

Custom-trained vs. Generic: a Comparative Study of Post-editing Performance

Alexander Gigga
CMO lengoo

#LTI19





Who am I?



Name:

Alexander Gigga

Location:

Berlin

**Professional
Background:**

2011 - B.Sc. Industrial Engineering, KIT

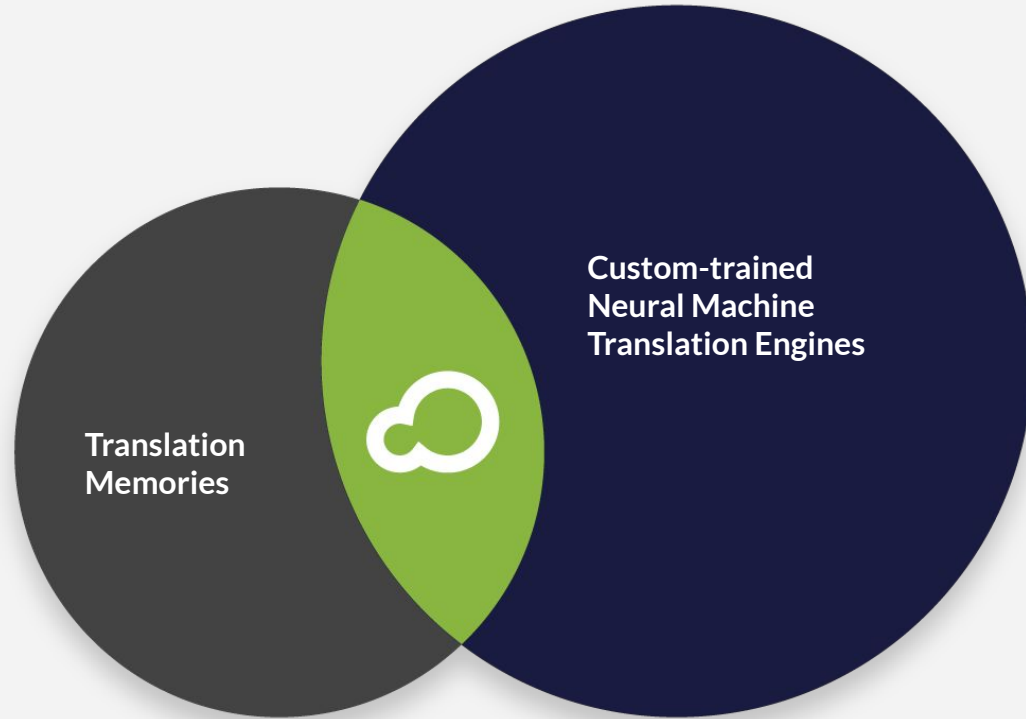
2016 - M.Sc. Industrial Engineering, KIT

2010 - Adidas, Human Resources

2013 - Co-Founder & CMO bei lengoo GmbH



What do we do?



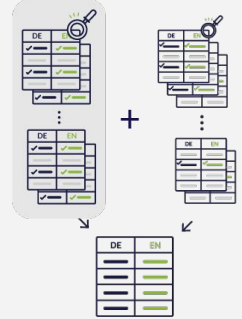
Translation
Memories

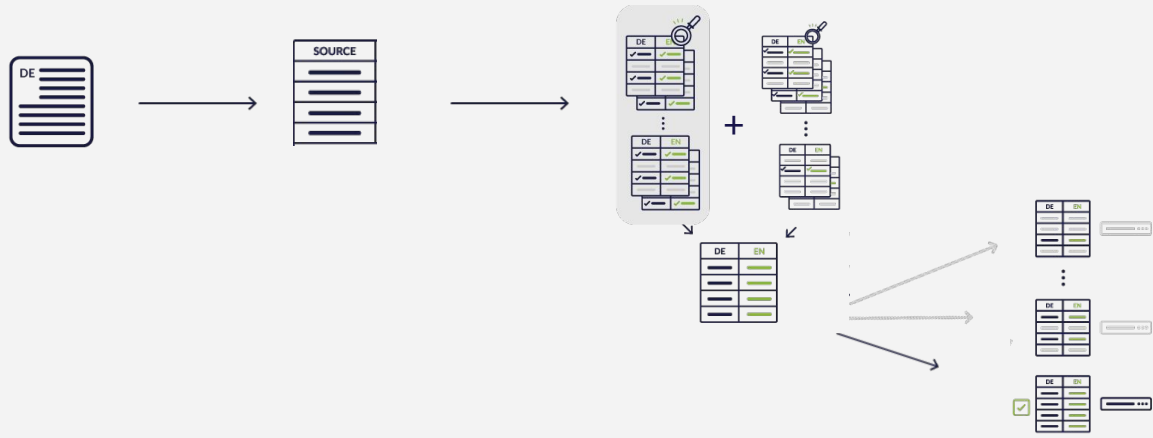
Custom-trained
Neural Machine
Translation Engines

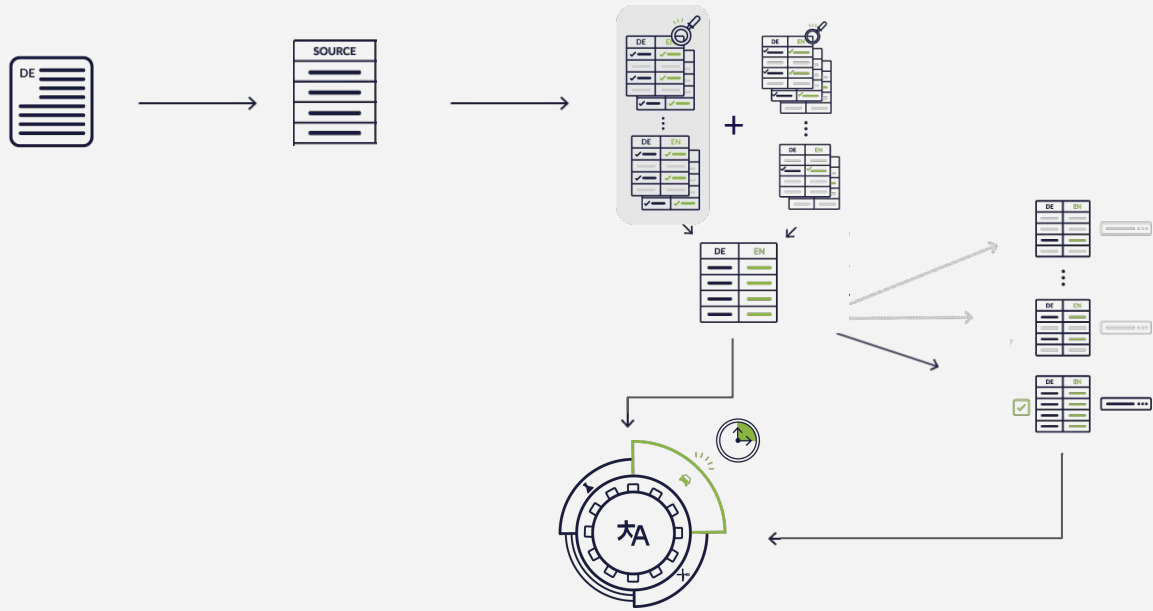


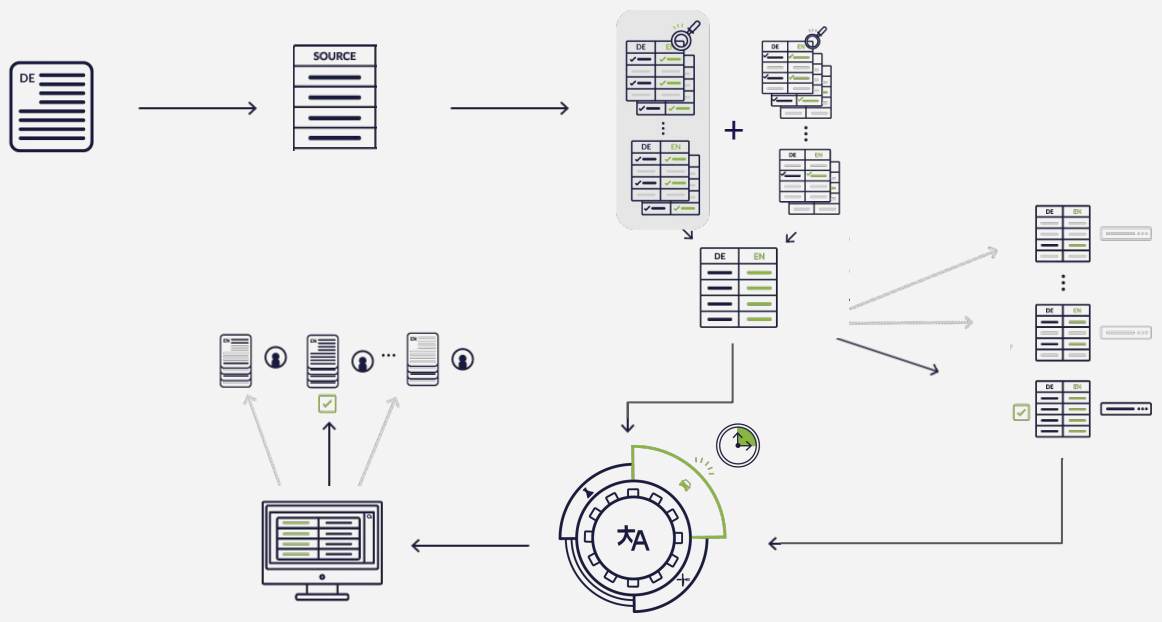
Training Custom NMT Engines

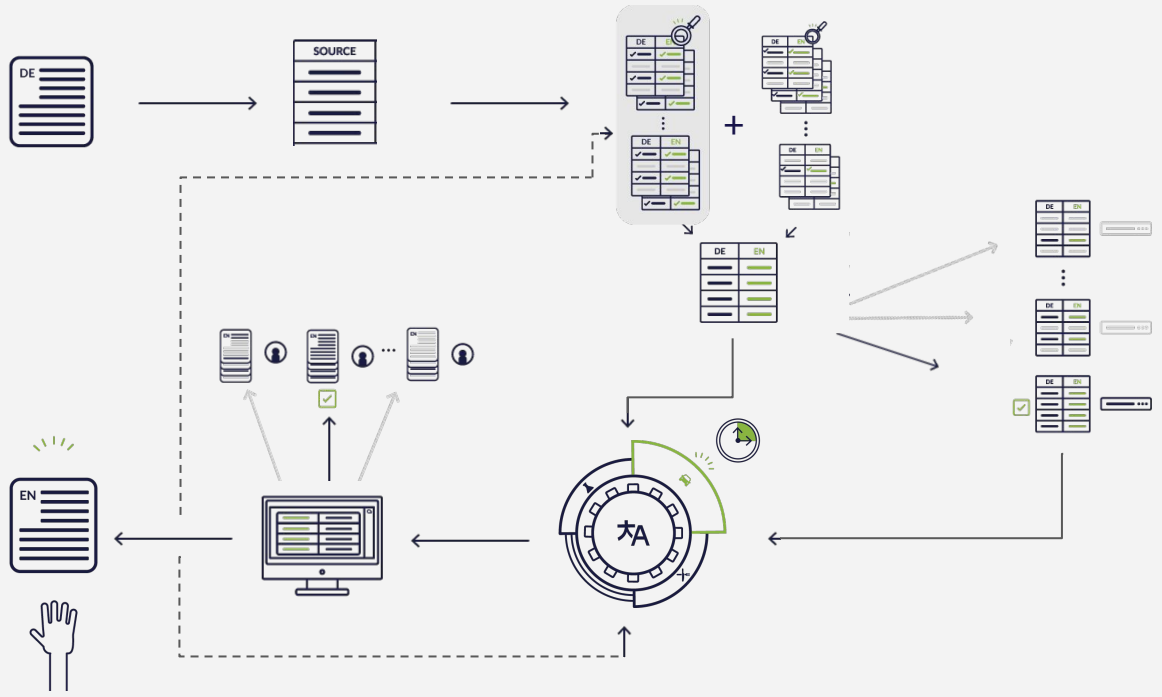














Hyper Customization



Training Automation



Measuring the Quality of Custom-Trained NMT



BLEU Score

Relative to the best generic NMT model available

Post-Editing Time

Time needed to post-edit 1,000 words

Quality Review

Syntax, Spelling, Usage of in-domain language



Case 1

Vacation Rental Descriptions



Client A | Vacation Rental Descriptions

Client is a vacation home rental platform.

Market leader for premium house rentals in Mallorca.

Case Details

Language Pair

DE → EN (UK)

Content

Rental home descriptions

Training & deployment of custom NMT

Very low amount of data available for the initial training of a custom NMT V1.

The industry standard “BLEU” measures the difference of a machine translation to the human reference translation.

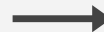
Operational results (based on 20k words)

Without custom NMT

Processing Time

14 days

-65%

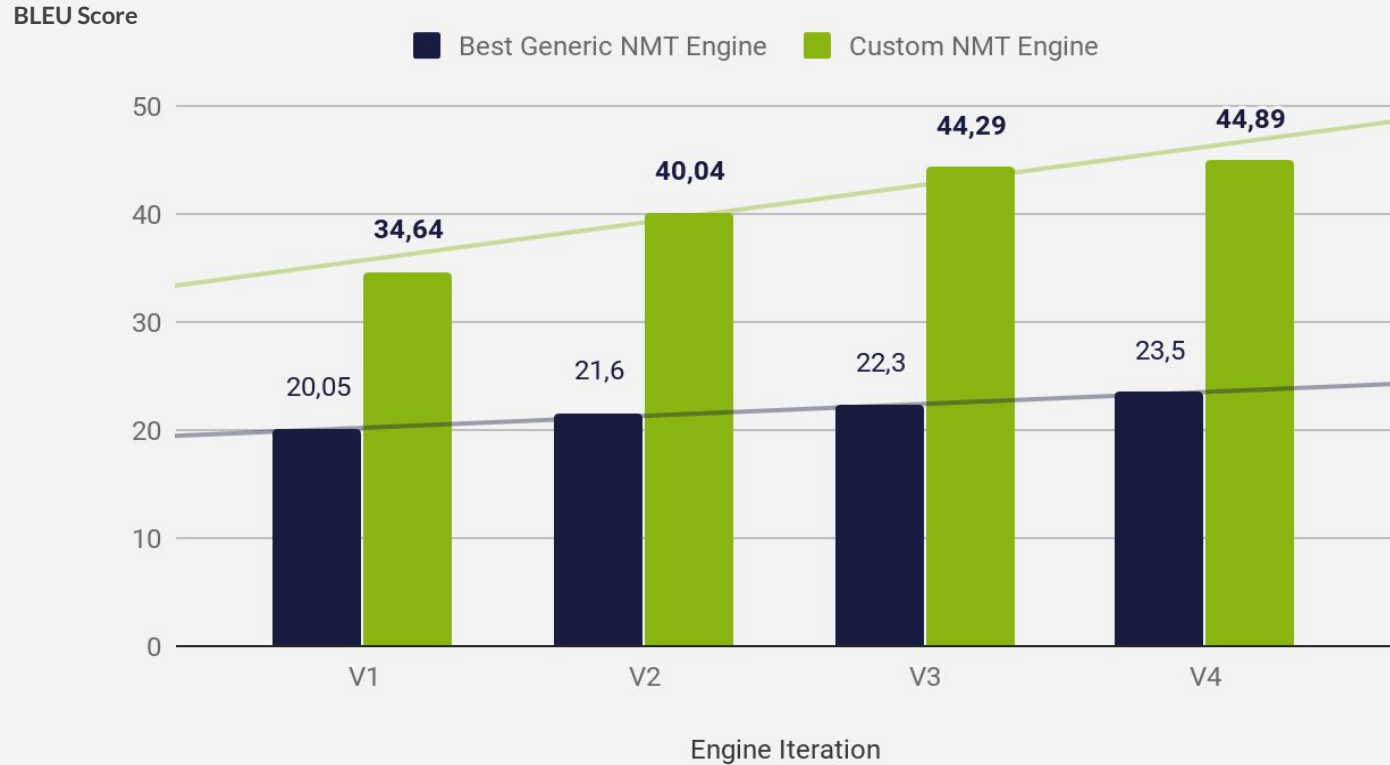


With custom NMT

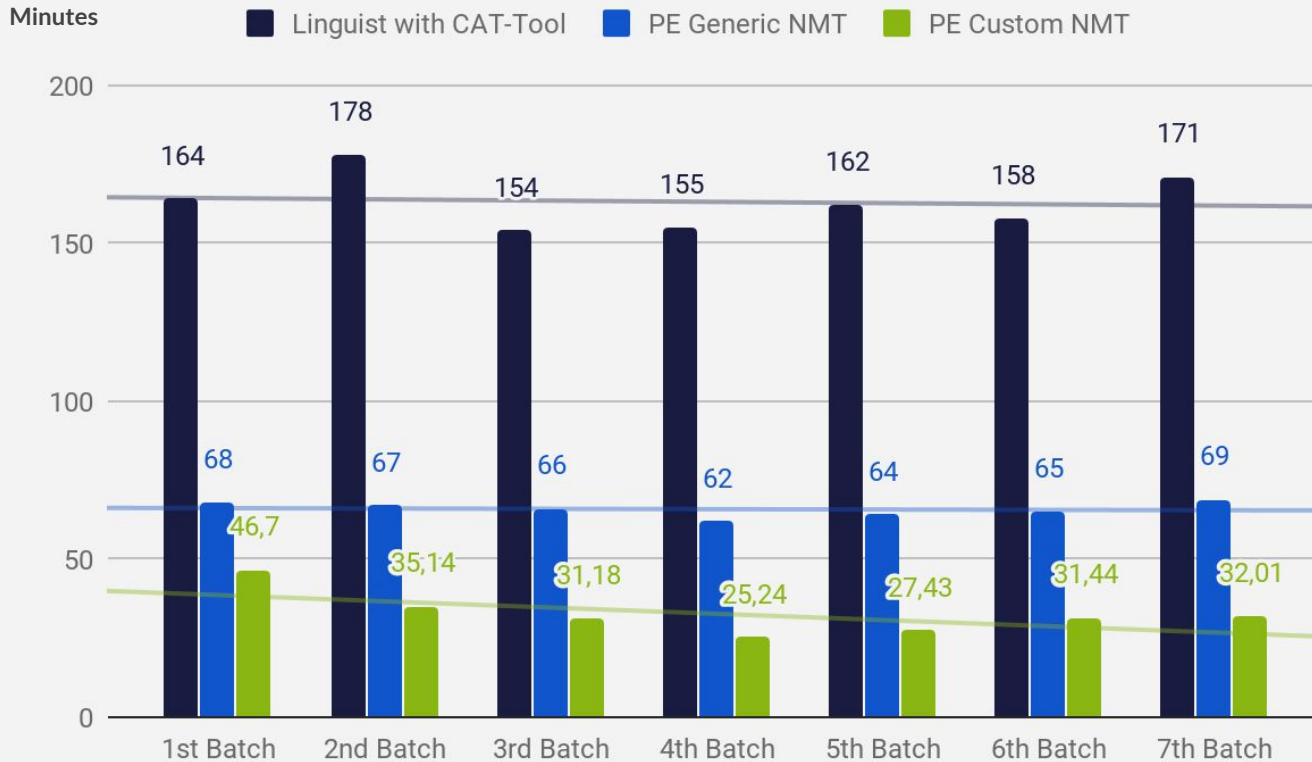
Processing Time

6 days

Case 1 | Comparison of BLEU Scores



Case 1 | Comparison of Processing Times





Case 2

Instructions for Use (Pharmaceuticals Company)



Client B | Pharmaceuticals Company

Client is a globally active pharmaceuticals producer from Switzerland.

Highly specialized use of language and intense use of tags.

Case Details

Language Pair
Content

EN → DE
Instructions for Use

Training & deployment of custom NMT

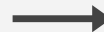
Very high amount of data available for initial training of V1.

Operational results (based on 20k words)

Without custom NMT

Processing Time
19 days

-70%



With custom NMT

Processing Time
6 days

Translation Examples B1



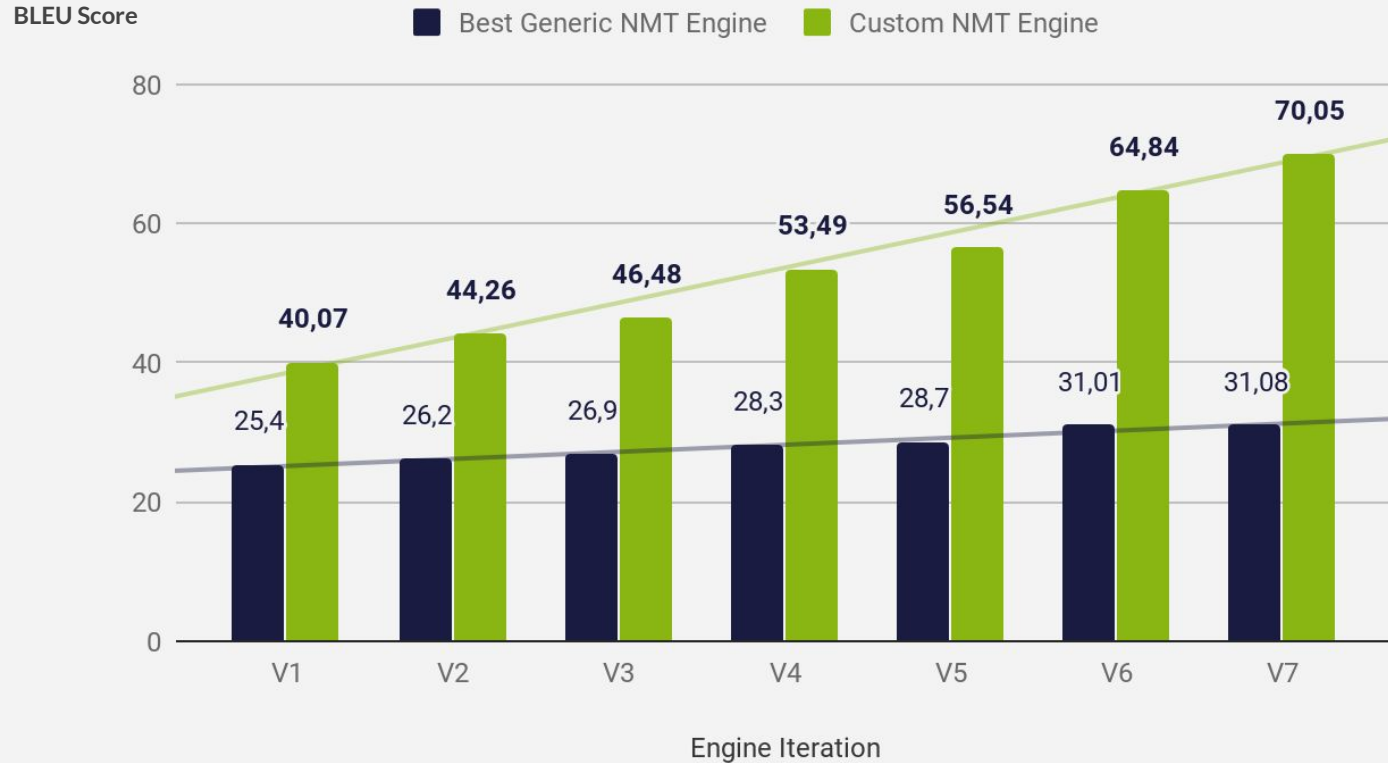
Source Gently swirl SR and allow vial to stand closed for a minimum of 10 minutes at 15<nbhy/>25 °C. Swirl the vial carefully to produce a homogeneous solution before use.

83% TM-Match R1 leicht schwenken und Flasche zur Rekonstitution verschlossen bei 15<nbhy/>25 °C 10 Minuten stehen lassen. Vor Gebrauch die Flasche vorsichtig schwenken, um eine homogene Lösung zu erhalten.

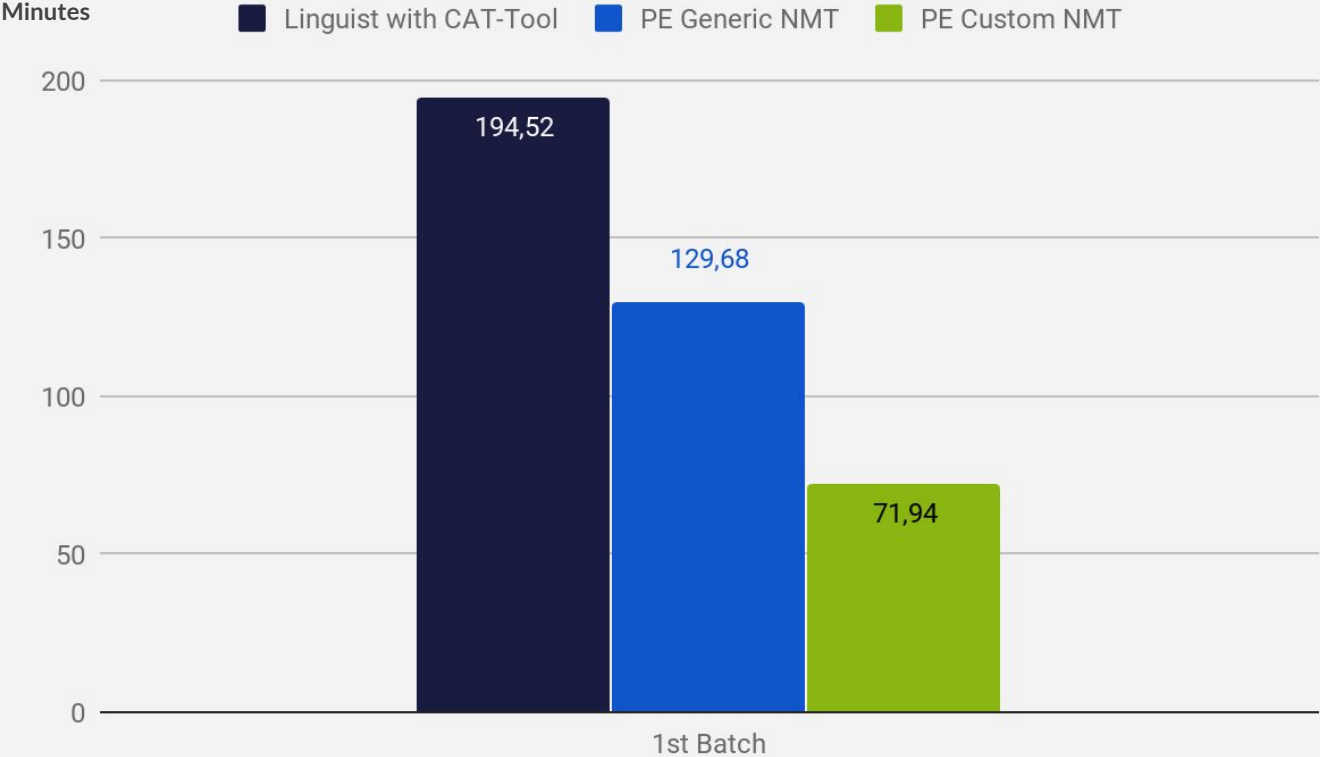
Best Generic Engine SR leicht **verwirbeln** und das **Fläschchen** mindestens 10 Minuten lang bei 15**Non Breaking Hyphen Tag Text**25 °C verschlossen stehen lassen. **Das Fläschchen vor der Anwendung** vorsichtig **umrühren**, um eine homogene Lösung zu erhalten.

lengoo NMT SR leicht **schwenken** und **Flasche** verschlossen für mindestens 10 Minuten bei 15<nbhy/>25 °C stehen lassen. **Vor Gebrauch die Flasche** vorsichtig **schwenken**, um eine homogene Lösung zu erhalten.

Case 2 | Comparison of BLEU Scores



Case 2 | Comparison of Processing Times





Learnings



Custom NMT does not equal custom NMT

Iterations of engines with a sensitive focus on data selection yields remarkable quality improvements over time.

Custom NMT is a viable solution

The technology has proven viable across various use cases and industries.

Selection is key

Focussing on the selection of data for training purposes is paramount to achieving better results.



Automated Hyper Customization

Are you already customizing NMT?

Alexander Gigga
Co-Founder and CMO
lengoo

+49.170.1244.179
alexander.gigga@lengoo.de

#LT119