



EMBRACE THE FUTURE

PEMT Productivity with domain-adapted NMT

*A methodology to evaluate the contribution of
DIY vs customized, commercial engines*

Markus Welsch

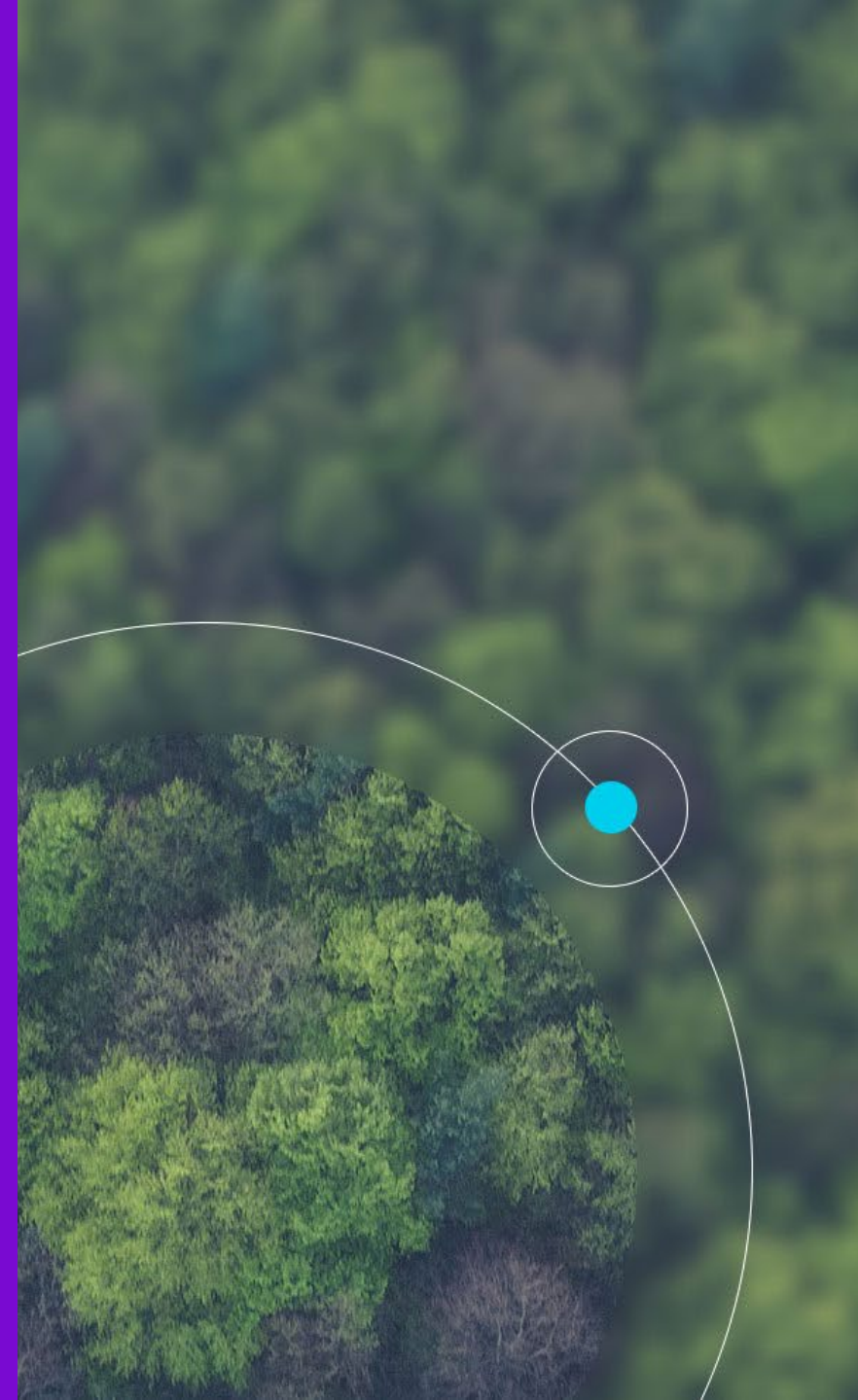
Vice President Content Intelligence

26 November 2019



AMPLEXOR in a Nutshell

We help customers manage their content and customer touchpoints to improve efficiency, increase revenue, reduce time to market and ensure quality and compliance.



Contact Data



**Markus Welsch***Vice President Content Intelligence*

AMPLEXOR International S.A.

55, rue de Luxembourg

8077 Bertrange

Luxembourg

 +352 314 411 420 markus.welsch@amplexor.com

About

Markus is Vice President Content Intelligence at AMPLEXOR and based in Luxembourg. During his more than 20 years within the AMPLEXOR group, he contributed in different roles and positions to the **design, architecture, implementation and operation** of numerous challenging, **multi-lingual Content and Information Management solutions** for customers in different industries.

Encouraged by his Computer Sciences background and a corresponding mindset, his passion and special attention always was - and continues being - around the **smart automation of content- and information-centric business processes and related cognitive activities**.

In his current position, Markus is responsible for managing a **comprehensive portfolio of smart solutions** that bring together the best of both worlds, combining the speed, scale and power of machines with a human-like approach to take advantage of information on a scale that would otherwise be impossible for people. Within their area of application, these solutions can understand language, recognize valuable patterns and relationships, learn from data and information and allow answering questions that would have seemed unimaginable only a few years ago.

AMPLEXOR

- | | |
|---------------------|------------------------|
| Augsburg | Lisbon |
| Bagnols-sur-Cèze | Lorient |
| Berlin | Lyon |
| Bertrange | Madrid |
| Brest | Montigny-le-Bretonneux |
| Broomfield (CO) | Montréal |
| Bucharest | Nantes |
| Budapest | Newton (MA) |
| Buenos Aires | Novo mesto |
| Cambridge | Paris |
| Chennai | Rīga |
| Cherbourg Octeville | River Falls (WI) |
| Cluj-Napoca | Shanghai |
| Dublin | Singapore |
| Düsseldorf | Sibiu |
| Eindhoven | Suzhou |
| Gent | Toulouse |
| Kraków | Vitoria-Gasteiz |
| Kreuzlingen | Zagreb |
| Leuven | Zürich |
| Kobe | |

Locations

41

OFFICES

23

COUNTRIES

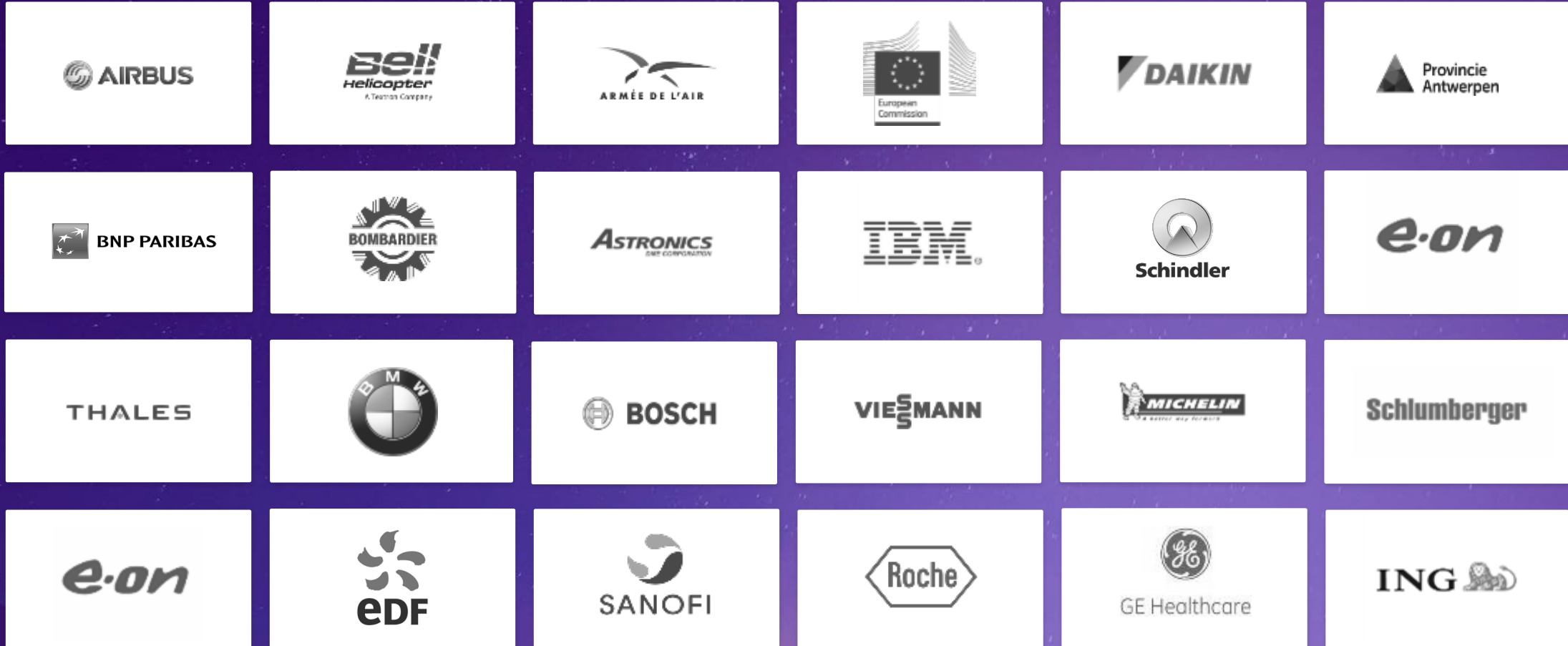
1.850

EMPLOYEES

Top
11

LSPs
WORLDWIDE

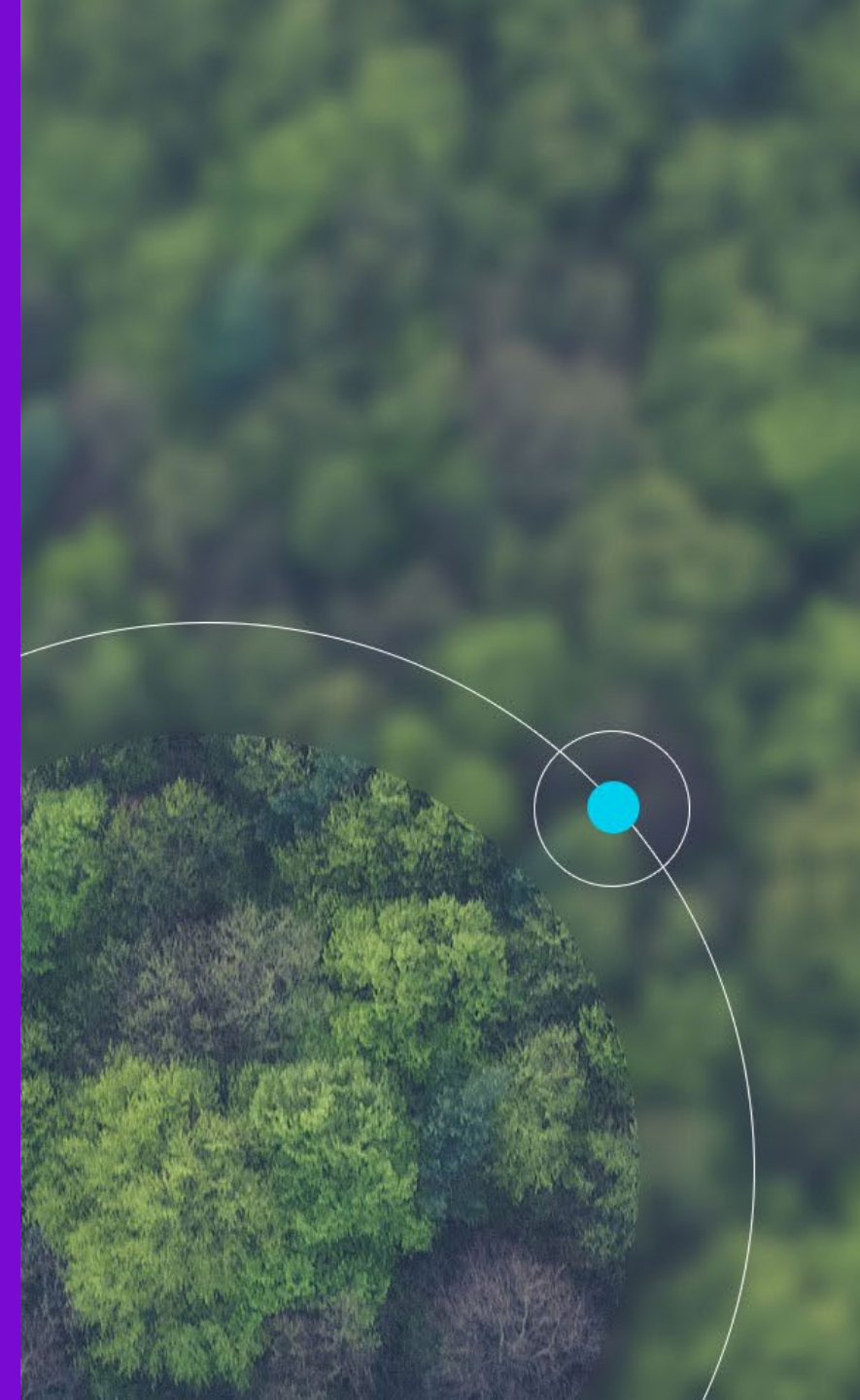

AMPLEXOR is a leading Digital Solution Provider – Globally present to support our Customers' core Business





Business Context

Neural Machine Translation
in an LSP context





Market Drivers

- Digital Transformation, AI & Machine Learning, Customer Experience, Big Data, IoT, etc.
- Time to Market / competitive Environment
- Data volumes growing at exponential rates
- 80+ % of data is unstructured



Enablers

- Algorithmic progress; public availability
- Computational Capacity
- Digitalization, Big Data (bulk of available data)
- Increasing interest and investments in R&D



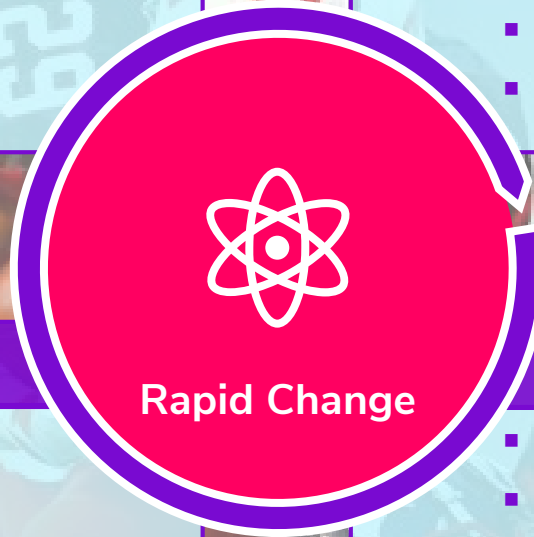
Technology Market

- NMT breakthrough at WMT 2016
- Big Tech becomes the innovation driver
- Ongoing extension of MT solution capabilities
- AI-driven innovations (Augmented Translation)



Language Services Market

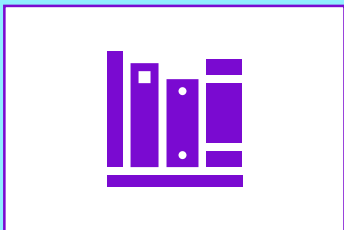
- Ongoing Mergers & Acquisitions
- Startups establish AI-based business models
- Big Tech enters the end-client market
- Continuous Localization
- Translation cost continuously decline





Domain

Medical Device



Content Types

Marketing,
Regulatory,
Technical, Others



Language Pairs

EN → DE
EN → PT
EN → JA



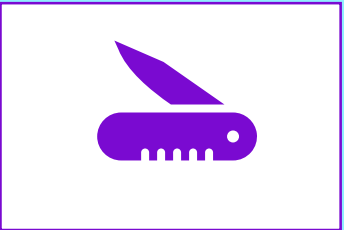
Training Data

10+ Mio
sentence pairs per
language pair



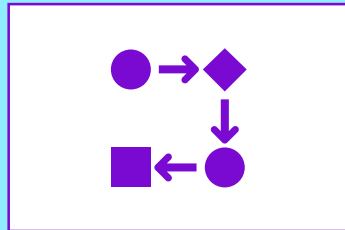
Data Manufacturing

In-domain Data
Out-domain Data
Synthetic Data



Framework

OpenNMT
Open source NMT
ecosystem

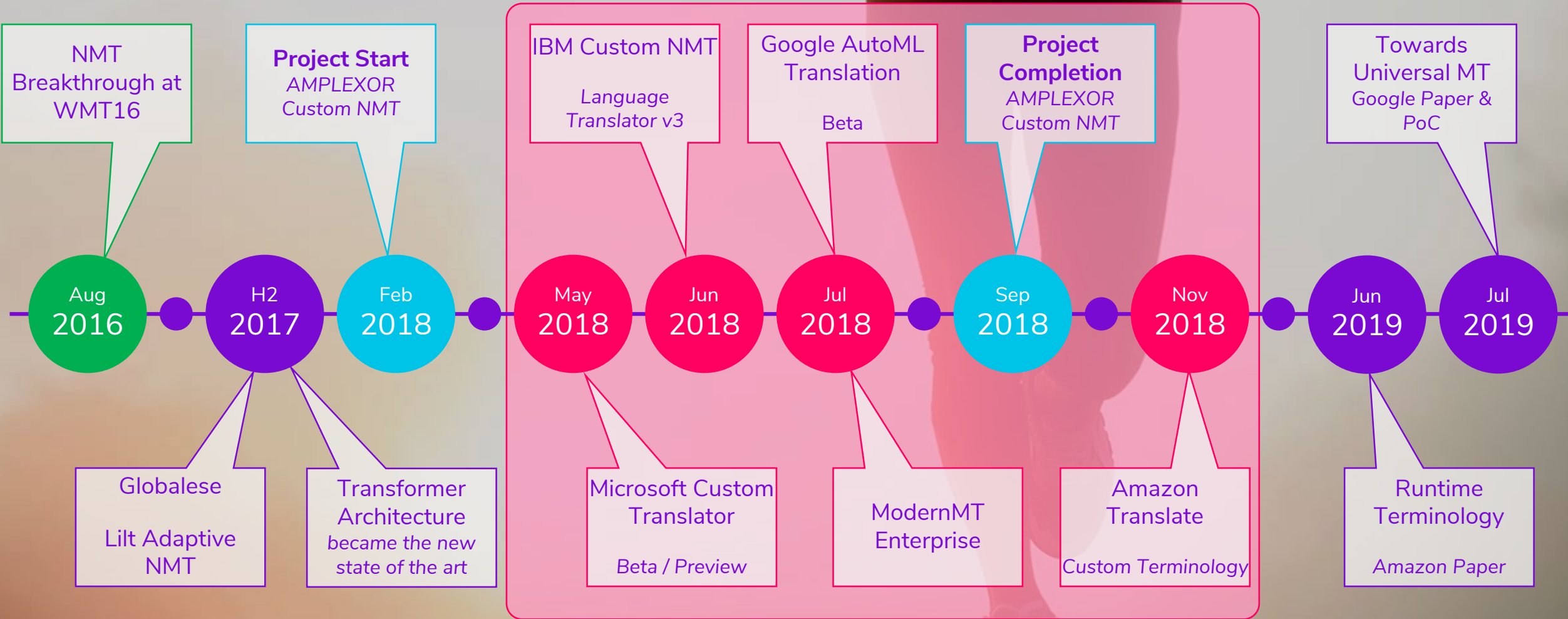


NN Architecture

RNN (Bi-LSTM)
with attention

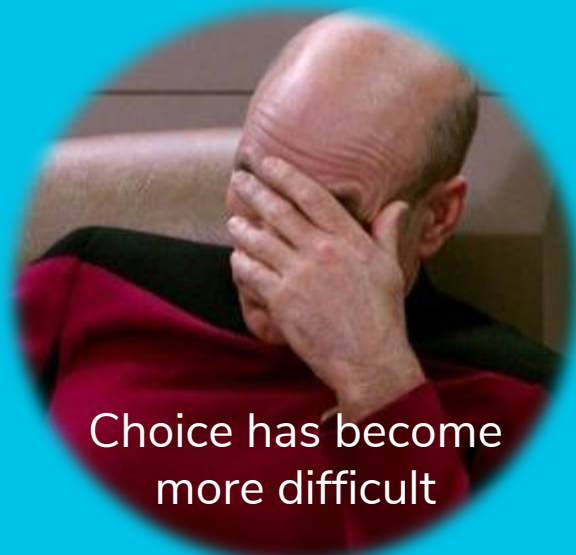
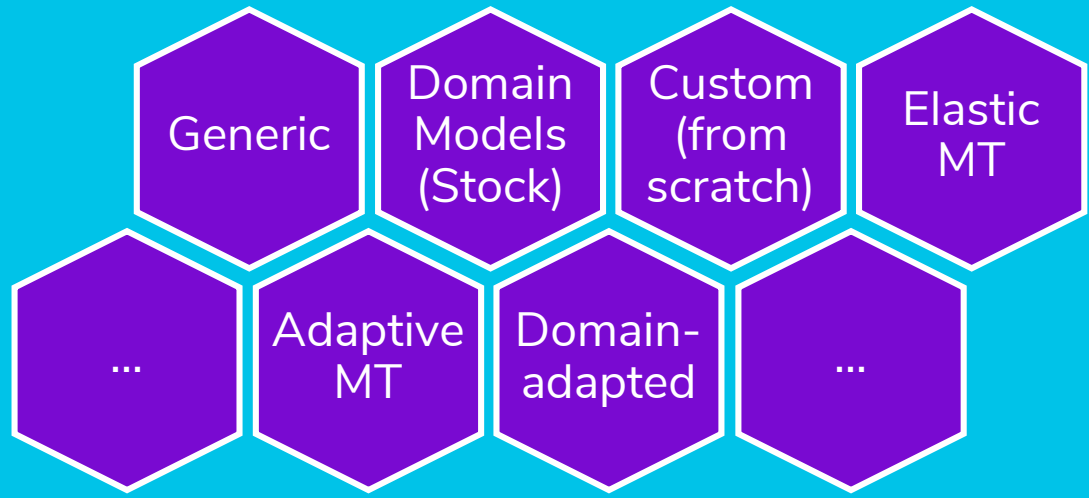


... and the uncontested winner is AMPLEXOR NMT





Types of NMT Engines





Alibaba Cloud
stock



Amazon
stock



Baidu
stock, custom



Kakao
stock



DeepL
stock



Google Cloud
stock, custom



Globalese
custom



eBay
stock



GTCOM
stock



IBM Watson
stock, custom



KantanMT
custom



Naver
stock



Lilt
stock, custom



Microsoft
stock, custom



ModernMT
stock, custom



NIUtrans
stock



Omniscien
custom



PangeaMT
custom



Prompsit
custom



Sogou
stock



PROMT
stock, custom



SAP
stock



SDL
stock, custom



Slate
custom



Systran
stock, custom



Tencent Cloud
stock



Tilde
custom



Yandex
stock



Youdao
stock

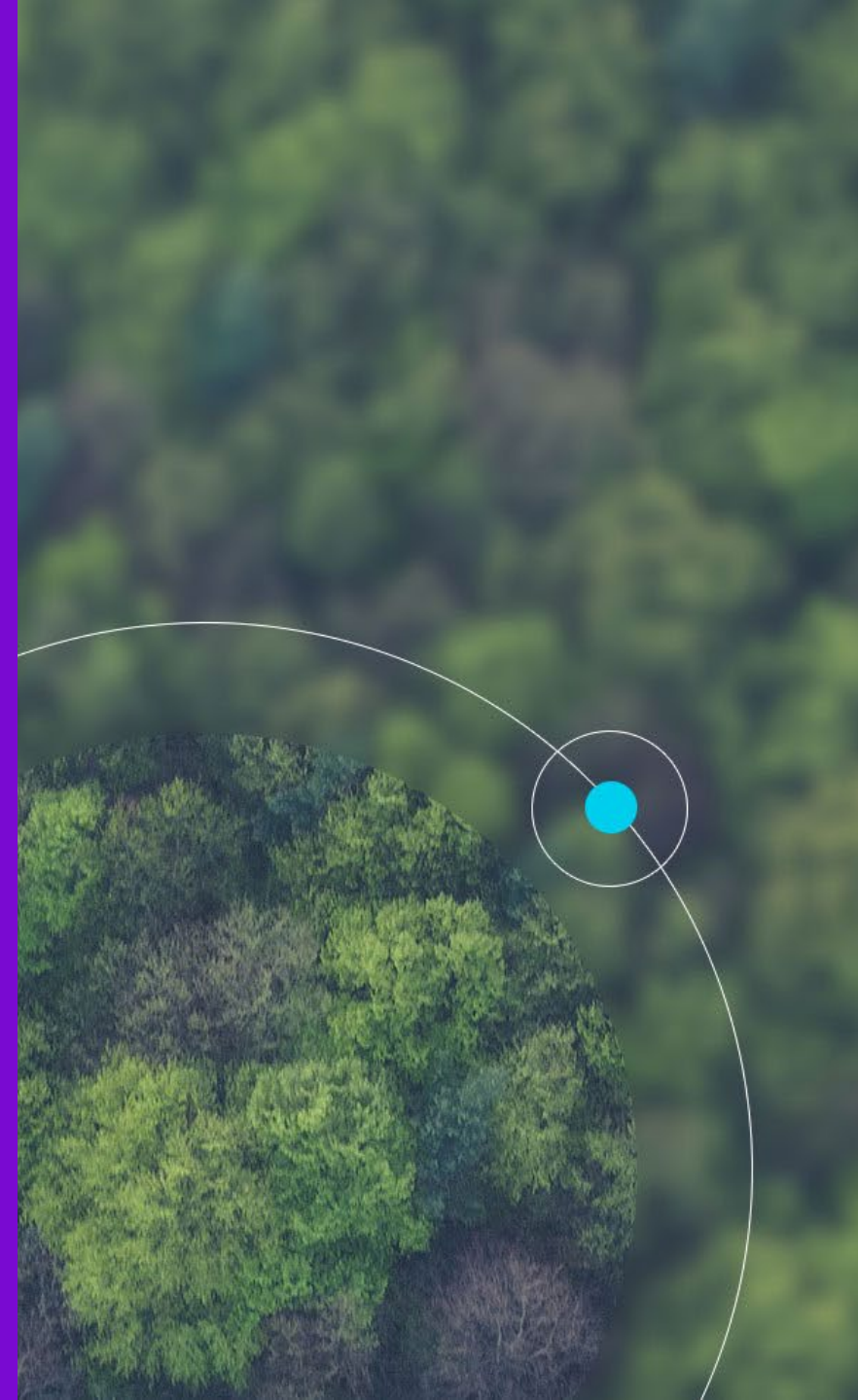


Aspect	DIY	Commercial Mainstream
Approach	native engine from scratch	Transfer Learning
Starting Point	NMT frameworks <i>(Open Source or proprietary)</i>	<ul style="list-style-type: none"> baseline models datasets as a service
Training Data Size	1M – 15+M segments	10k – 1M segments
Training Process	heavily curated	automated
Tuning Process	empirical, interactive	automated (black box)
Setup Costs	€€€€ – €€€€€ <ul style="list-style-type: none"> heavy data manufacturing human involvement computing cost 	€ – €€ <ul style="list-style-type: none"> data manufacturing
Translation Costs	€ – €€	€€€ – €€€€ <i>Pay-As-You-Go</i>
Required Profile	Professional (Data Scientist)	Business User




Productivity Analysis

Approach and Methodology






We need trustable results!



Domain

Medical Device



Evaluation Targets


AMPLEXOR NMT
vs
Google AutoML



Language Pairs


EN → DE

Later: EN → PT, EN → JA



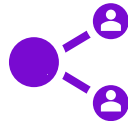
Comparative Evaluation

Contribution to PE
Productivity



Impact of Content Types

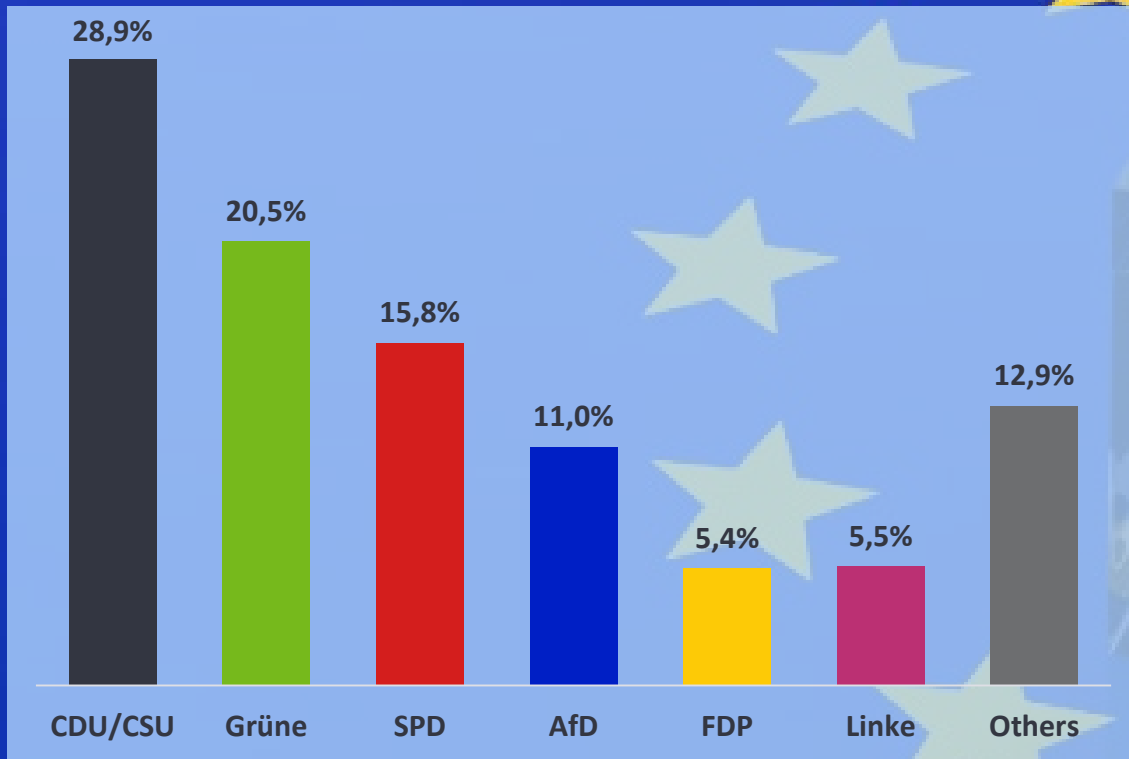
2 Experiments
All medical content types;
Only technical content



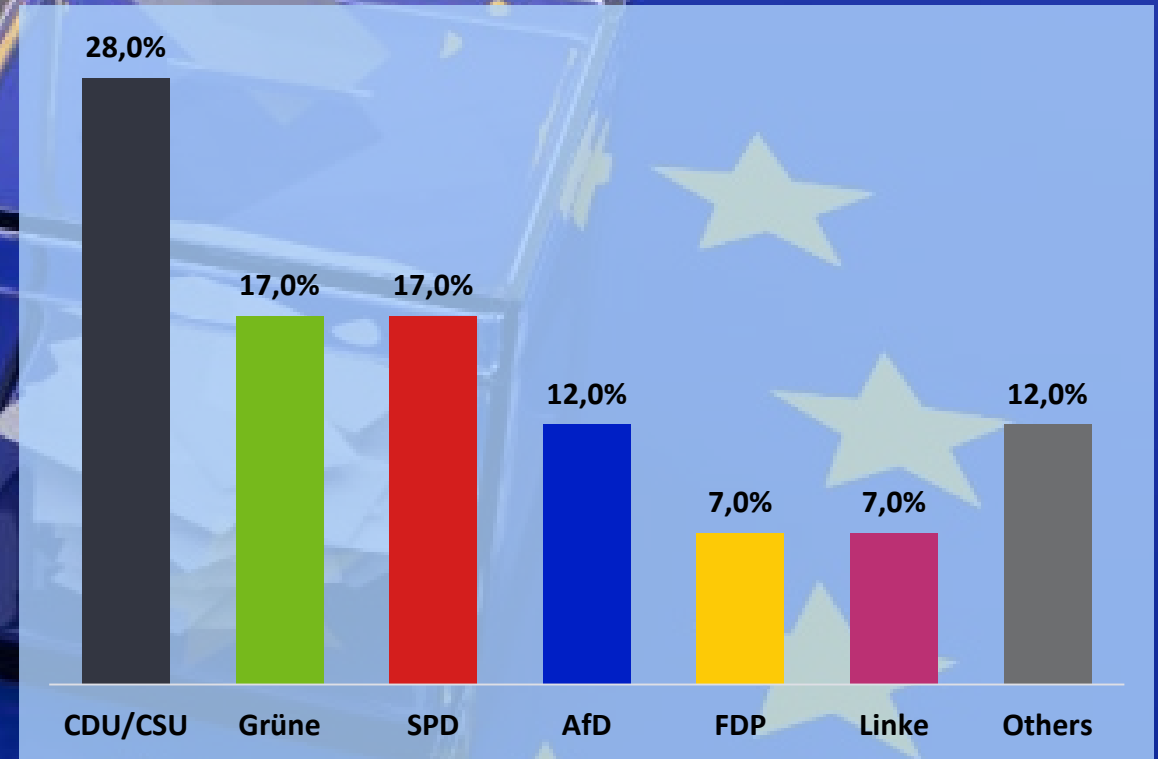
Impact of Post-Editors

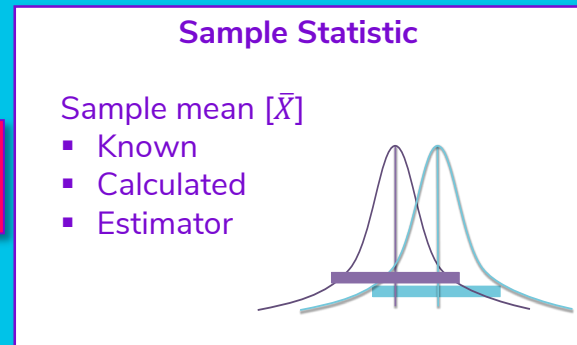
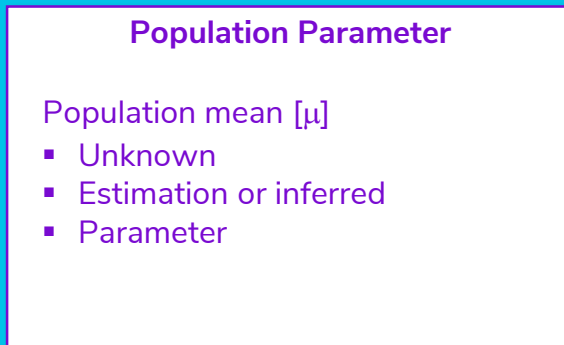
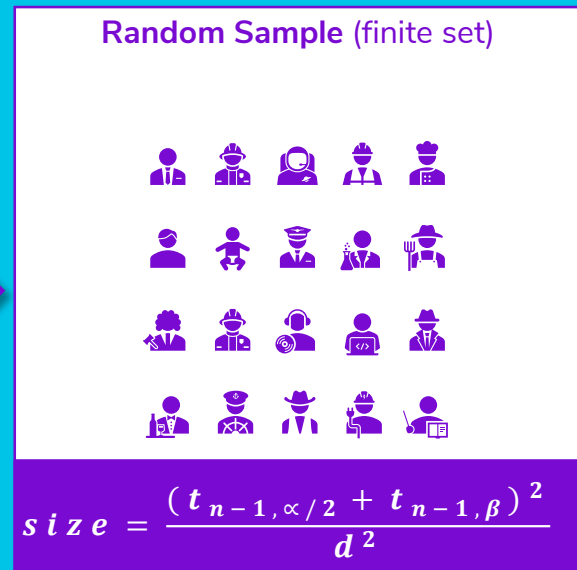
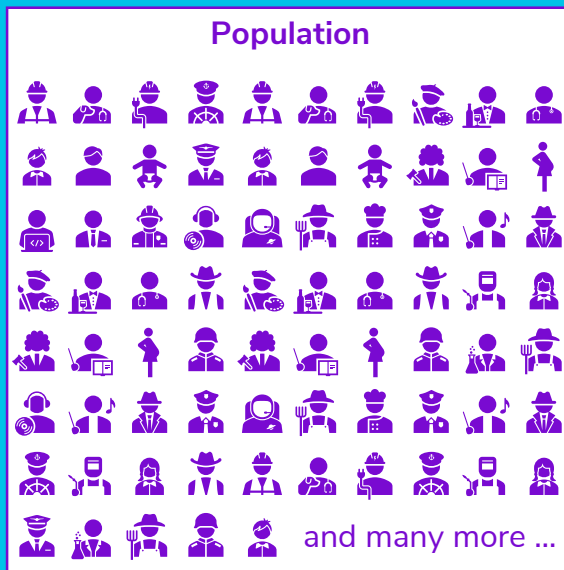
Throughput
Variance

Result on 26.05.2019



Opinion poll on 16.05.2019 by infratest dimap





An “appropriate” sample size is of vital importance to achieve statistical significance

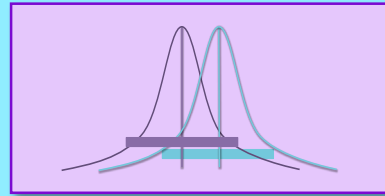
Experiment Phases

Design
Methodology ensuring
Statistical Significance

Execution
Post-Editing

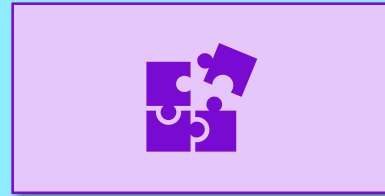
Linguistic QC
Linguistic Assessment
of PE Products

Statistical Analysis
Results / Inference / Conclusions



Stat. Significance

Observed differences
must unlikely have
occurred by chance



Approach

Product-based
vs
Process-based



Measurements

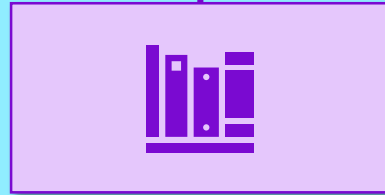
Data Point Collection
PE Time
Edit Distance



Post-Editing Task

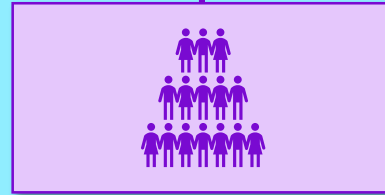
Full Post-Editing
Instructions & Guidelines

Methodology – Design of Experiments (DoE)
Validity – Reliability – Replicability – Statistical Power - Sensitivity



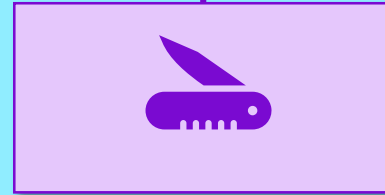
Content / Types

Relevancy and
Representativeness



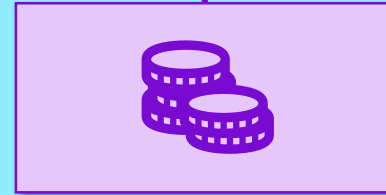
Post-Editors

Qualified PEs
Subject Matter Expertise, PE
experience, open-minded, etc.



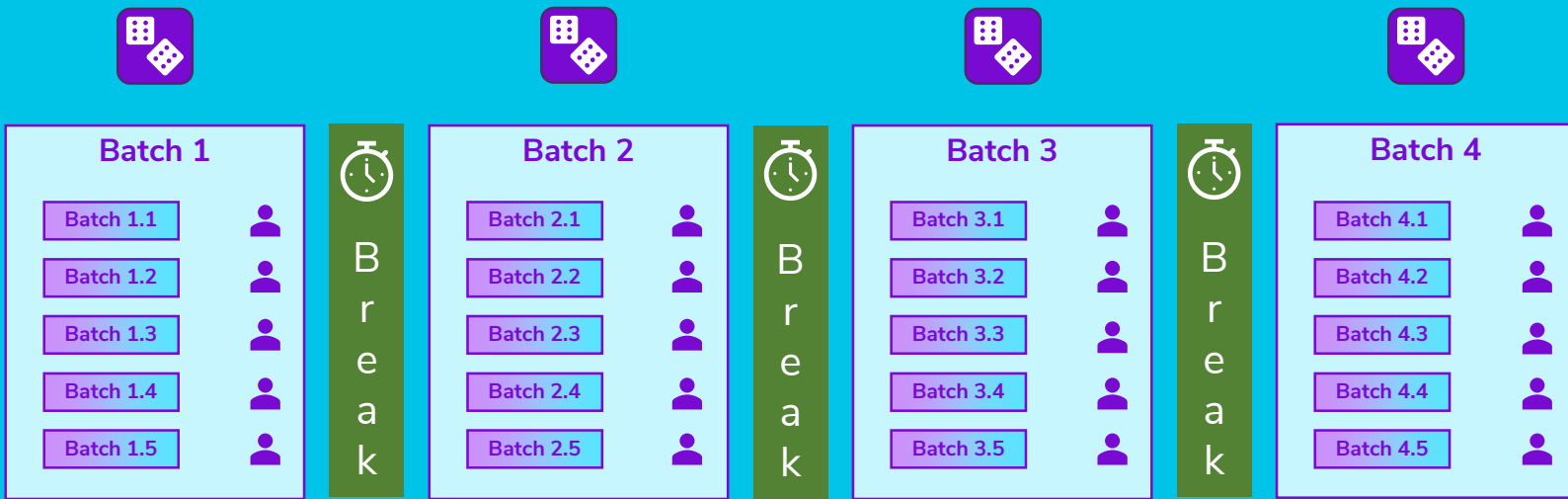
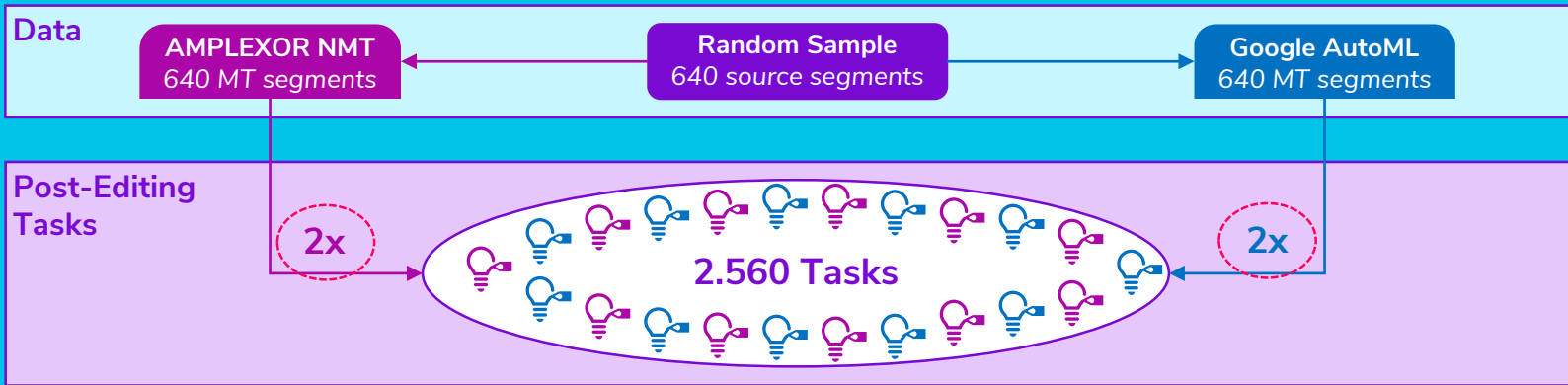
CAT Tool

MateCat
Customized
Open Source Version



Experiment Costs

Cost Effectiveness

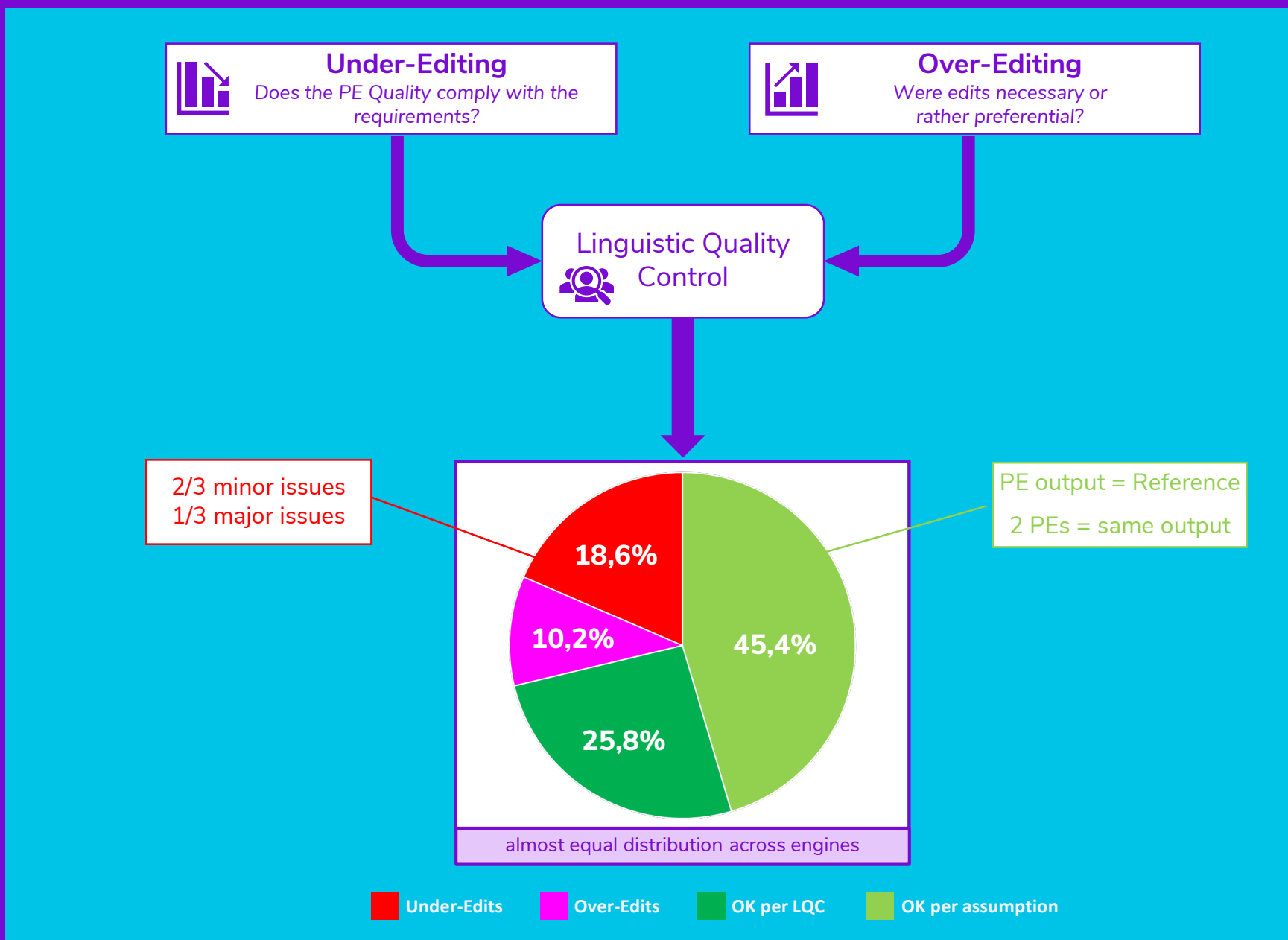
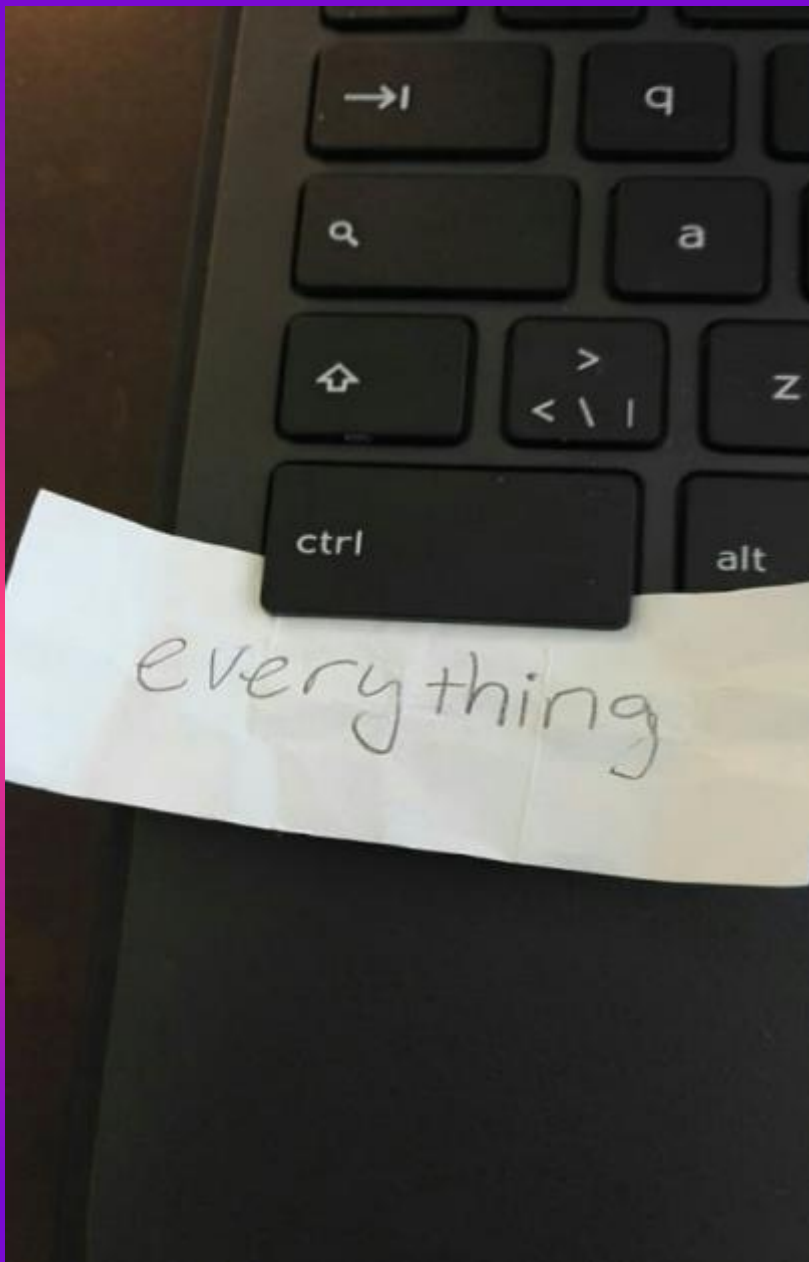


Allocation Requirements & Constraints

- Same MT Segment → same Post-Editor
- Engine balance per batch and PE
- Randomized sequences per batch
- Balanced Post-Editor pairs

Post-Editor bias and variance

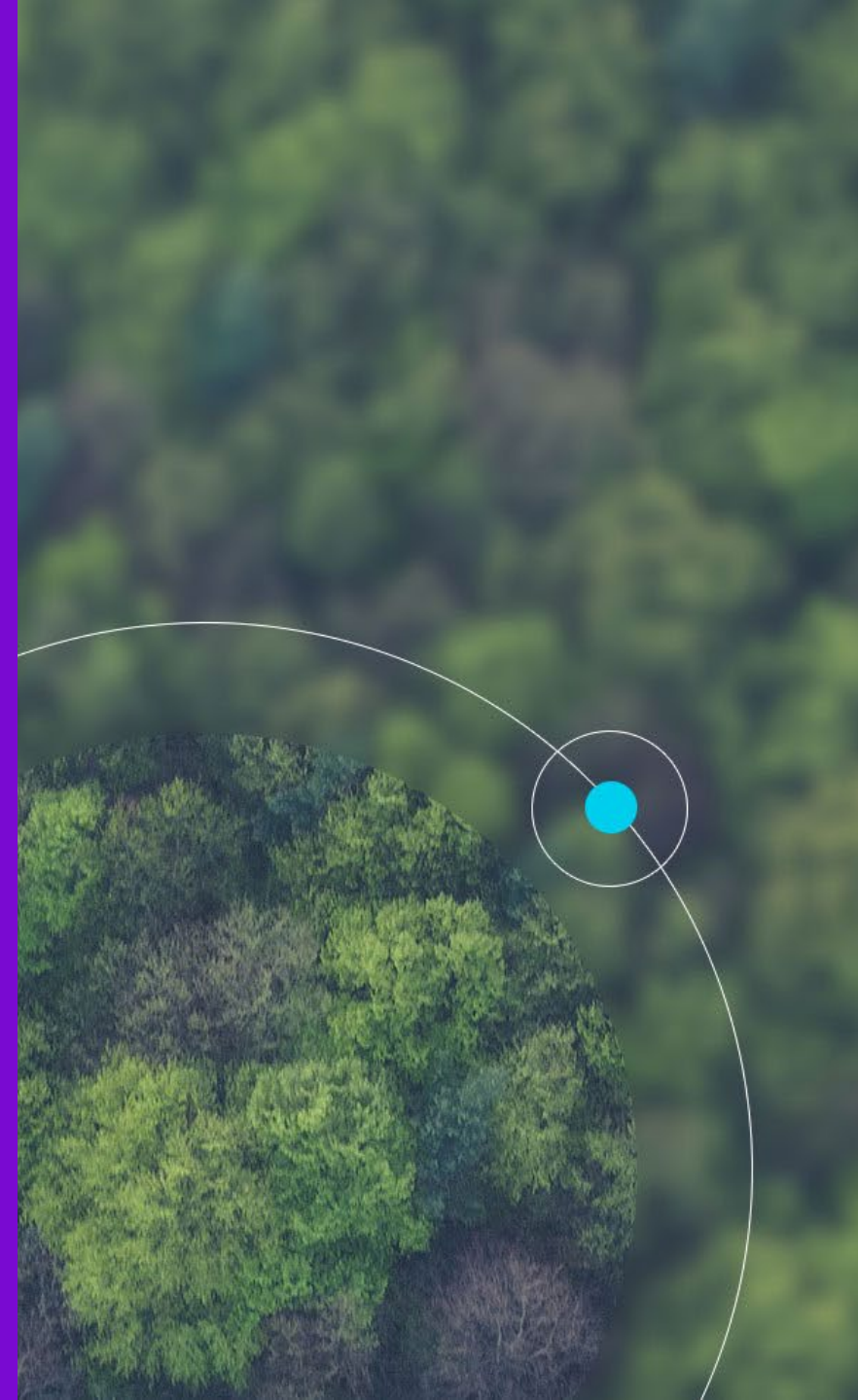


- Unawareness about engine
- Minimize Learning Effects
- Avoid mental Fatigue
- Open Mindedness





Findings

Results, Conclusions and bigger Context





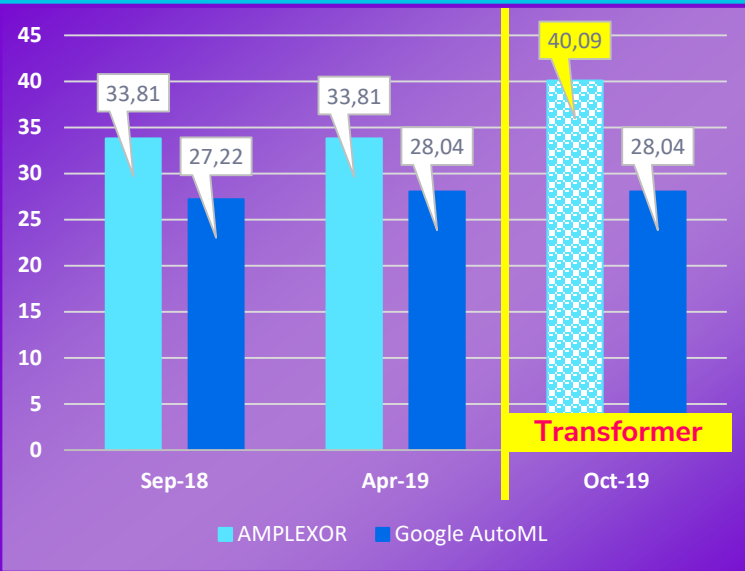
Dimension	Scope	Experiments	Description	Duration	Distance
Experiment	All	#1, #2	Experiments #1 & #2	Not for publication	
	Experiment #1	#1	Content types: Marketing, Technical, Other		
	Experiment #2	#2	Content type: Technical		
Content Type	Technical (1, 2)	#1, #2	Technical content from experiments #1 & #2		
	Technical (1)	#1	Technical content from experiment #1		
	Marketing (1)	#1	Marketing content from experiment #1		
	Other (1)	#1	Other content from experiment #1		

Executive Summary

- **Post-Editing Time / Throughput**
 - **Overall:** Google AutoML outperforms AMPLEXOR NMT
 - **Marketing:** AMPLEXOR NMT is significantly short of Google AutoML
 - **Technical:** AMPLEXOR NMT shows slight (non-significant) advantage over Google AutoML
- **Edit Distance:** AMPLEXOR NMT output consistently requires more edits than Google AutoML output
- **Throughput variance** between Post-Editors (fastest vs slowest): **factor 2,5+**

BLEU

higher score = better



AMPLEXOR (Apr-19 vs Oct-19)

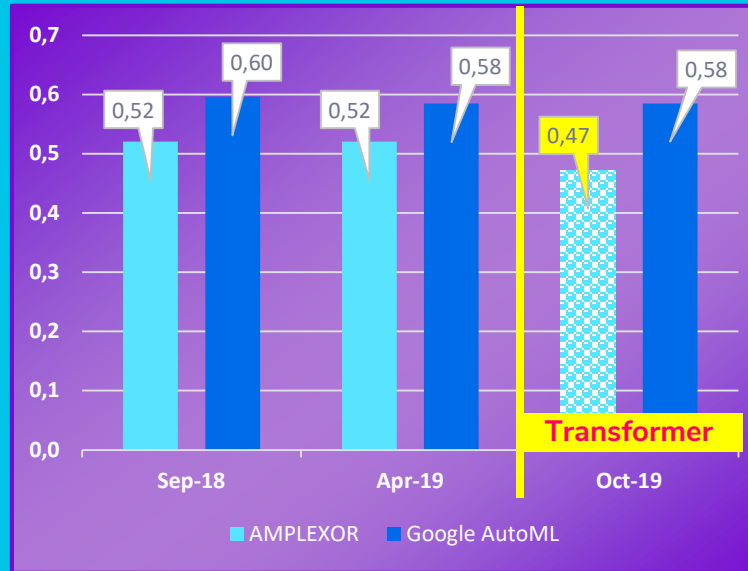
- Delta: 6,28 (+18,6%)

AMPLEXOR vs Google AutoML (Oct-19)

- Delta: 12,05
- Google AutoML: -30,0%

TER - Translation Error Rate

lower score = better



AMPLEXOR (Apr-19 vs Oct-19)

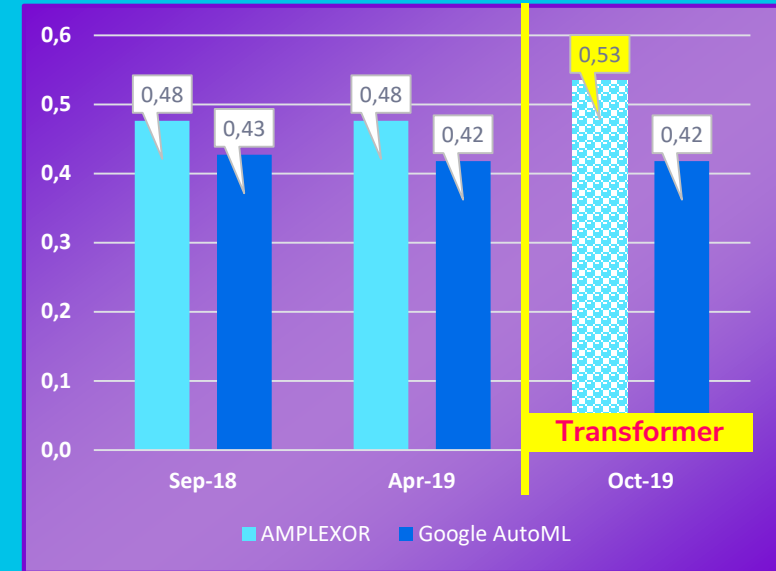
- Delta: 0,05 (-9,6%)

AMPLEXOR vs Google AutoML (Oct-19)

- Delta: 0,11
- Google AutoML: +23,4%

METEOR

higher score = better



AMPLEXOR (Apr-19 vs Oct-19)

- Delta: 0,05 (+10,4%)

AMPLEXOR vs Google AutoML (Oct-19)

- Delta: 0,11
- Google AutoML: -20,8%





Engine Portfolio

Best of Breed (based on Use Case)



Continuous Benchmarking

Productivity Analysis



Enterprise-grade NMT

Complementary capabilities



Linguistic Data Asset Management



Augmented Translation

AI-driven Localization Ecosystem



End-to-End Process

Project Mgt, Desktop Publishing, etc.



People

Committed, experienced, etc.



AMPLEXOR

EMBRACE THE FUTURE



Thank you for your time.

Markus Welsch

Vice President Content Intelligence

+352 314 411 420

markus.welsch@amplexor.com

