pre-trained BERT** models, it's important to note that BERT pre-training is usually conducted on monolingual data, potentially limiting its inherent generalization to languages not explicitly included in the training set.

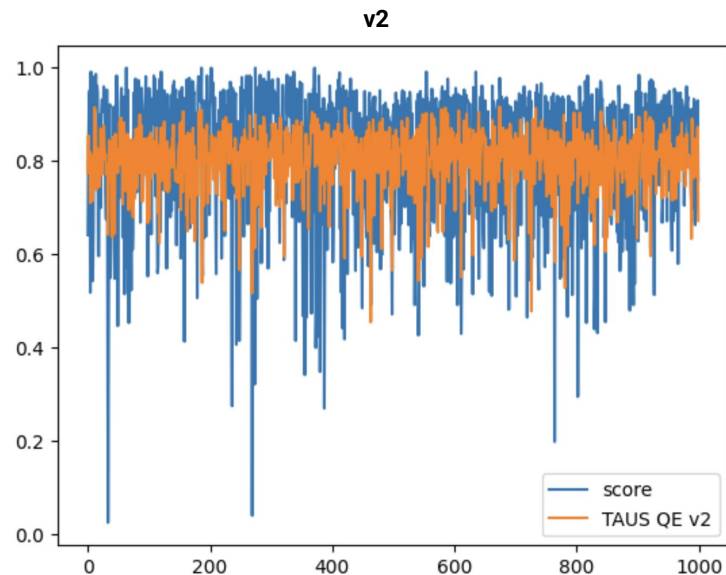
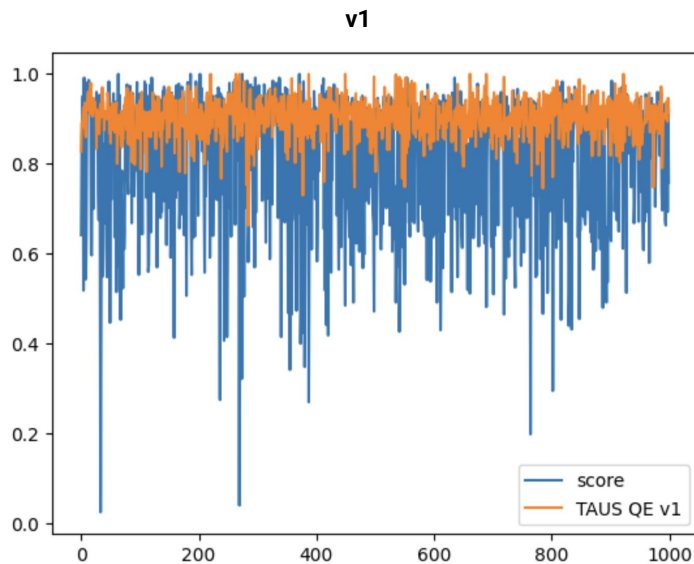
The utilized architecture is the **Bi-Encoder**, wherein each sentence is independently encoded, followed by the calculation of similarity.



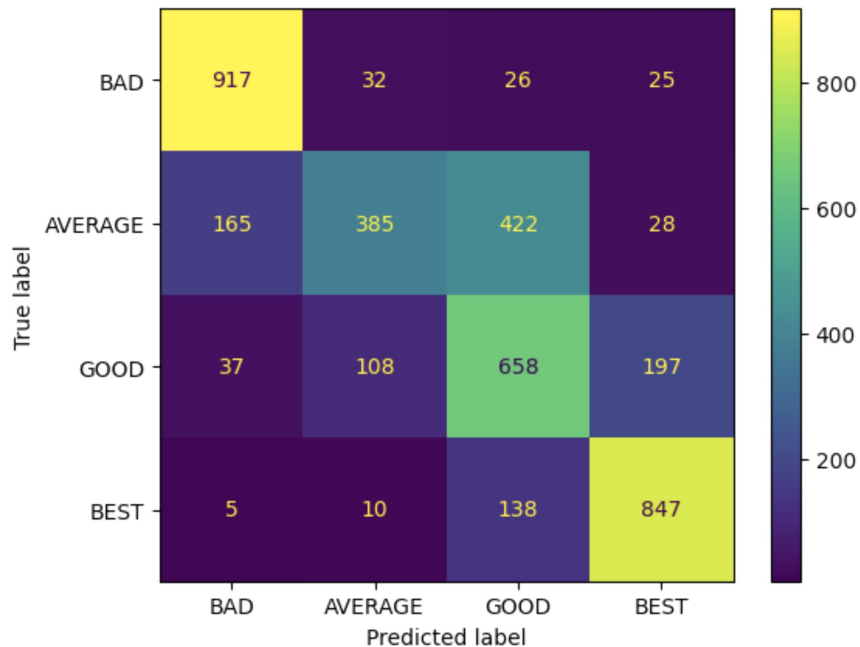
Based on the customization of **pre-trained XML-R** models, which is explicitly designed for cross-lingual tasks. It aims to learn universal language representations that can be effective across multiple languages, making it particularly useful for multilingual applications.

The utilized architecture is the **Cross-Encoder**, wherein both sentence concatenated are encoded using the same model and then a classification head is added on top.

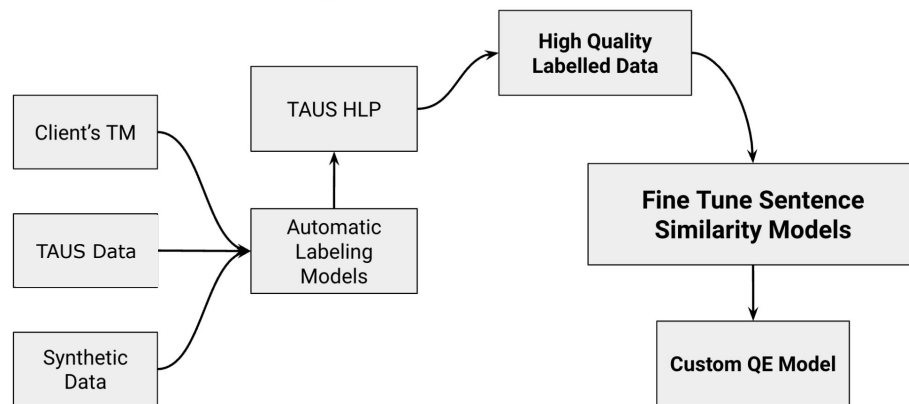




Example of English-German Test Data



Error Rate | BEST 15% | BAD 8%



- Indicate the metric version in the API call.
- If the metric specification is omitted in your request, the QE v1.0.0 will be returned by default.
- You can also specify both versions in the request to compare the results.
- If a custom-trained version is available to your account for the requested language pair, then by default the custom version will be returned.

```
import requests
import json

url = "https://api.sandbox.taus.net/1.0/estimate"

payload = json.dumps({
    "source": {
        "value": "This is a test.",
        "language": "en"
    },
    "targets": [
        {
            "value": "Dat ist ein Test.",
            "language": "de"
        }
    ],
    "metrics": [
        {
            "uid": "taus_qe",
            "version": "2.0.0"
        }
    ]
})

headers = {
    'Content-Type': 'application/json',
    'Api-Key': 'YOUR-API-KEY'
}

response = requests.request("POST", url, headers=headers, data=payload)

print(response.text)
```

Pricing: volume- or subscription-based

- Pay for what you use
- Recommended for relatively low volumes of up to 20M characters per year
- Base price: €200 per 1M characters
- Custom model training available at extra charge

Bundle	Price
Free Trial - 500K characters	€0
2 million characters	€400
5 million characters	€1,000
20 million characters	€4,000
50 million characters	€10,000
100 million characters	€16,000

Includes:

- Up to 100 million characters per language pair per month
- Training and maintenance of customer-specific QE models
- NLP retainer: on-demand development of custom NLP features, such as error auto-correction, error categorization, automatic PE, smart translation routing, etc.
- Annual commitment, from €21,600 per language pair per year

For more information, please visit

<https://www.taus.net/data-solutions/quality-estimation>

See API specifications <https://developer.taus.net/>

Or contact us at sales@taus.net
