Maxim Khalilov, Ph.D.

Machine translation that makes sense: the Booking.com use case

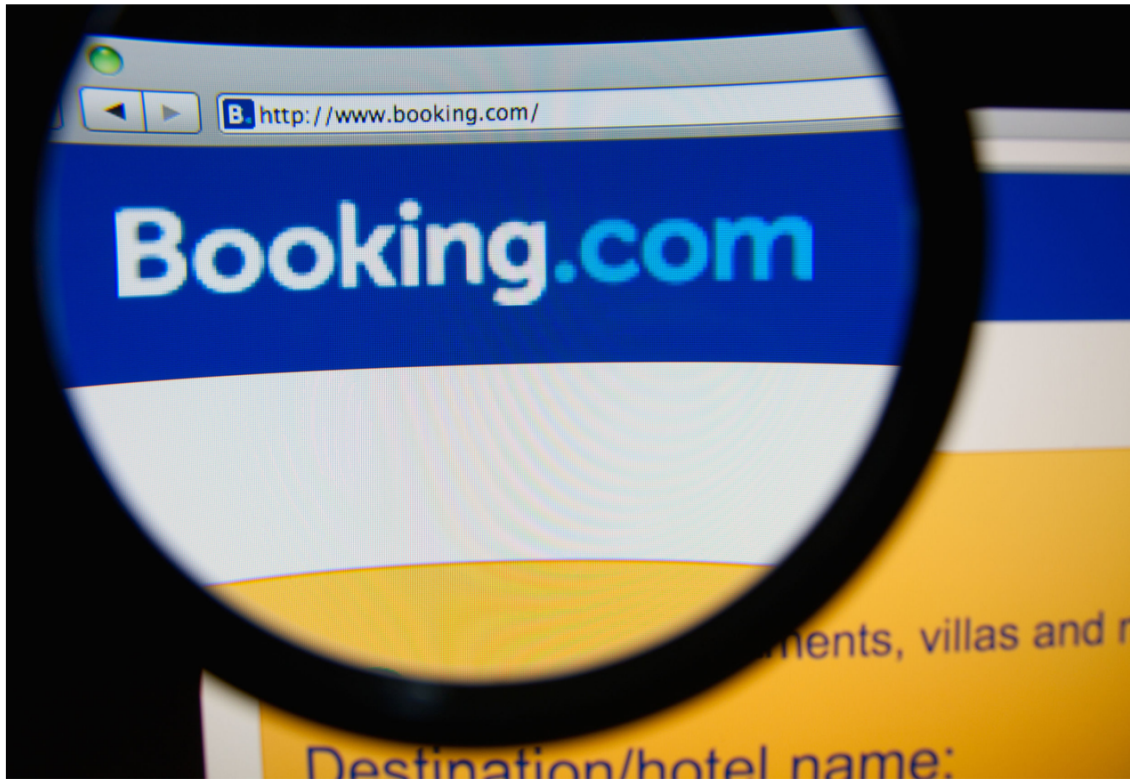
Technical presentation

March 6, 2018
Cambridge, UK

Booking.com.

The world's #1 website for booking hotels and other accommodations.

- Founded in **1996** in Amsterdam
- Part of the Priceline Group (NASDAQ: **PCLN**) since **2005**
- **1,500,000+** properties in more than 220 countries and territories representing over **27M** rooms
- Over **1,550,000** room nights every 24 hours
- Number of unique destinations worldwide: **120,000+**
- Total number of guest reviews: **173,000,000+**
- **43** languages
- **198** offices worldwide
- More than **15,500** employees



Use case of MT at Booking.com

Mission: Empower people to experience the world without any language barrier.

of daily bookings on Booking.com is made in a language other than English

... thus it is important to have **locally relevant content at scale**

How Locally Relevant?

Allow partners and guests to **consume and produce content in their own language**

- Hotel Descriptions
- Customer Reviews
- Customer Service Support

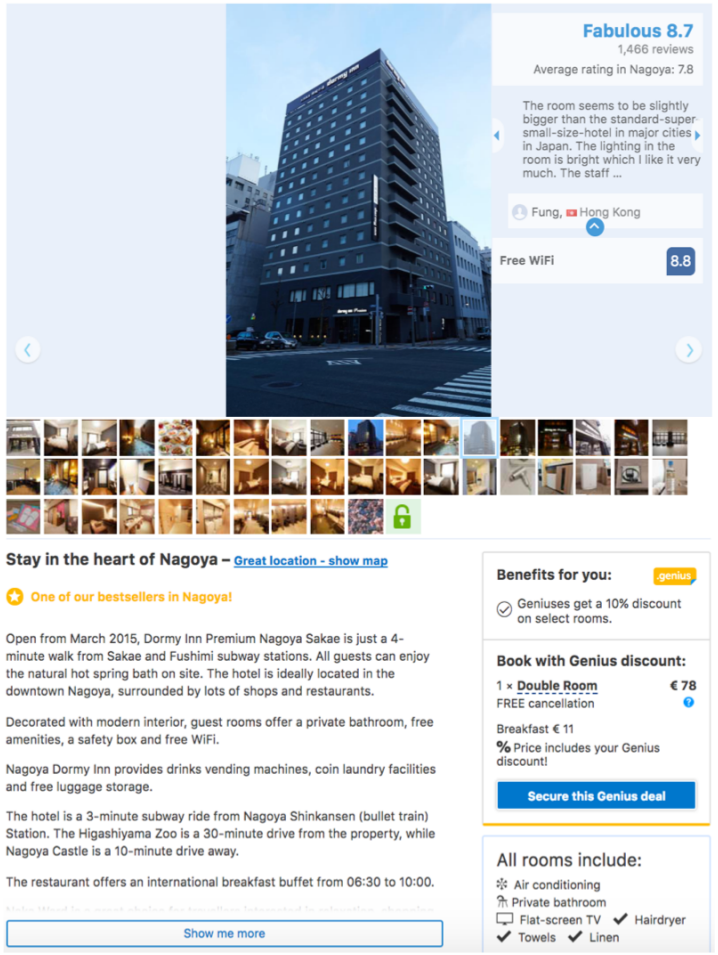
Why At Scale?

- **One Million+ properties** and growing very fast
- **Frequent change requests** to update the content
- **43 languages** and more
- New user-generated **customer reviews / tickets** every second

Why MT?

Limited domain	One product
Lots of in-domain data	Av. 10M parallel sent. for big languages
Language expertise	In-house evaluators for 43 languages

Use Case #1: Hotel descriptions – currently translated by human in 43 languages based on visitor demand.



Most often used by people in Japan

日本語 | English (US) | 한국어
English (UK) | 简体中文 | 繁體中文

All languages

English (UK)	Čeština	Українська
English (US)	Magyar	Bahasa Indonesia
Deutsch	Română	Bahasa Malaysia
Nederlands	日本語	ภาษาไทย
Français	简体中文	Eesti
Español	繁體中文	Hrvatski
Español (AR)	Polski	Lietuvių
Català	Ελληνικά	Slovenčina
Italiano	Русский	Srpski
Português (PT)	Türkçe	Slovenščina
Português (BR)	Български	Tiếng Việt
Norsk	العربية	Filipino
Suomi	한국어	Íslenska
Svenska	עברית	
Dansk	Latviski	

Hotel in Japan

German Visitor

Sees English description

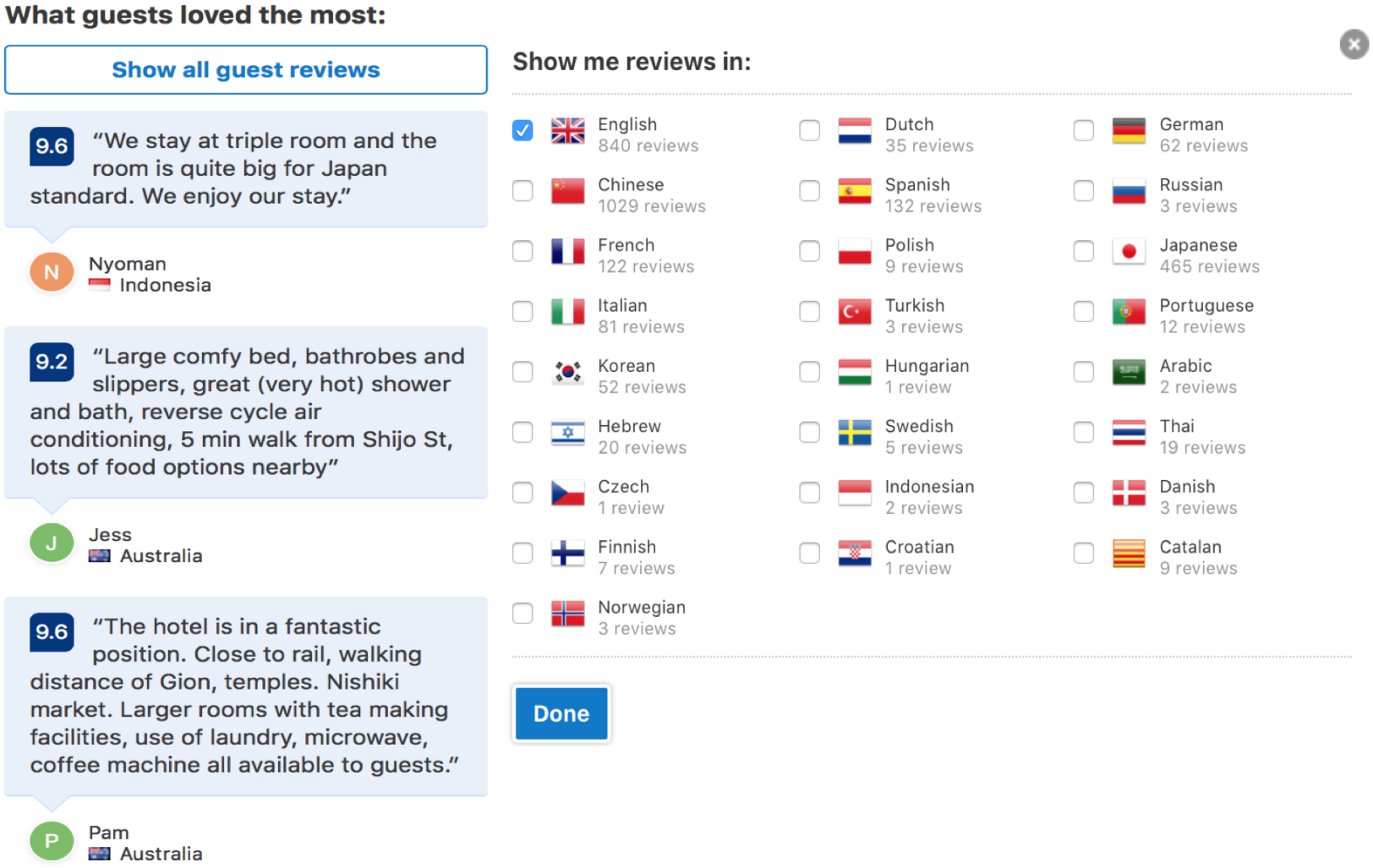
Drops Off

Lost Business

Human Translation Pipeline

Machine Translation

Use Case #2: Customer Reviews – currently not translated; available only if user leaves a review in that language.



Hotel in Japan

German Visitor

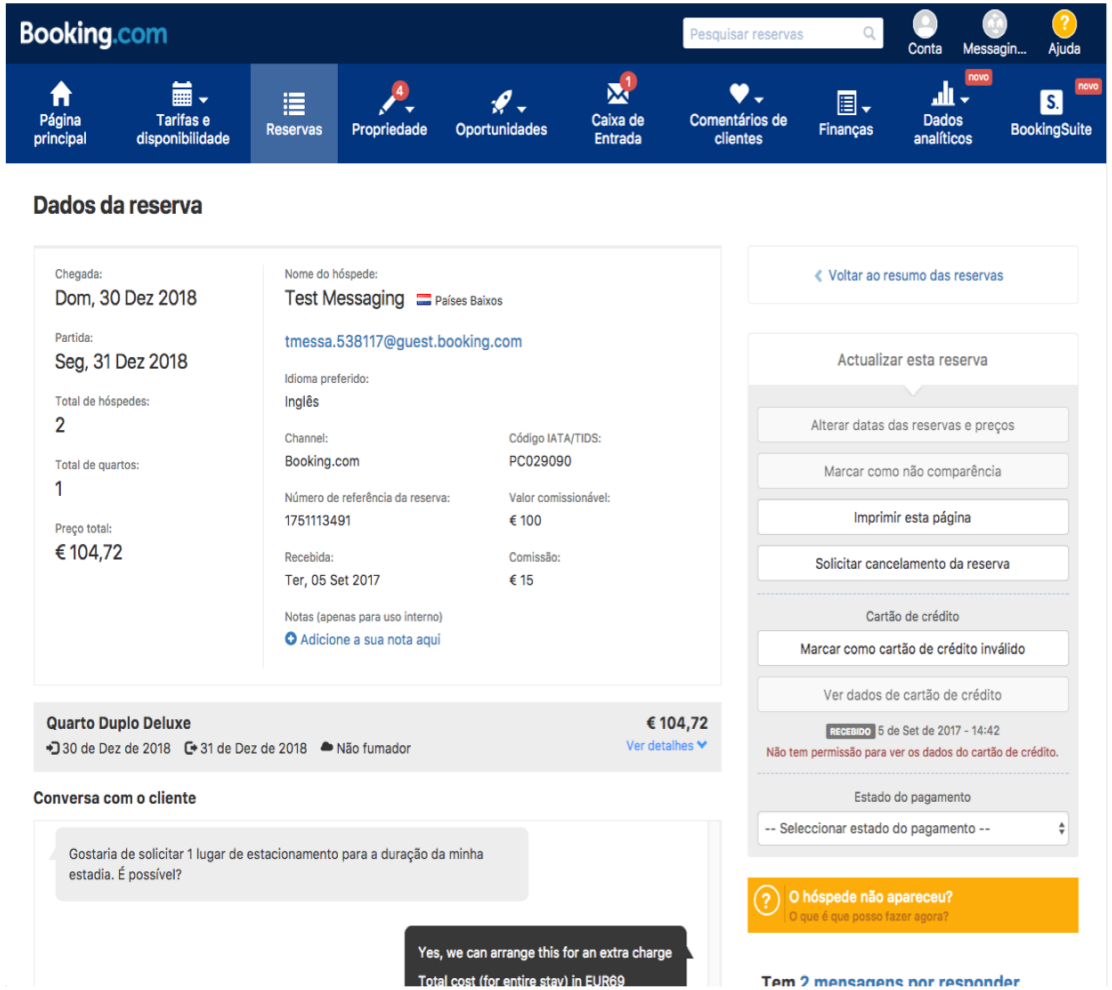
No German Reviews

Drops Off

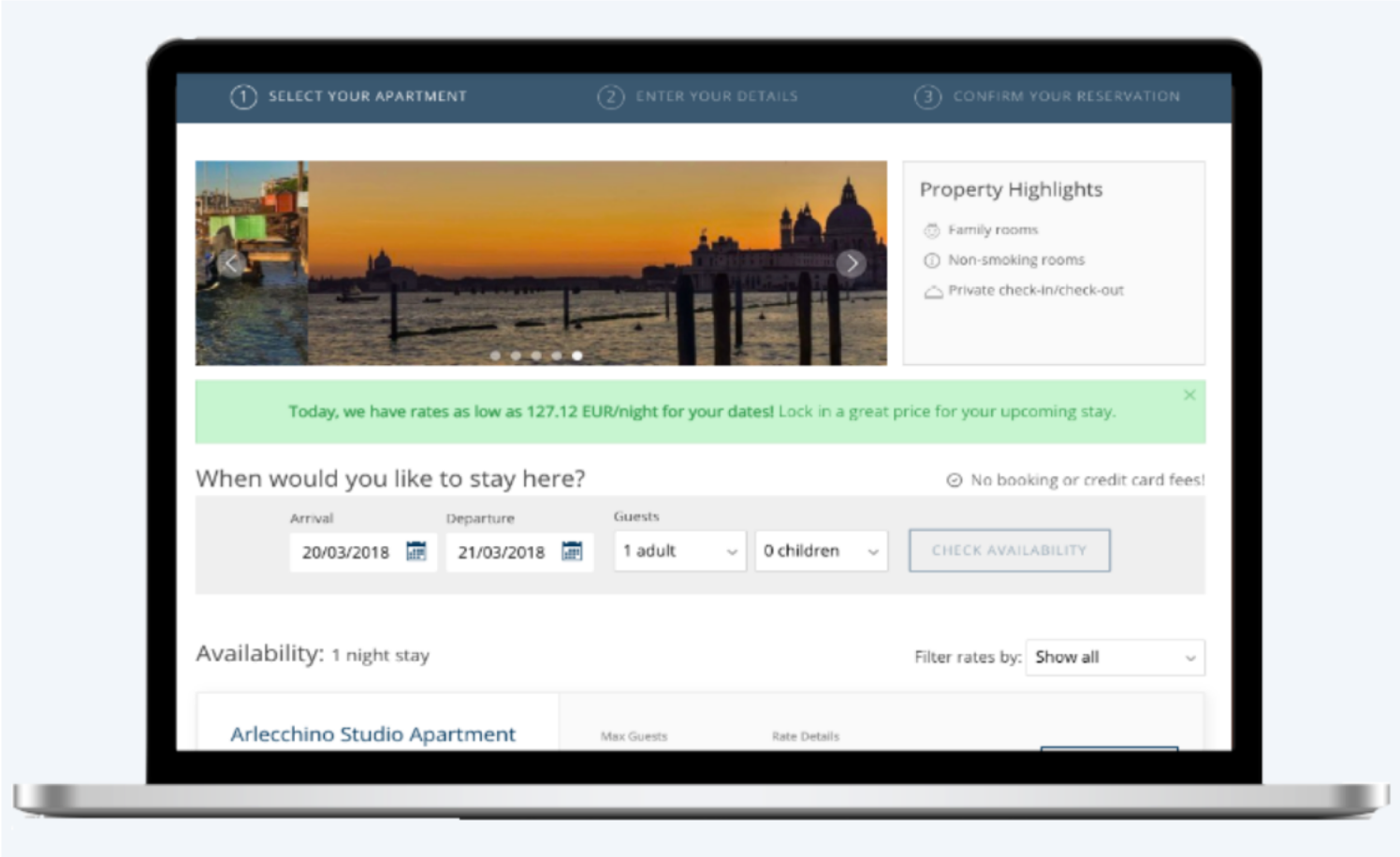
Lost Business

Machine Translation

Use Case #3: Partner support – Partner-facing localization and customer/partner support.



Use Case #4: Translation support – make translation cheaper by providing high-quality productivity tools.



And there is even more..

Messages.

Ask a question

Say hello to your host or send a request!

Please write your requests in English or German.
Special requests cannot be guaranteed–but the accommodation will do its best to meet your needs.
You can always make a special request after your booking is complete!

我想要一个安静的房间

- Introduce yourself to your host
- What brings you to the area?
- Who are you travelling with?

Attractions .

Attractions Recommended by Locals:

Hackescher Höfe
(0.7 km)
This collection of 8 courtyards has developed into a real entertainment hub. Both locals and tourists flock to the art galleries, independent shops and lively bars.

Mauerpark
(1.9 km)
With outdoor karaoke and dozens of market stalls, weekends at Mauerpark are certainly not a quiet affair. The basketball courts are where sports fans can show off their skills.

Tiergarten
(2.9 km)
This park is popular with locals who want to picnic on the grass or kick a ball around. In the winter, skaters spin circles on the frozen lakes.

Room descriptions.

Doppelzimmer mit Gemeinschaftsbad

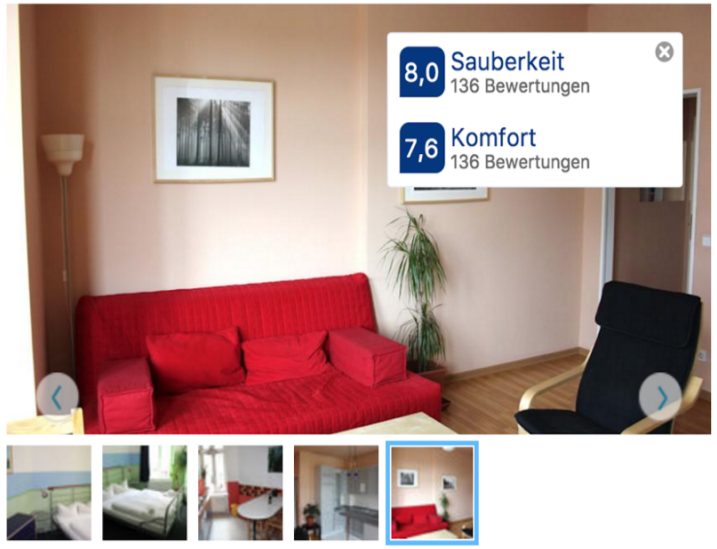
✓ KOSTENLOSE Stornierung vor 23:59 Uhr am 23. Januar 2018

✓ KEINE VORAUS/-ANZAHLUNG NOTWENDIG – Zahlen Sie in der Unterkunft

Frühstück € 4

€ 60,90 für 1 Nacht

Tolles Schnäppchen heute



Reservieren

Es dauert nur 2 Minuten

Zimmergröße 12 m²

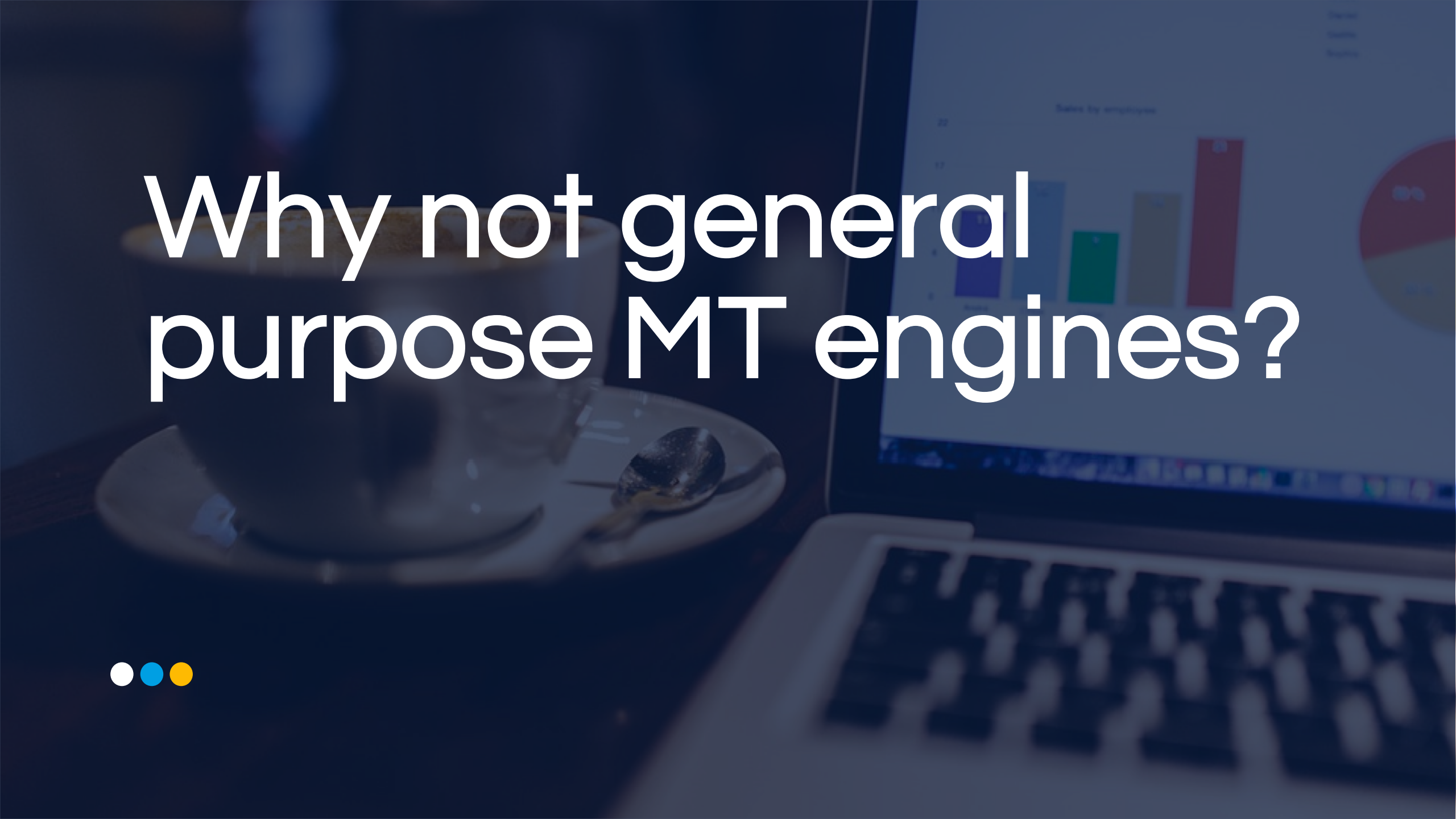
Dieses modern und gemütlich eingerichtete Doppelzimmer befindet sich in einem gemeinschaftlich genutzten Apartment und bietet Zugang zu einem gemeinschaftlich genutzten Wohnbereich mit Küchenzeile und einem Gemeinschaftsbad.

Zimmerausstattung:
Küchenzeile, Gemeinschaftsbad, Heizung, Handtücher, Bettwäsche, Toilettenpapier

Kostenfreies WLAN!

P Parkplätze stehen zur Verfügung

Vor Kurzem gebucht!



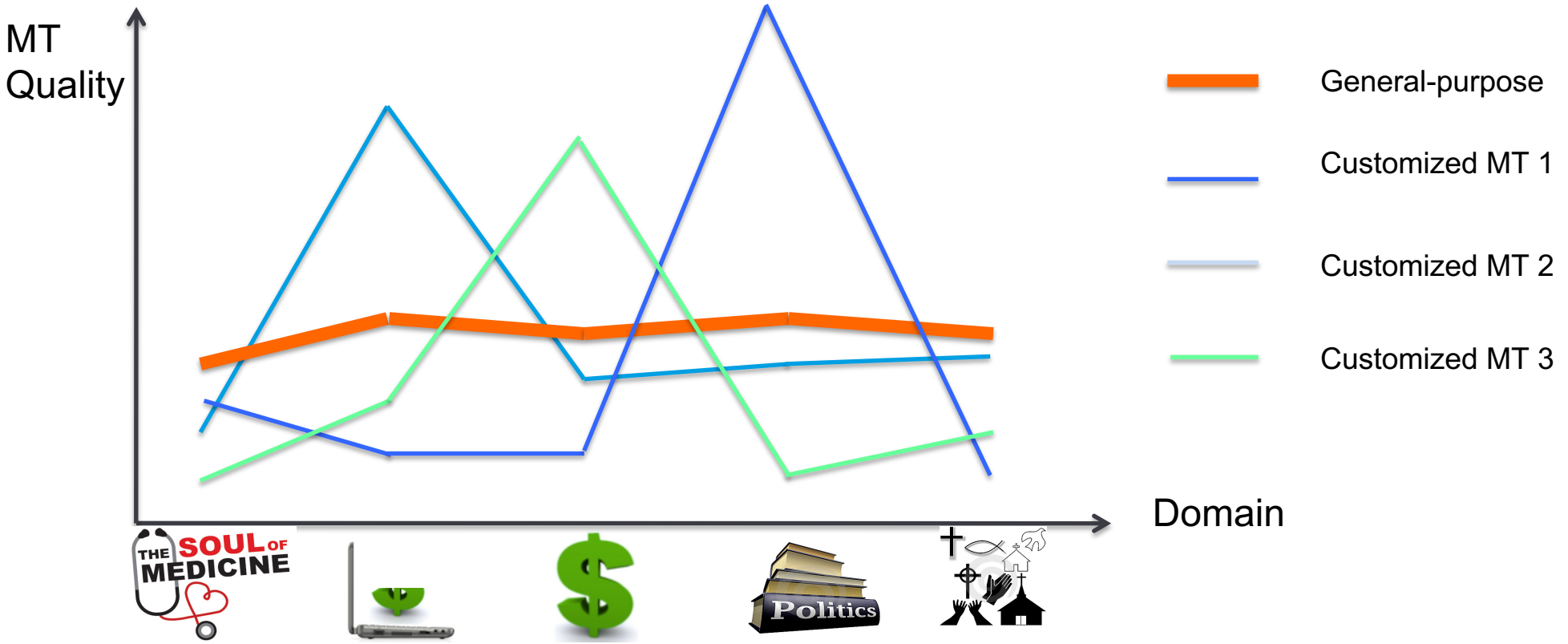
Why not general purpose MT engines?

3

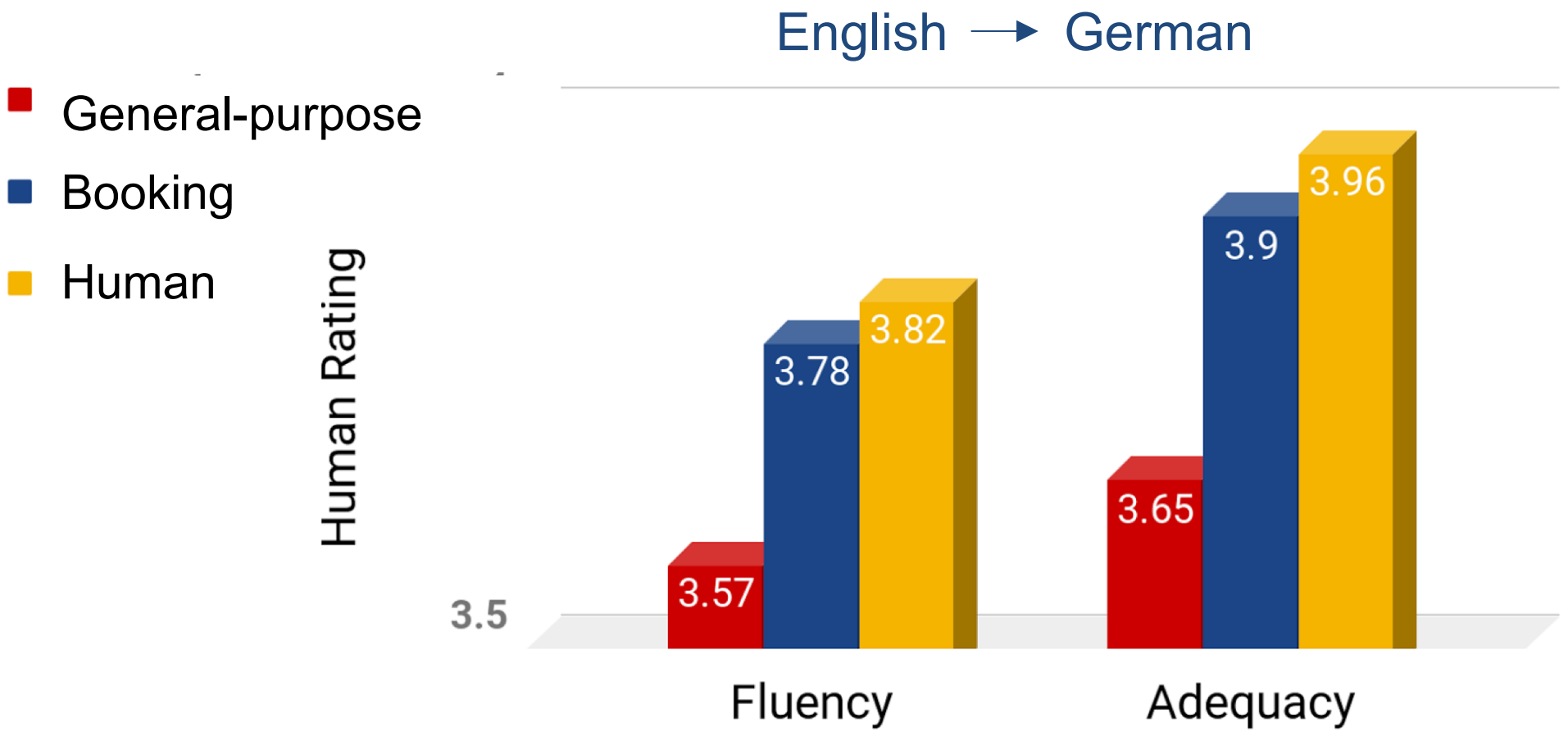
Reasons

1. Quality

Customized MT can do much better for our own content.

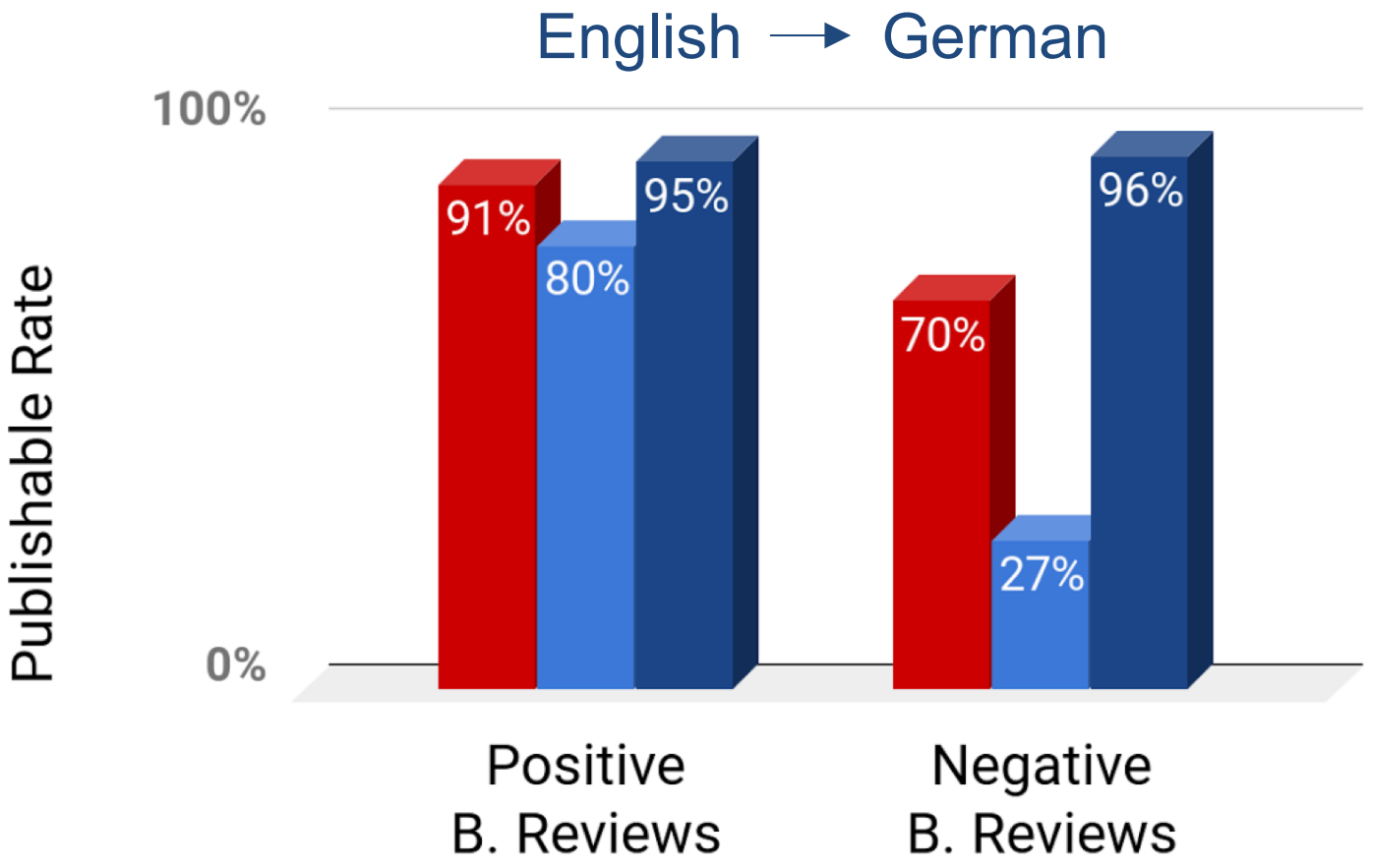


Hotel Description: Evaluation Results



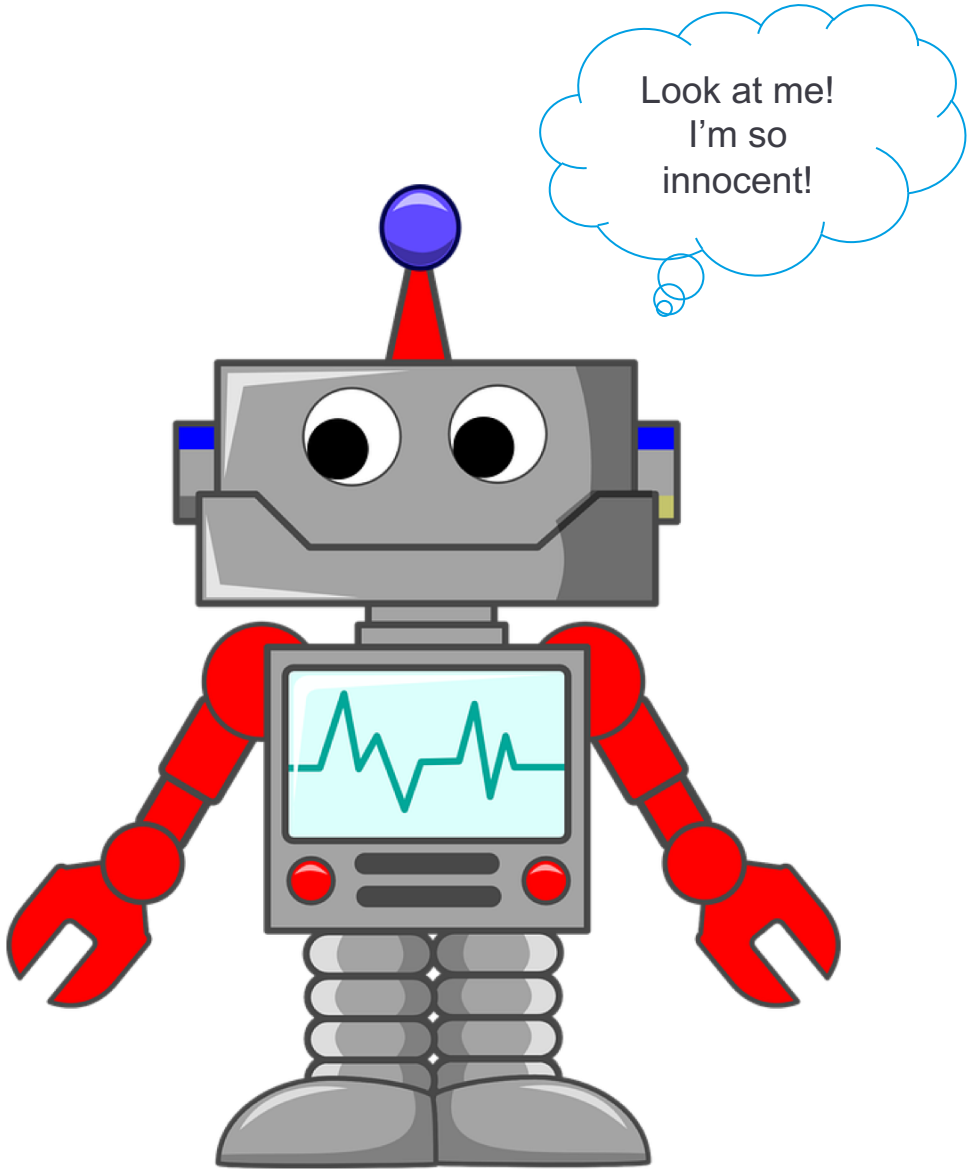
Customer Review: Evaluation Results

- General purpose
- Booking:
  - Hotel Description
- Booking:
  - Customer Review



1. Quality
2. **Risk**

Can machine translation be dangerous?

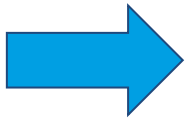


Booking.com

The imperfection of MT might mislead users, have legal consequences for the company or damage brand's reputation and customer’s confidence of translated content.

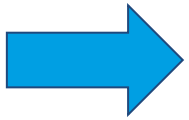
Examples of business sensitive errors

Offering a restaurant with WiFi, Hodor Ecolodge is located in Winterfell. On-site **parking is free**.



Die Hodor Ecolodge in Winterfell bietet ein Restaurant mit WLAN. **Parkplatz vor Ort ist verfügbar**.

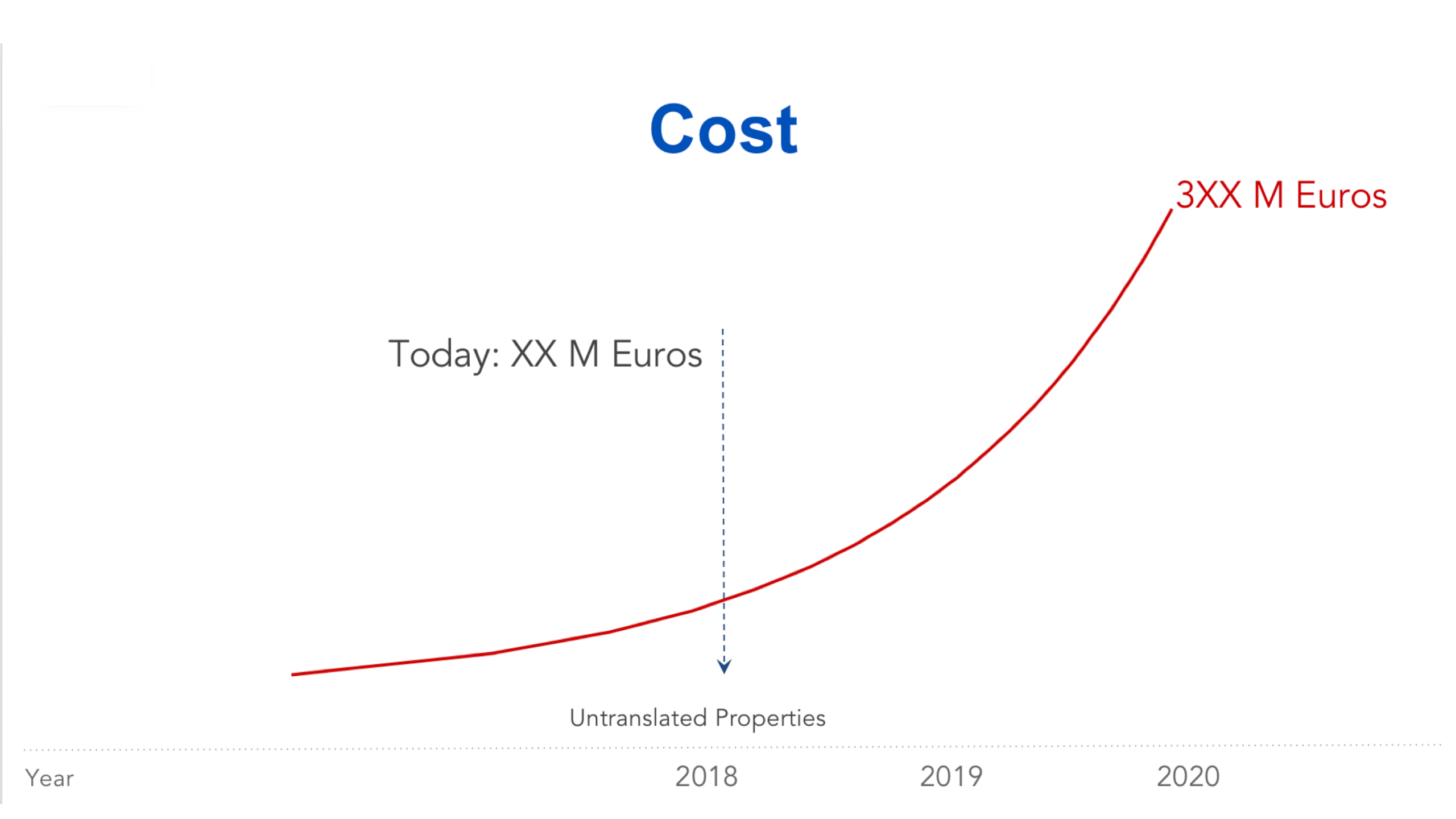
The hotel offers 24-hour concierge service and free-use bicycles. **Pets can be accommodated** with advance reservation.



Der Conciergeservice steht rund um die Uhr zu Ihrer Verfügung und die Leihfahrräder nutzen Sie kostenfrei.

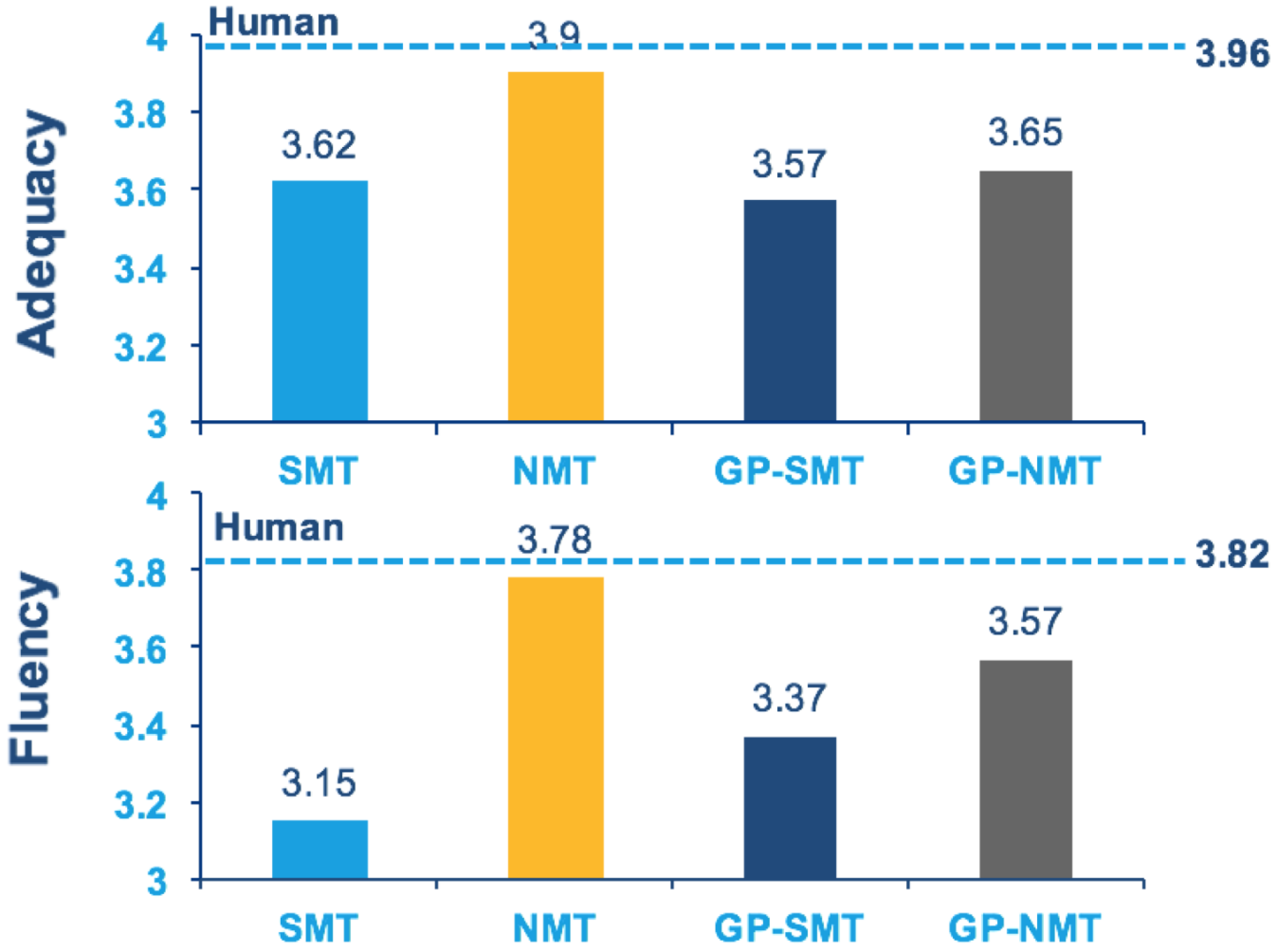
1. Quality
2. Risk
3. **Cost**

Cost


3XX M Euros
Today: XX M Euros
Untranslated Properties
Year
2018
2019
2020

But why neural?

Adequacy / Fluency Scores for EN->DE hotel description translations



Our In-domain NMT system outperforms all other MT engines

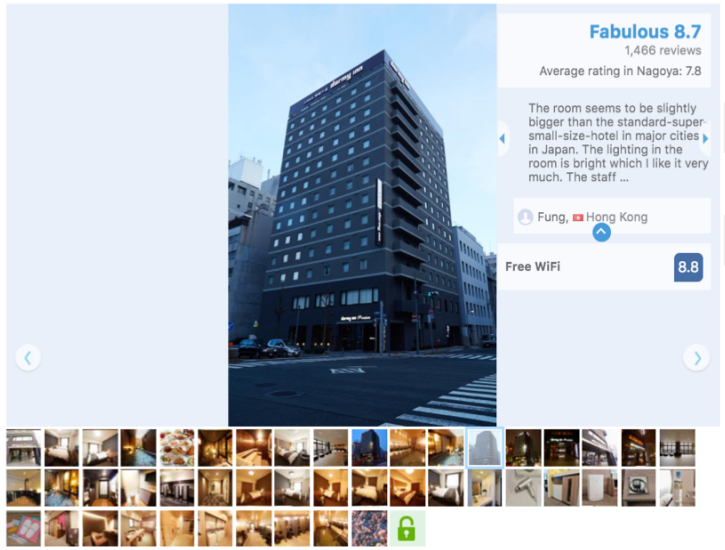
Both Neural systems still consistently outperform their statistical counterparts

General Purpose NMT beats In-domain SMT

Particularly fluency score of our NMT engine is close to human level

The Data

Hotel descriptions translated by human in 43 languages resulting in lots of in-domain data for MT



Most often used by people in Japan

- 日本語
- English (UK)
- English (US)
- 简体中文
- 한국어
- 繁體中文

All languages

- English (UK)
- English (US)
- Deutsch
- Nederlands
- Français
- Español
- Español (AR)
- Català
- Italiano
- Português (PT)
- Português (BR)
- Norsk
- Suomi
- Svenska
- Dansk
- Čeština
- Magyar
- Română
- 日本語
- 简体中文
- 繁體中文
- Polski
- Ελληνικά
- Русский
- Türkçe
- Български
- العربية
- 한국어
- עברית
- Latviski
- Українська
- Bahasa Indonesia
- Bahasa Malaysia
- ภาษาไทย
- Eesti
- Hrvatski
- Lietuvių
- Slovenčina
- Srpski
- Slovenščina
- Tiếng Việt
- Filipino
- Íslenska

50%
Translation Coverage

90%
Demand Coverage

10M
Average Corpus Size

* Approximate numbers based on average of some languages

Monolingual reviews never translated in 43 languages resulting in lots of out-of-domain data potentially useful for MT

„Es war alles ziemlich nach vorne, das Zimmer hatte eine schöne Größe, die Betten waren bequem, wir brauchten keine Aussicht."

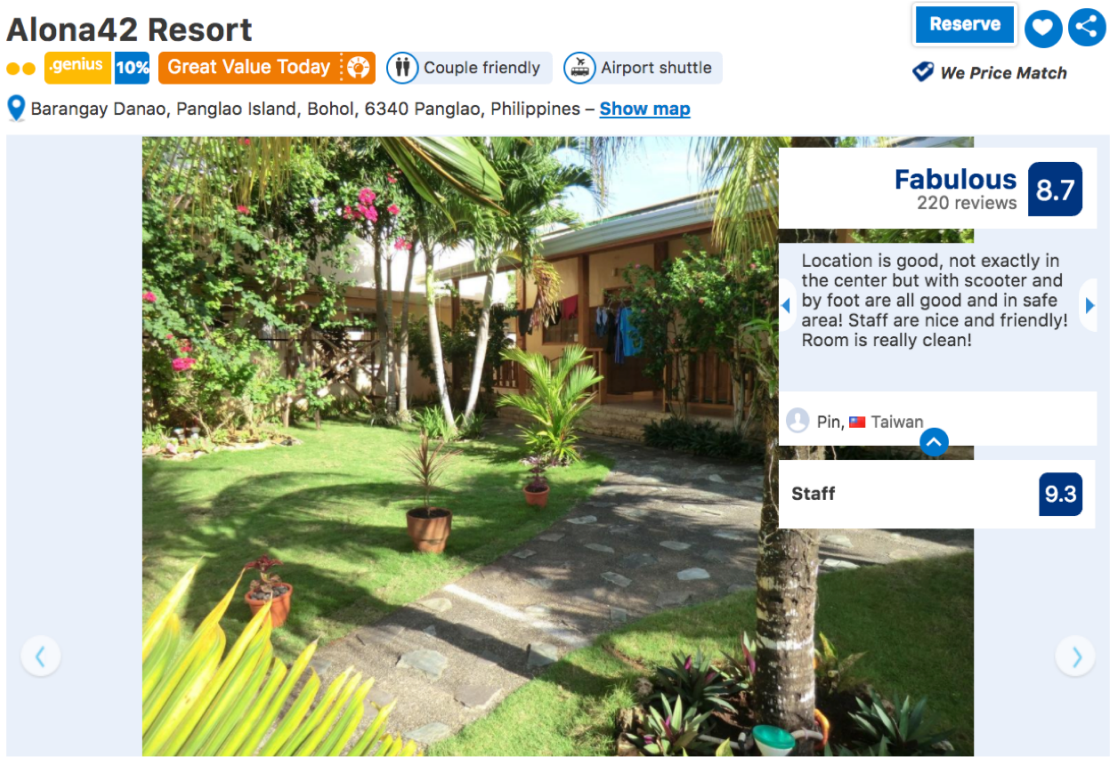
Übersetzt aus: English - Original anzeigen

M Morgan
Großbritannien

„Sauber, tolle Lage, wunderbare und große Bar für die Gäste mit ausgezeichneten großen Bildschirmen (Football an diesem Abend)"

Übersetzt aus: English - Original anzeigen

M Markus
Deutschland



173M
Total reviews

17
Languages >1M reviews

37%
Properties w/o reviews

Few specific challenges and proposed solutions

Our NMT Model Configuration Details

Data Preparation	
Split Data	Train, Val, Test
Input Text Unit	Word Level
Tokenization	Aggressive
Max Sentence Length	50
Vocabulary Size	50,000

Model	
Model Type	seq2seq
Input Embedding Dimension	1,000
RNN Type	LSTM
# of hidden layers	4
Hidden Layer Dimension	1,000
Attention Mechanism	Global Attention

** Approx. 220 Million Parameters

Training	
Optimization Method	Stochastic Gradient Descent
Initial Learning Rate	1
Decay Rate	0.5
Decay Strategy	Decrease in Validation Perplexity <=0
Number of Epochs	5 - 13
Stopping Criteria	BLEU + sensitive sentences +constraints
Dropout	0.3
Batch Size	250

** 1 Epoch takes approx. 2 days on a single NVIDIA Tesla K80 GPU

Translate	
Beam Size	10
Unknown Words Handling	Source with Highest Attention

Evaluate	
Auto	BLEU
Human	A/F
Other	A/B Test

** MT pipeline based on Harvard implementation of OpenNMT

Our challenges

Real-world content

- Named entities
- Rare words

Customer facing output

- Human loop
- BLEU & human evaluation correlation
- Business sensitive issues

Lack of parallel training data

- Use and sources of data
- Domain adaptation

Our challenges

Real-world content

- Named entities
- Rare words

Customer facing output

- Human loop
- BLEU & human evaluation correlation
- Business sensitive issues

Lack of parallel training data

- Use and sources of data
- Domain adaptation

End-to-end approach insufficient to handle Named Entities, pre-processing improves performance

Problem

Hotel name | Time | Landmark name | Landmark name

Hotel Shirakawa is just a **5-minute** walk from **Fushimi Subway Station**. **Nagoya Castle** is a **10-minute** drive, and the **Sakae shopping area** is **500 m** away.

Time | Landmark name | Distance

Approach

Search for named entities in source → **Replace with placeholders** → **Translate** → **Substitute as per target format**

- Regular expression for distance, date, time
- Hybrid dictionary, conditional random field NER for names

Results

Raw source	Winterfell Railway Station can be reached in a **55-minute** car ride.
Pure NMT Translation	Den Bahnhof Winterfell erreichen Sie nach einer **5-mintigen** Autofahrt.
NMT with distance placeholders	Den Bahnhof Winterfell erreichen Sie nach einer **55-mintigen** Autofahrt.

Better handling of rare words and 4 points BLEU score improvement with Byte Pair Encoding (BPE)

Raw source	Offering a restaurant with WiFi, Hodor Ecolodge is located in Winterfell.
Tokenized source	offering^{C} a^{L} restaurant^{L} with^{L} wi ■^{C} fi^{C} ■,^{N} ho ■^{C} dor^{L} ecolodge^{L} is^{L} located^{L} in^{L} winter ■^{C} fell^{L} ■.^{N}
Tokenized output	die^{C} ho ■^{C} dor^{L} ecolodge^{C} in^{L} winter ■^{C} fell^{L} bietet^{L} ein^{L} restaurant^{C} mit^{L} wlan^{U} ■.^{N}
De-tokenized output	Die Hodor Ecolodge in Winterfell bietet ein Restaurant mit WLAN.

BLEU	50K-Vocab baseline	Joint BPE				Separate BPE			
		30K	**50K**	**70K**	**90K**	**30K**	**50K**	**70K**	**90K**
Epoch 5	39.54	**43.75**	43.46	43.40	41.23	42.81	42.35	39.73	N/A
Epoch 10	40.95	**44.55**	44.52	43.81	43.81	43.39	43.48	43.51	
Epoch 15	42.01	45.08	45.91	**46.14**	45.75	43.58	43.23	45.17	
Epoch 20	42.15	46.31	46.43	**46.61**	45.62	45.22	46.00	45.90	

Translation of informal language of customer reviews and partner-(company)-user comms

Examples

- The stuff
- The night guy aund the girl in the morning who looks like manage the hotel
- They keep your luggage for free if you for some days to Sapa
- And as well the offered us a breakfast in the morning asap
- Thans for the detail

Approach

Correct typos which are easy to fix → **Adapt to the UGC domain** → **Translate** → **Iterate**

Results

Adequacy score	Positive reviews	Negative reviews
Baseline	80 %	27 %
+typos correction+DA	95 %	96 %

Our challenges

Real-world content

- Named entities
- Rare words

Customer facing output

- Human loop
- BLEU & human evaluation correlation
- Business sensitive issues

Lack of parallel training data

- Use and sources of data
- Domain adaptation

How can we control (M)T quality in eCommerce environment?

Integrated approach to MT evaluation.

BLEU

Applicable to make sure there are no new bugs introduced as the result of the MT engine retraining and some experiments.

Entity analysis

Scoring the quality of entity handling.

Adequacy/Fluency scoring

Rough assessment of the MT-ed content in terms of its publishability

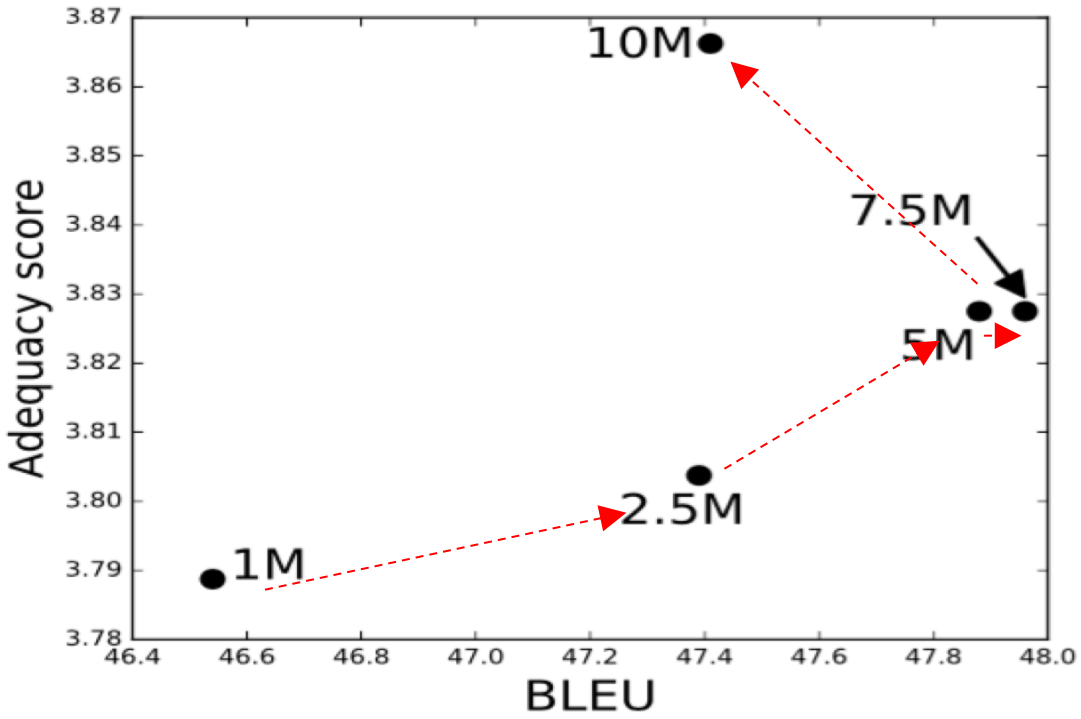
Business Sensitivity Analysis

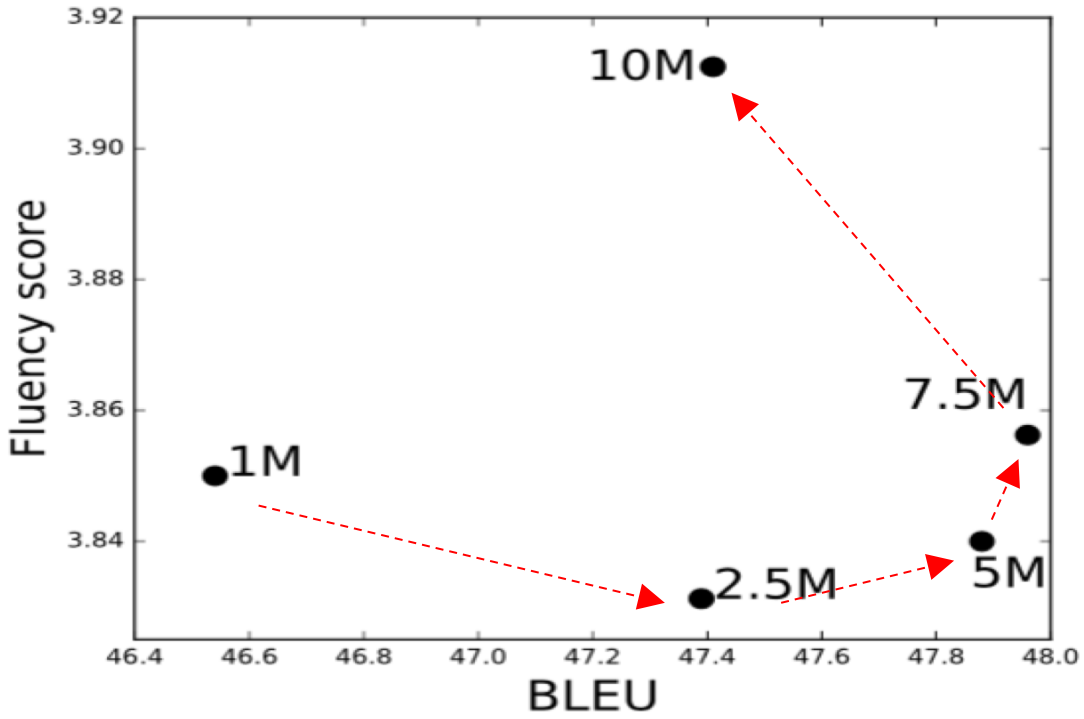
Links MT quality with potential threats for the business

A/B testing

Two-sample hypothesis testing where business metrics are to be optimized

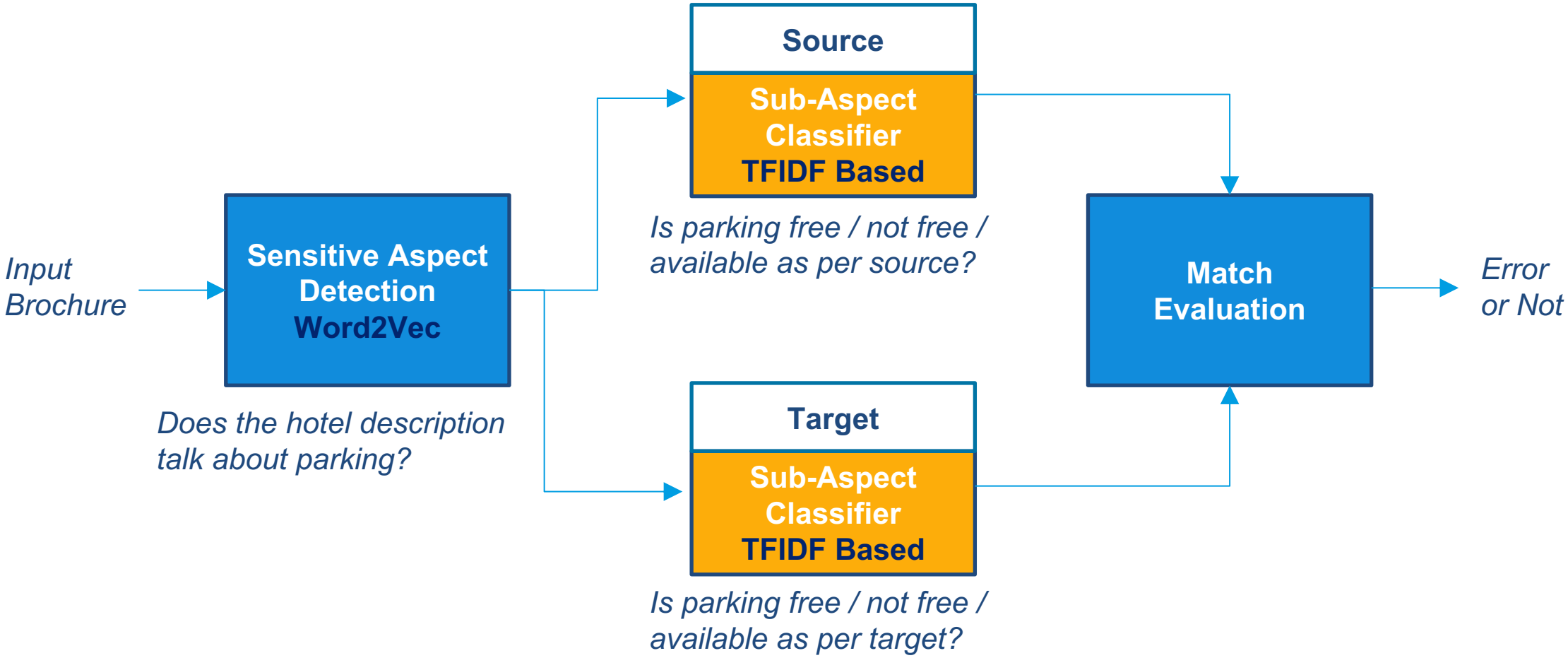
Improvement with more data is better seen from human evaluation...





...which doesn't seem to be completely aligned with BLEU

Business Sensitivity Framework to detect if aspects and sub-aspects match between source & translated content



Business Sensitivity Framework: results

FREE/NOT FREE PARKING	translation			
source		free parking	not free parking	not about parking
	free parking	*99.4%*	*0.5%*	*0.1%*
	not free parking	*5.1%*	*94.6%*	*0.3%*
	not about parking	*<0.1%*	*<0.1%*	*99.9%*

Our challenges

Real-world content

- Named entities
- Rare words

Customer facing output

- Human loop
- BLEU & human evaluation correlation
- Business sensitive issues

Lack of parallel training data

- Use and sources of data
- Domain adaptation

Method.

- A few thousand of in domain sentences.

-In addition to the hotel descriptions data, available external open data is used including data from:

- *Movie subtitles*
- *Wikipedia*
- *TED talks*
- *New commentary*
- *EuroParl*

-Synthetic Data

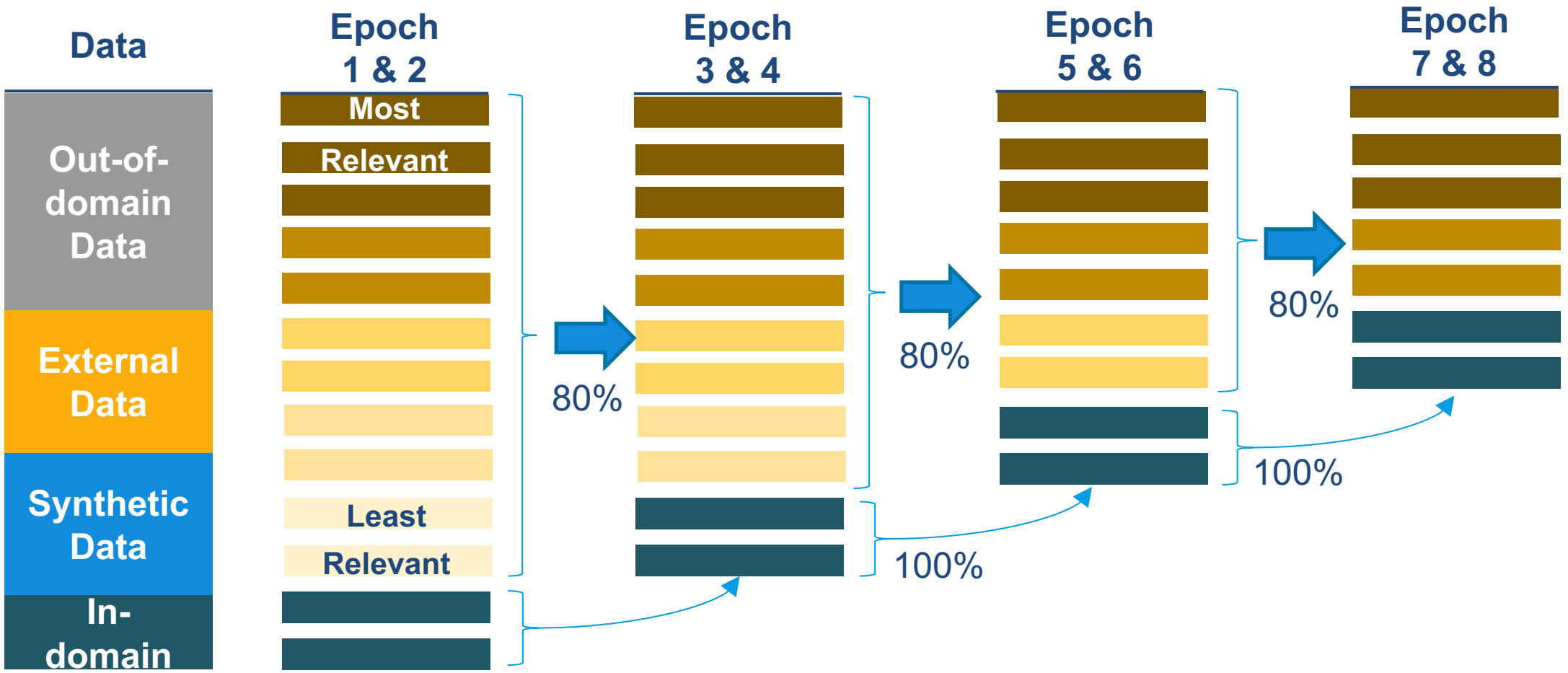
-Gradual downsampling (Wees et al., 2017)

Booking.com

Data generation for customer reviews based on mono - lingual / non-parallel bilingual data

Data	Idea	Methodology
External Corpus	Use in-domain language model to select most relevant sentences from external corpus	**Bilingual Cross Entropy Difference** (Axelrod et al) - To select sentences that are most similar to in-domain but different to out-of-domain.
Synthetic Data	Use large amount of mono-lingual data to create some synthetic in-domain data	**Rico Sennrich et al**. – Back translate target language in-domain data into source by reversing our MT model.
In-domain Data	Create a small amount of in-domain corpus as well, to test for additional impact	**Human Translation**

Domain Adaptation using gradual downsampling to most relevant data selected by in-domain language model

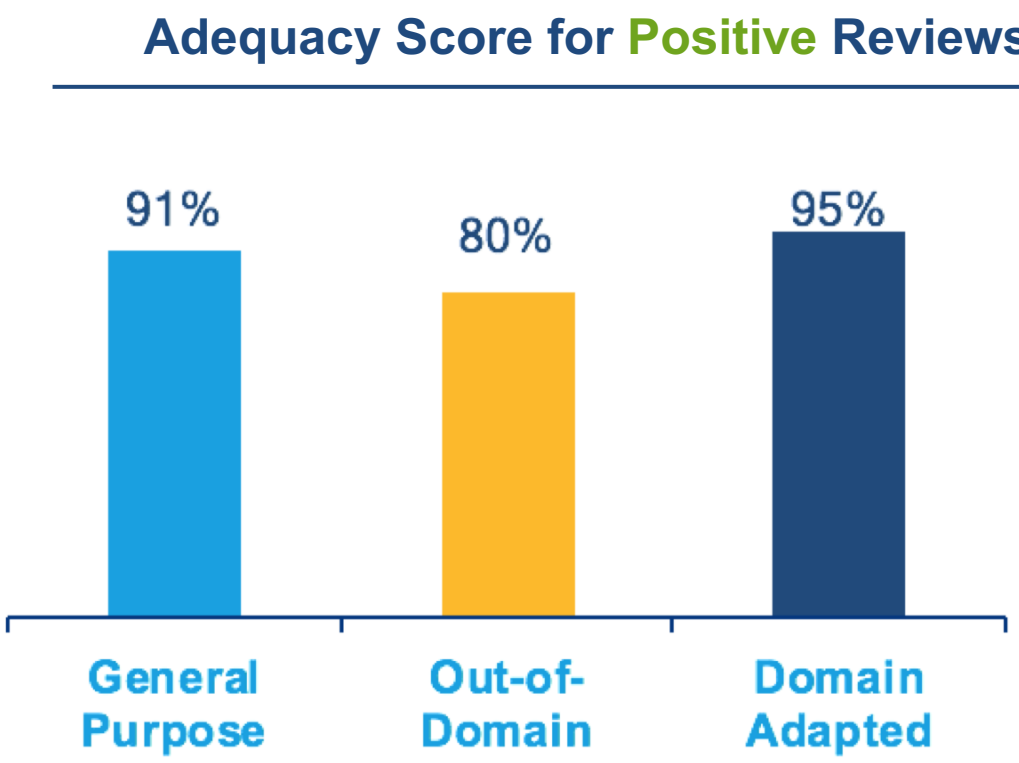


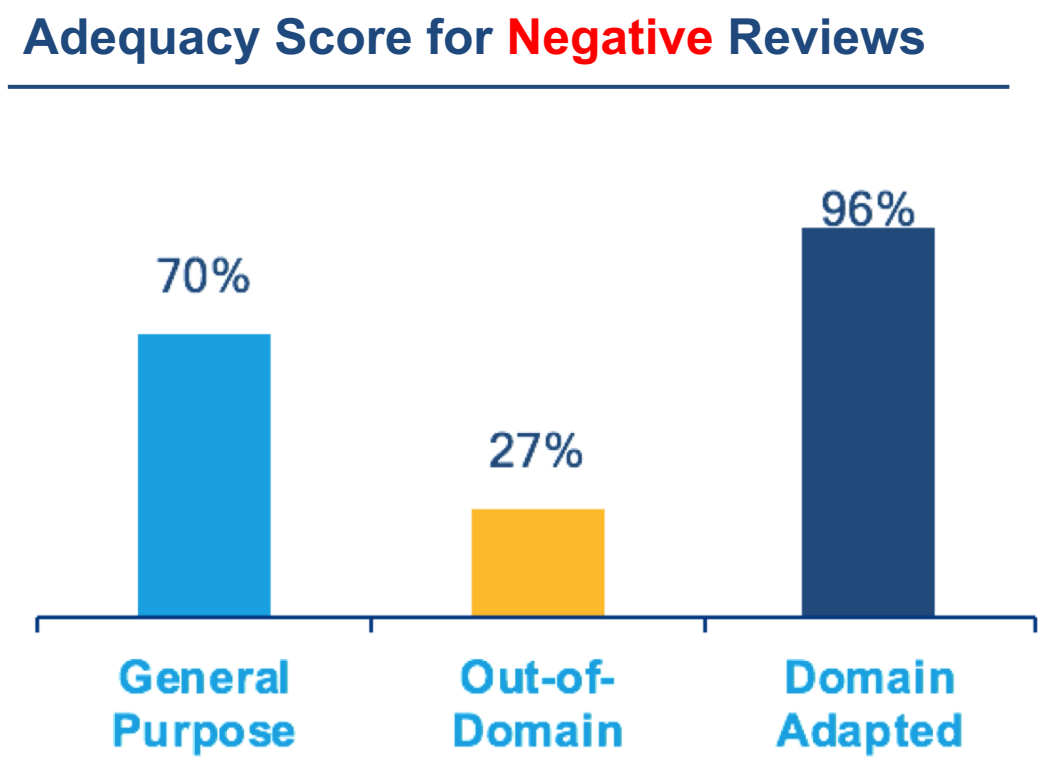
Gradual downsampling vs fine tuning

Gradual downsampling	Fine tuning
Faster iteration	Takes time to get the General Model trained
Trained for specific use case from the beginning	Can be adapted to multiple use cases
Applicable without In-domain parallel data	Needs In-domain parallel data
Less accurate	More accurate

No answer yet

Human Evaluation Results for Domain Adapted Model to translate customer reviews (gradual downsampling)





Want to know more?

Machine Translation at Booking.com: Journey and Lessons Learned
EAMT (User Track)
Prague, May 2017
Best Paper Award

Toward a full-scale neural machine translation in production: the Booking.com use case
MT Summit XVI (Commercial Track)
Nagoya, Sep 2017

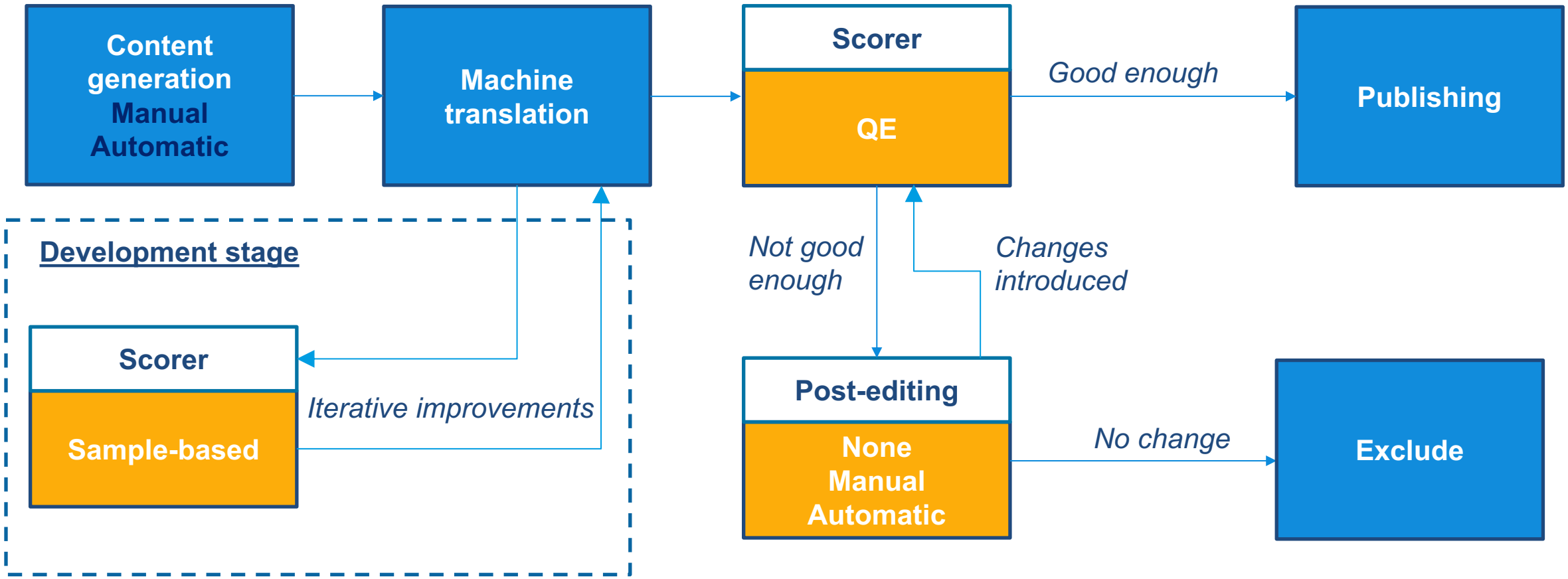
Automatic post-editing and Quality Estimation

What is the business rationale?

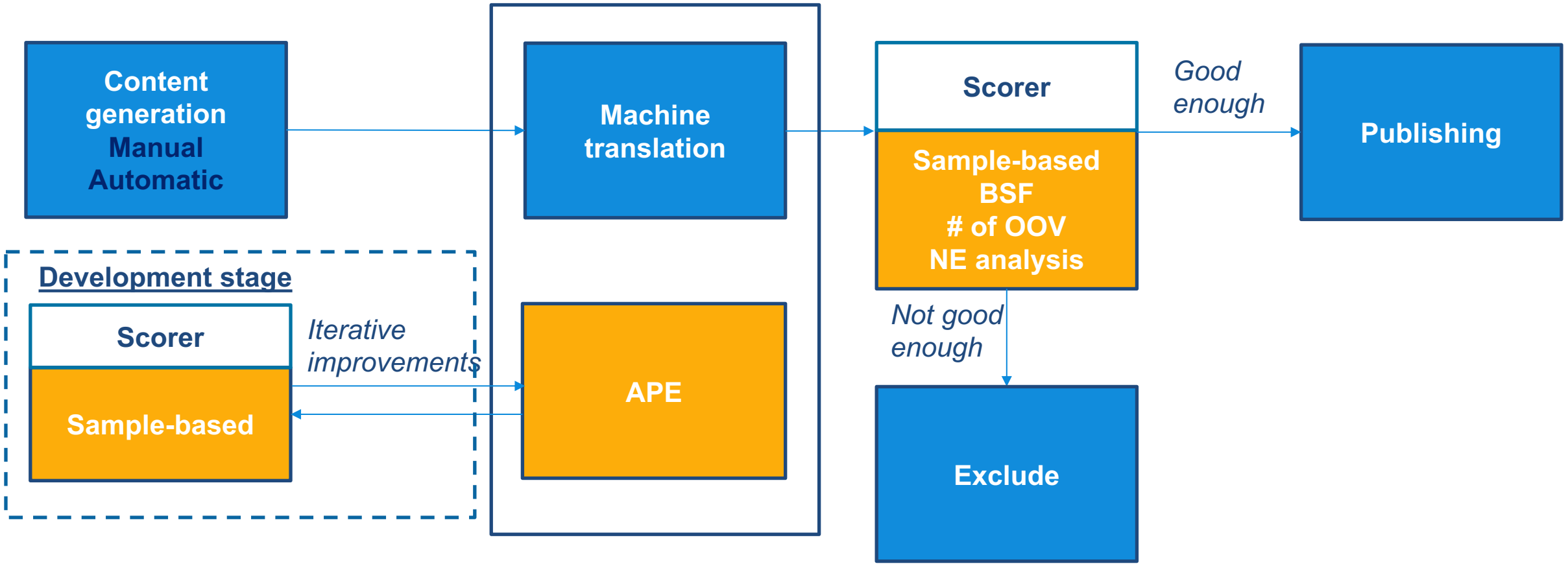
- **The Whys:**

  - Reduce monetary and legal risks
  - Increase user trust
  - Increase traction with partners and customers (B2B and B2C)
  - As a part of the better integrated MT system, improve user experience

Complete MT-QE-APE architecture

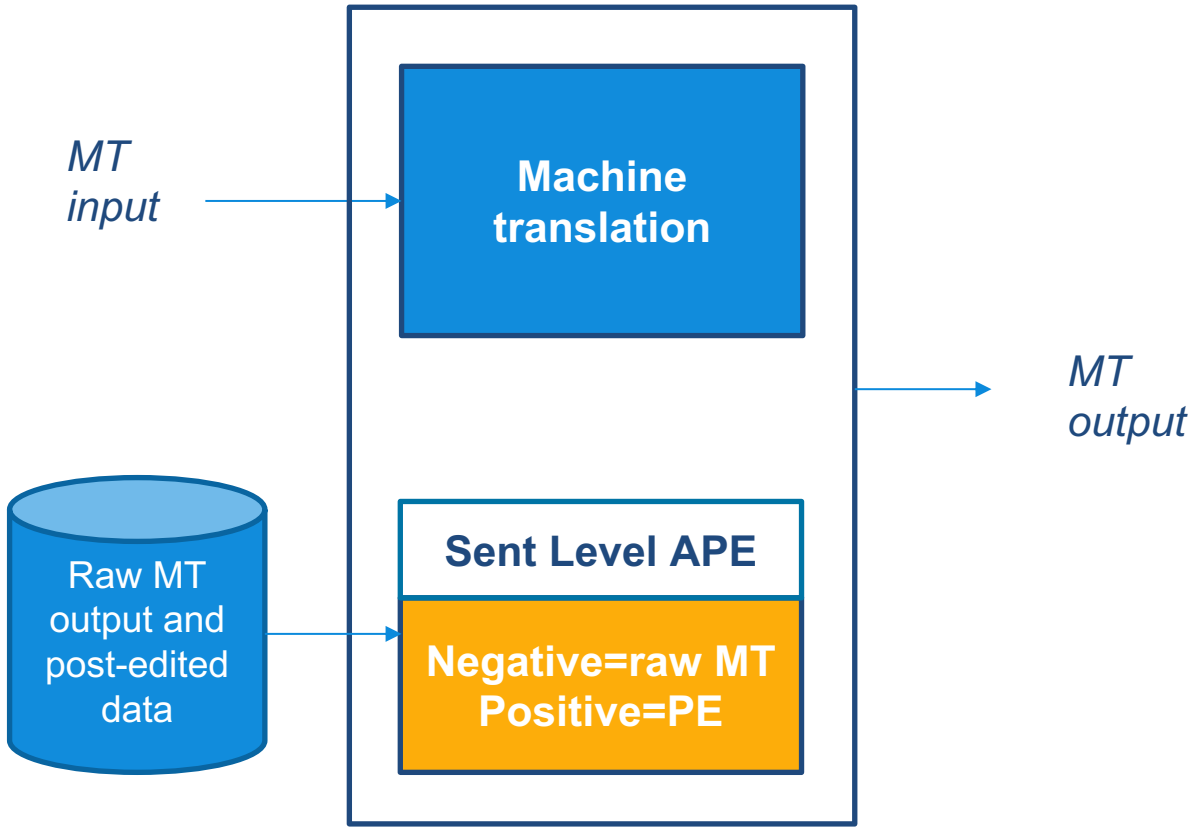


How can we validate?



How can we design an APE system, which would address the most important problems?

Sentence level APE



Credit: MT research group at the University of Edinburgh

Negative and Positive training examples

Source

Offering a restaurant with WiFi, Hodor Ecolodge is located in Winterfell. On-site **parking is free**.

Raw MT

Die Hodor Ecolodge in Winterfell bietet ein Restaurant mit WLAN. **Parkplatz vor Ort ist verfügbar**.

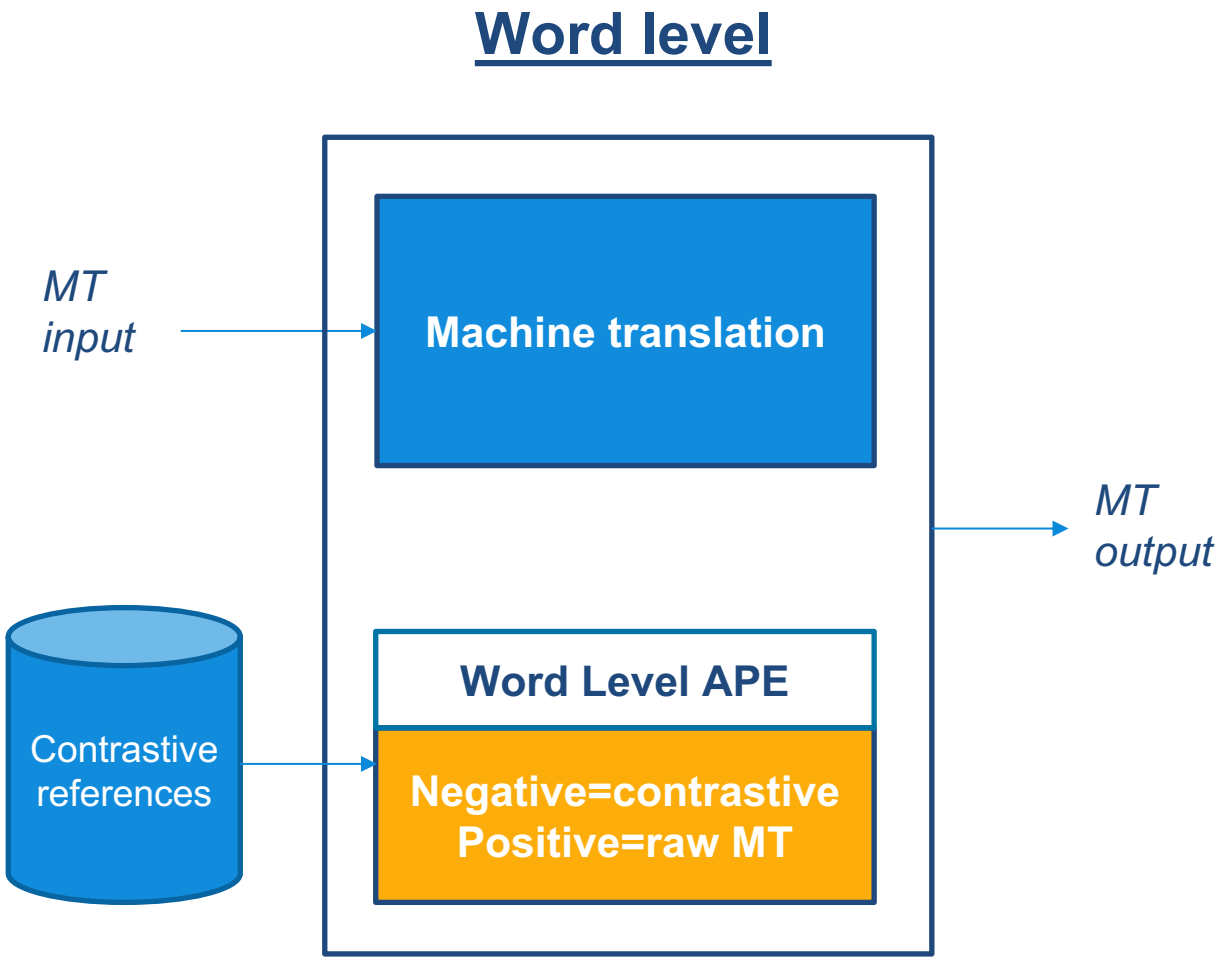
Negative example

Post-edited MT

Die Hodor Ecolodge in Winterfell bietet ein Restaurant mit WLAN. **Parkplatz vor Ort ist kostenlos**.

Positive example

How can we design an APE system, which would address the most important problems?



Credit: MT research group at the University of Edinburgh

Contrastive references

Source

On-site **parking is free**.

Translation

Parkplatz vor Ort ist verfügbar.

Positive example

Contrastive

Parkplatz vor Ort ist nicht verfügbar
or
Parkplatz vor Ort ist kostenlos.

Negative example

Future Directions (applied research and technology)

Explore alternative NMT technologies

- "Transformer" by (Vaswani et al., 2017)

Ensure high quality of translations

- Named Entities
- NMT with reconstruction (Tu et al., 2017)
- Optimization for UGC
- Conditioning MT output on structured data

Reinforcement learning (Nguyen et al., 2017)

TAUS
the language data network

TAUS MT Survey 2018

http://info.taus.net/taus-mt-survey-2018

Deadline: Friday, April 14th

Thank You

Questions?

Maxim Khalilov

maxim.khalilov@booking.com
www.linkedin.com/nl/maximkhalilov