

16th Biennial Conference of the Association for Machine Translation in the Americas ([AMTA](#))

Presentations - Users & Providers Track

PREDICT Methodology

Machine Translation (MT) Eligibility Criteria

Paula Manzur (She/Her)

September 30th – October 2nd 2024

Chicago, Illinois, USA



Machine Translation State

Language Tier	Impact Level 1 (Reviews)		Impact Level 1 (others)		Impact Level 2		Impact Level 3		Impact Level 4		Impact Level 5		TOTAL	
	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%	Count	%
Tier 1	22 B	32.4%	7 B	11.1%	171.7 M	0.3%	1.6 M	0.0%	20.8 M	0.0%	2.8 M	0.0%	29.3 B	43.8%
Tier 2	26 B	38.8%	3 B	5.2%	59.1 M	0.1%	0.9 M	0.0%	17.7 M	0.0%	3.3 M	0.0%	29.6 B	44.1%
Tier 3	6 B	8.2%	3 B	3.8%	11.8 M	0.0%	0.6 M	0.0%	8.6 M	0.0%	1.7 M	0.0%	8.1 B	12.1%
TOTAL	53 B	79.4%	13.5 B	20.1%	243 M	0.4%	3 M	0.0%	47 M	0.1%	8 M	0.0%	67.0 B	100%

99.5 % = 66.5 Billions



0.5 % = 301 Millions

RAW MT

 Booking.com MT
  Google Translate

Localization Solution

Human Translation with varied degree of Translation Technology support

Localization Solution

Traditionally, the decision to use Machine Translation is driven by a **non-systematic** human approach, mainly based on **budget** and **speed** needs, with limited consideration for **MT content suitability**.

A content type is a set of **linguistic**^{*} and **business**^{**} characteristics that inform the selection of a **localization solution**.

* *register, terminology, sentence structure and complexity, formatting, etc.*

** *target audience, business impact, content lifespan, legal requirements, budget and resources, etc.*

Localizing every piece of content to the **right level of quality** for our customers enhances their **international experience.**

The “Right” Level of Quality: Definitions & Localization Solutions

Content Prioritization	Quality Level	Quality Level Definition	Applicable Error Categories	Suggested Localization Solution
Content Types	1	Factual information (including but not limited to any numeric value, currencies and location or personal names) is conveyed clearly without distortions, even if the phrasing is unnatural.	Accuracy	RAW MT
Content Types	2	Translation that conveys the meaning without any accuracy or fluency errors, uses inclusive language where relevant and maintains consistent terminology and local conventions, although it may not be stylistically perfect.	Accuracy, Fluency, Terminology, Verity, Locale convention, Markup	MTPE
Content Types	3	Translation that conveys the meaning, is free from any formal errors and is written in a polished style.	Accuracy, Fluency, Terminology, Style, Design, Locale convention, Verity	TEP
Content Types	4	Fully adapted and polished content that reads naturally and seamlessly to resonate with the target audience.	All Categories	TRANSCREATION

PREDICT is the **methodology** used within a **framework** called **Content Profiling** to analyze and prioritize **content types** based on **MT risks**.

The goal is to match **content type** with its optimal **localization solution** (**right level of quality**).

Risk Scale & Weights for Machine Translation Eligibility

Content Prioritization	Quality Level	Quality Level Definition	Applicable Error Categories	Suggested Localization Solution	MT Risk Scale	MT Risk Weights
Content Types	1	Factual information (including but not limited to any numeric value, currencies and location or personal names) is conveyed clearly without distortions, even if the phrasing is unnatural.	Accuracy	RAW MT	1 - 25	1
Content Types	2	Translation that conveys the meaning without any accuracy or fluency errors, uses inclusive language where relevant and maintains consistent terminology and local conventions, although it may not be stylistically perfect.	Accuracy, Fluency, Terminology, Verity, Locale convention, Markup	MTPE	26 - 50	2
Content Types	3	Translation that conveys the meaning, is free from any formal errors and is written in a polished style.	Accuracy, Fluency, Terminology, Style, Design, Locale convention, Verity	TEP	51 - 75	3
Content Types	4	Fully adapted and polished content that reads naturally and seamlessly to resonate with the target audience.	All Categories	TRANSCREATION	76 - 100	4

Binary Questions about the Source Content and Localization Use Case

Text Characteristics (20)

- Structure
- Style
- Intent

Terminology (3)

- Industry-Specific Jargon
- Inconsistencies
- Abbreviations/Acronyms

Legal & Regulatory (2)

- Confidentiality
- Regulatory-constrained

Target Audience (2)

- Internal use only
- Specific cultural or focus group & purpose

Project-Specific (2)

- Budget
- Turn-Around Time (TAT)

Field of Expertise (1)

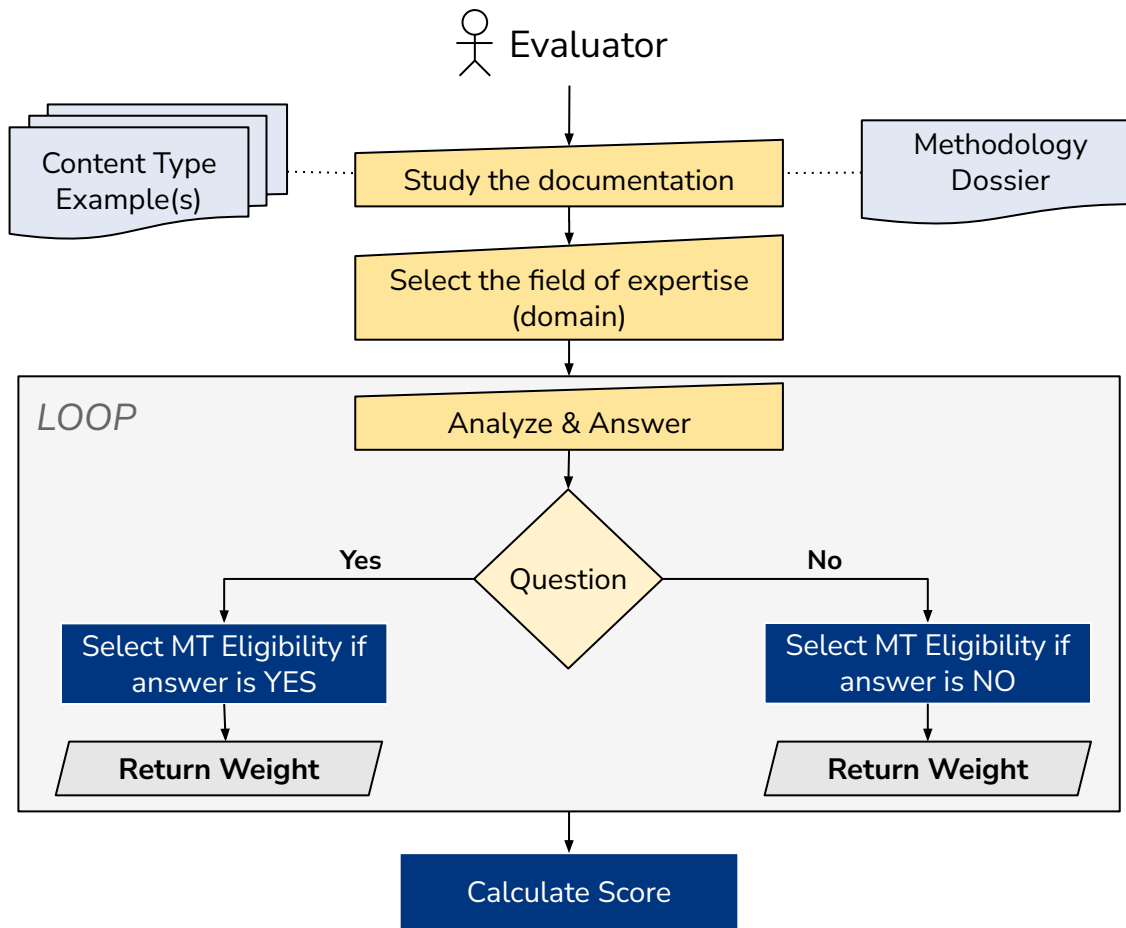
(Always True)

- Legal
- Marketing
- User Interface

The "Yes" or "No" answer to the questions helps decide if Machine Translation is risky (**At Risk**) or eligible (**Optimal**) for use.

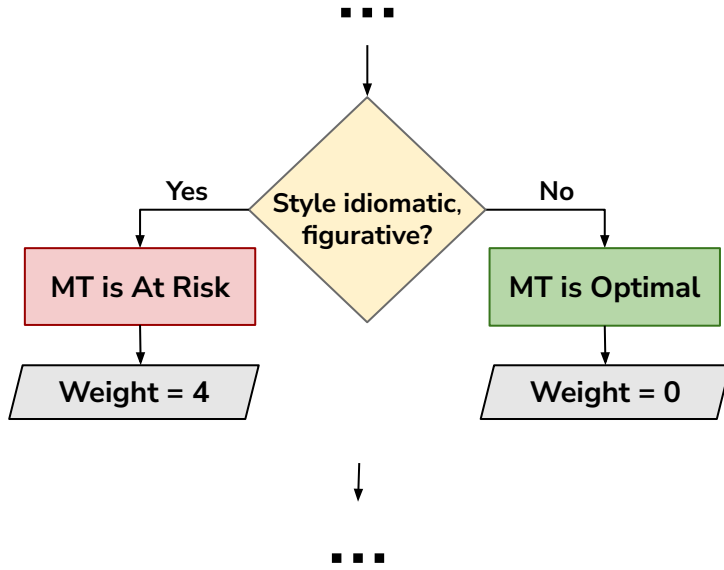
The **1 to 4 risk weights** (based on the Quality Levels) show how **critical** the risk might be.

How PREDICT Works In a Nutshell

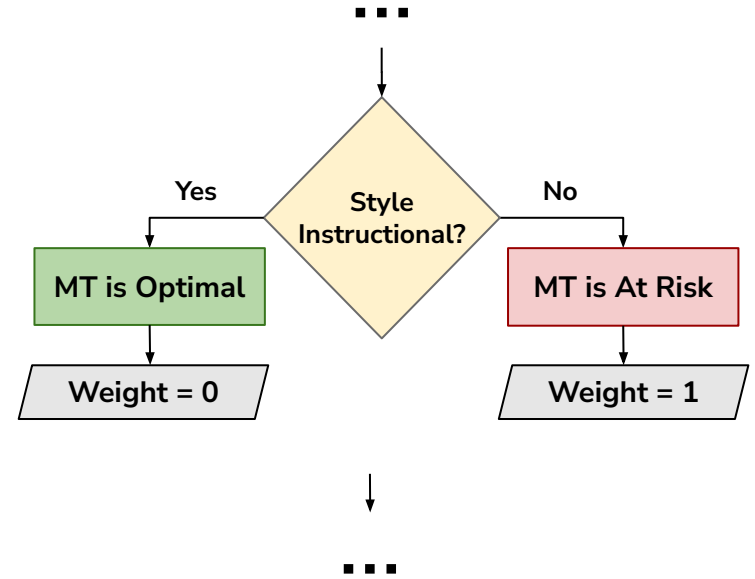


The risk depends on the question







Is the **style idiomatic**, figurative? (at least one expression)



Is the **style instructional** (precise directions & guidance)?



PREDICT Methodology Dossier

-  **Question** Source text is **idiomatic, figurative** (contains **at least one** expression)
 -  **Definition** An idiomatic or figurative writing style employs expressions or **phrases that hold specific meanings within a language, dialect, or culture**. These phrases **convey ideas or emotions uniquely**, relying on **cultural or contextual** significance to convey a deeper, more **nuanced meaning than their literal interpretation** might suggest.
 -  **General Example** **It's raining cats and dogs!**
 -  **Specific Example** Calling all adventurous travellers: **this is right up your alley**.
 -  **Additional Resource** Grammarly: [What Is an Idiom? Definition and Examples](#)
-
-  **Content Type Example(s)** **References (text, visuals, mockups, etc.)** for **evaluators** when answering the questions.

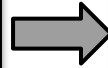
The “Weights Dictionary”

Questions <i>Source Text Structure</i>	Risk Weight	TRUE		FALSE		MT Eligibility Explanation
		MT Eligibility is	Weighted Impact	MT Eligibility is	Weighted Impact	
Well-written/grammatically sound (i.e. adheres to correct grammar and language conventions)	1	Optimal	0	At Risk	1	Machine translation excels when the source text has proper grammar and language structure. It relies on consistent grammar to produce accurate translations, and well-written content facilitates clearer understanding for machine translation systems.
High level of sentence complexity (i.e. long sentences with multiple clauses or dependent constructions)	4	At Risk	4	Optimal	0	Long, intricate sentences challenge accurate parsing and meaning conveyance. Errors and inaccuracies are likely. Machine translation prefers simple, consistent language. Additionally, if the text is overly wordy, machine translation systems may struggle to identify the most important information and may produce a translation that is similarly verbose and difficult to read.
Dialogue, Messaging Format	2	Optimal	0	At Risk	2	Dialogue or messaging format tends to be straightforward, aiding machine translation. Conversational style often follows common language patterns, facilitating accurate rendering for gist purposes mainly. Some challenges: ambiguity: when the intended meaning of a message can be unclear without additional context. Specialized Vocabulary: dialogue or messaging involving domain-specific terms (e.g., medical, legal).
Tables, Graphics, Charts (at least 1)	4	At Risk	4	Optimal	0	For the user to comprehend the main idea of the message (get the gist)The questions are designed to predict MT performance based on text characteristics Instructional (precise directions & guidance) Tables, Charts, Graphics, Footers, Headers, and Footnotes can affect the quality of machine translation output because they contain information that is context-dependent, domain-specific, and often essential to the overall meaning of a document. The nature of these elements may hinder accurate translation if not adequately handled by the system.
Headers, Subheaders, Footers, Footnotes (at least 4)	4	At Risk	4	Optimal	0	Similar to complex elements such as tables and charts, these components might contain critical context outside the main text body. Overlooking these parts could lead to a loss of essential information in the translation.
Placeholders, Tags, Links, HTML Markup Text (at least 2)	4	At Risk	4	Optimal	0	Machine translation systems might struggle with placeholders, tags, or HTML markup text, potentially causing errors in format or structure. This can especially affect productivity when dealing with Unicode Plurals in languages requiring additional code elements beyond the source language.
Restricted Character Count Format, Social Media, Ads Format	4	At Risk	4	Optimal	0	Texts constrained by character count or specific formats might lack contextual cues, leading to ambiguity in translation. The brevity sacrifices nuances, idiomatic expressions, and cultural references necessary for accurate translation, particularly challenging when languages differ in verbosity. Maintaining accuracy while adapting to character constraints becomes an added difficulty for machine translation in such cases.

A score is calculated

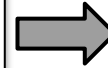
Maximum Weights

Question weights are multiplied by either 1 (MT is at risk) or 0 (optimal conditions for MT), and the cumulative weighted impact on MT Eligibility is computed.



Impacted Weight

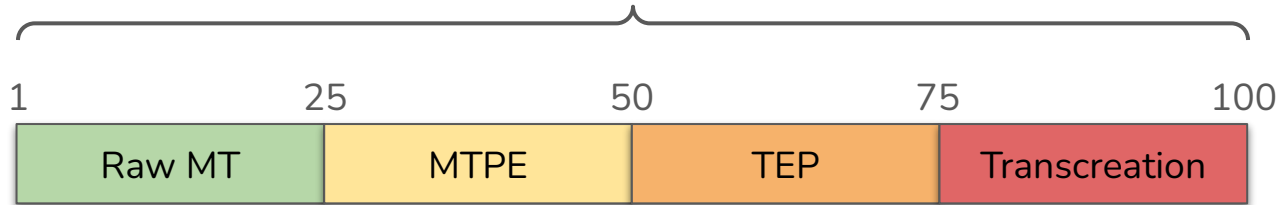
This weighted impact is divided by the maximum conceivable score, that is, the total score achievable if all questions were answered as "At Risk".



Internal Use

If the answer to "Only internally within the Company (and will not be published)?" is YES, the weighted impact is divided into half, taking it closer to Optimal MT Eligibility.

$$\text{Final Score} = \frac{\text{Impacted Weight}}{\text{Maximum possible risk to MT}}$$



Case Study With 1 Evaluator


Evaluated 10 content types to test the methodology effectiveness




Content Type	Current Localization Solution
Marketing PR Articles	TEP
Marketing Editorial Articles	TEP
Marketing (Partners)	TEP
Attractions Descriptions	MTPE
Scaled Property Descriptions	RAW MT, MTPE, AutoGen
Legal Documents (Internal)	TEP
Informative Communication (Legal Emails)	TEP
Informative Communication (Non-Legal Emails)	TEP
Guest Reviews (UGC)	RAW MT
User Research (Surveys)	TEP

Case Study Example: Scaled Property Descriptions

Carrer de Montgomit, 6, 03560 La Venteta, Spain – [Show map](#)



Wonderful location ⓘ



Entire apartment Swimming pool Balcony Free Wifi View

Air conditioning Private pool Terrace Kitchen Washing machine

Apartmento playa Muchavista is located in La Venteta, just 1.8 miles from Carrer La Mar Beach and 3.5 miles from Alicante Golf Club. Guests staying at this apartment have access to a balcony. Free Wifi is available throughout the property and Muchavista Beach is a 2-minute walk away.

With direct access to a terrace, this air-conditioned apartment comes with 1 separate bedroom and a fully equipped kitchen. A flat-screen TV is featured. The accommodation is non-smoking.

Alicante Train Station is 8.4 miles from the apartment, while Terra Natura is 22 miles away. The nearest airport is Alicante–Elche Miguel Hernández Airport, 19 miles from Apartamento playa Muchavista.

Distance in property description is calculated using © OpenStreetMap

Most popular facilities

- Swimming pool
- Free Wifi

Apartmento playa Muchavista is located in La Venteta, just 1.8 miles from Carrer La Mar Beach and 3.5 miles from Alicante Golf Club. Guests staying at this apartment have access to a balcony. Free Wifi is available throughout the property and Muchavista Beach is a 2-minute walk away.

With direct access to a terrace, this air-conditioned apartment comes with 1 separate bedroom and a fully equipped kitchen. A flat-screen TV is featured. The accommodation is non-smoking.

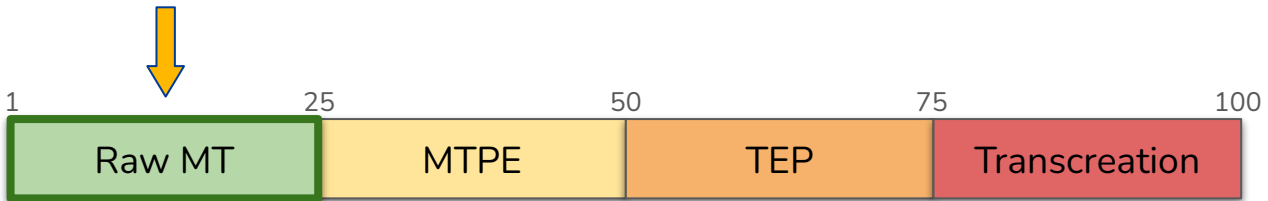
Alicante Train Station is 8.4 miles from the apartment, while Terra Natura is 22 miles away. The nearest airport is Alicante–Elche Miguel Hernández Airport, 19 miles from Apartamento playa Muchavista.

Question	Answer (YES/NO)	MT Eligibility is...	Weight	Weighted Impact
Structure				
Well-written/grammatically sound (i.e. adheres to correct grammar and language conventions)	YES	Optimal	1	0
High level of sentence complexity (i.e. long sentences with multiple clauses or dependent constructions)	NO	Optimal	4	0
Dialogue, Messaging Format	NO	At Risk	2	2
Tables, Graphics, Charts (at least 1)	YES	At Risk	4	4
Headers, Subheaders, Footers, Footnotes (at least 4)	YES	At Risk	3	3
Placeholders, Tags, Links, HTML Markup Text (at least 2)	NO	Optimal	4	0
Restricted Character Count Format, Social Media, Ads Format	NO	Optimal	4	0
Style				
Creative (inventive language)	NO	Optimal	4	0
Descriptive (detailed specifics of a product or item)	YES	Optimal	2	0
Instructional (precise directions & guidance)	NO	At Risk	1	1
Declaratory, informative (straightforward, factual)	YES	Optimal	1	0
Idiomatic, figurative writing (at least one idiomatic or figurative expression)	NO	Optimal	4	0
Nuanced (intricate layers of meaning, cultural subtleties, connotations)	NO	Optimal	4	0
Informal writing (conversational tone, colloquialisms, a more "relaxed" approach to grammar rules)	YES	At Risk	3	3
Formal writing (more sophisticated vocabulary, impersonal tone, priority is adherence to grammar rules)	NO	At Risk	2	2

Question	Answer (YES/NO)	MT Eligibility is...	Weight	Weighted Impact
Intent				
Convince, persuade the user to take action or behave in a certain way (e.g. convince the user to subscribe to a newsletter)	NO	Optimal	4	0
Interact; engage with the user (e.g. shopping cart on website)	NO	At Risk	3	3
Entertain; ignite inspiration, motivation or enthusiasm in the user (e.g. marketing case study)	NO	Optimal	4	0
For the user to comprehend the main idea of the message (get the gist)	YES	Optimal	1	0
Provide user with information (e.g. a product description in retail)	YES	Optimal	2	0
Domain & Subdomain				
What's the content field of expertise?	Scaled Product Descriptions	At Risk	2	2
Terminology				
Use of industry-specific jargon (specialized terminology, expressions, or language unique to a particular field, i.e. patents, dish/cuisine names (at least 3 specialized terms)	NO	Optimal	4	0
Use of abbreviations and/or acronyms (at least 3)	NO	Optimal	4	0
Inconsistent use of terminology, contradictory word choices and expressions (at least 2)	NO	Optimal	4	0

Question	Answer (YES/NO)	MT Eligibility is...	Weight	Weighted Impact
Legal and Regulatory Requirements				
Highly sensitive, confidential	NO	Optimal	4	0
Regulatory-constrained (i.e. patents)	NO	Optimal	4	0
Target Audience				
Only internally within the Company (and will not be published)	NO	At Risk	1	1
Amongst a culturally-specific or focus group (i.e. teenagers, kids, parents, students, etc.) with a specific purpose	NO	Optimal	4	0
Project-Specific				
Needs the translation with a fast turn-around time to meet with specific deadlines (e.g. TAT shorter than standard delivery SLA for the volumes to be translated)	YES	Optimal	1	0
Has budget concerns	YES	Optimal	1	0
			86	21

$$Score = \frac{\text{Impacted Weight}}{\text{Maximum possible risk to MT}} = \frac{21}{86} = 0,2442 * 100 = 24,42\%$$



Case Study Results

Content Type	Current Localization Solution	PREDICT Recommended Localization Solution	Score
Marketing PR Articles	TEP	TEP	56.8%
Marketing Editorial Articles	TEP	TEP	53.4%
Marketing (Partners)	TEP	TEP	56.8%
Attractions Descriptions	MTPE	RAW MT	19.8%
Scaled Property Descriptions	RAW MT, MTPE, AutoGen	RAW MT	24.4%
Legal Documents (Internal)	TEP	RAW MT	23.3%
Informative Communication (Legal Emails)	TEP	MTPE	31.8%
Informative Communication (Non-Legal Emails)	TEP	MTPE	29.5%
Guest Reviews (UGC)	RAW MT	RAW MT	22.1%
User Research (Surveys)	TEP	MTPE	36.8%

Access the Case Study (adapted version) here: paulamanzur.github.io

From a Case Study to Real Application

Operationalization



PREDICT in Action

Currently evaluating approximately **116 content types** (existent and new) with this methodology. By end of October 2024, a total of **580 responses** is expected (**5 evaluators per content type**).



Analytics New Formula

50% weight is assigned to the **domain** of the content type (e.g. Marketing) and the rest **50%** is based on **evaluators weighted opinion** to understand the **risk to MT**.



Preliminary Results

While a high percentage of content types are a **good fit for MT**, most require **some level of human touch**. Thus, MT is no universal solution, we still need **diverse localization solutions**.

The localization team will use **PREDICT Methodology recommendations** to make a **data-driven decision** on the **final localization solution** for each **content type**.

Takeaways

The localization & MT industry are assessing MT suitability in different ways, with no unified system in place.

With the PREDICT Methodology, we've made a simple yet important attempt at a more robust, detailed, and systematic approach to MT eligibility criteria.

This is not the final answer but a starting point to reopen the conversation within the MT community about the need for common best practices that empower people to make informed, data-driven decisions.

Thank you
More info in the appendix.

Questions?

[Paula Manzur](#)

Appendix

Some Considerations

01

Human-Centered Approach

It is recommended that language and localization experts answer the questionnaire. The more evaluators are involved in the methodology, the more objective the evaluation process is.

02

Dynamic Aspect

Parameters, such as the questions, examples, domains and associated risk weights can be adjusted dynamically according to different Customer needs.

03

Using PREDICT's Dossier to prompt LLMs

Some experiments have been done with ChatGPT as an evaluator. Preliminary conclusions show that humans are more reliable. More tests are needed.

04

Prediction Validation

If the methodology recommends Full MTPE for one content type and A/B testing is used to compare this solution versus RAW MT, for example, this can help validate or reject the prediction from a customer/business perspective.

Additional Information

Why was this methodology developed?

The creation of this methodology was part of a strategic localization initiative at Booking.com. It introduced a more robust, detailed, and systematic approach for the objective evaluation of diverse content types using multiple evaluators.

Adoption

The Methodology aims to facilitate informed decision-making, aiding stakeholders in determining the optimal localization solution based on MT risk assessment.

Industry Experts: Validation & Insightful Advice

- [Mikolaj Szaina](#), Localization Manager at Booking.com for his support and trust to initiate and guide this endeavor.
- Arle Lommel, Don de Palma and Alison Toon from [CSA Research](#) for providing insightful feedback.
- Adjunct Professor [Jon Ritzdorf](#) and Director of Operations at [machinetranslate.org](#) [Cecilia Yalangozian](#) for validating the methodology as a valuable asset to the localization industry.

Evaluation & Cadence

“Evaluators” actively provide their insight by answering the questionnaire. The advised number is 3 or 5 (preferably an odd number). Stakeholders and language professionals who are familiar with the content can participate in the questionnaire.

1. **Exploratory** evaluations for new content types
2. **Periodic** ones for validation (recommendation is once a year) depending on the business context

Access the Methodology via the Case Study

A slightly different version of the case study is shared [open source](#) as a contribution from Booking.com’s localization team, aiming to bring value for both suppliers and buyers in the localization industry.