DO ANY OF THESE APPLY TO YOU?

❑ Internal AI solution

❑ AI goals

❑ Quality concerns or complaints

❑ Want to use AI but don't know how

❑ ...
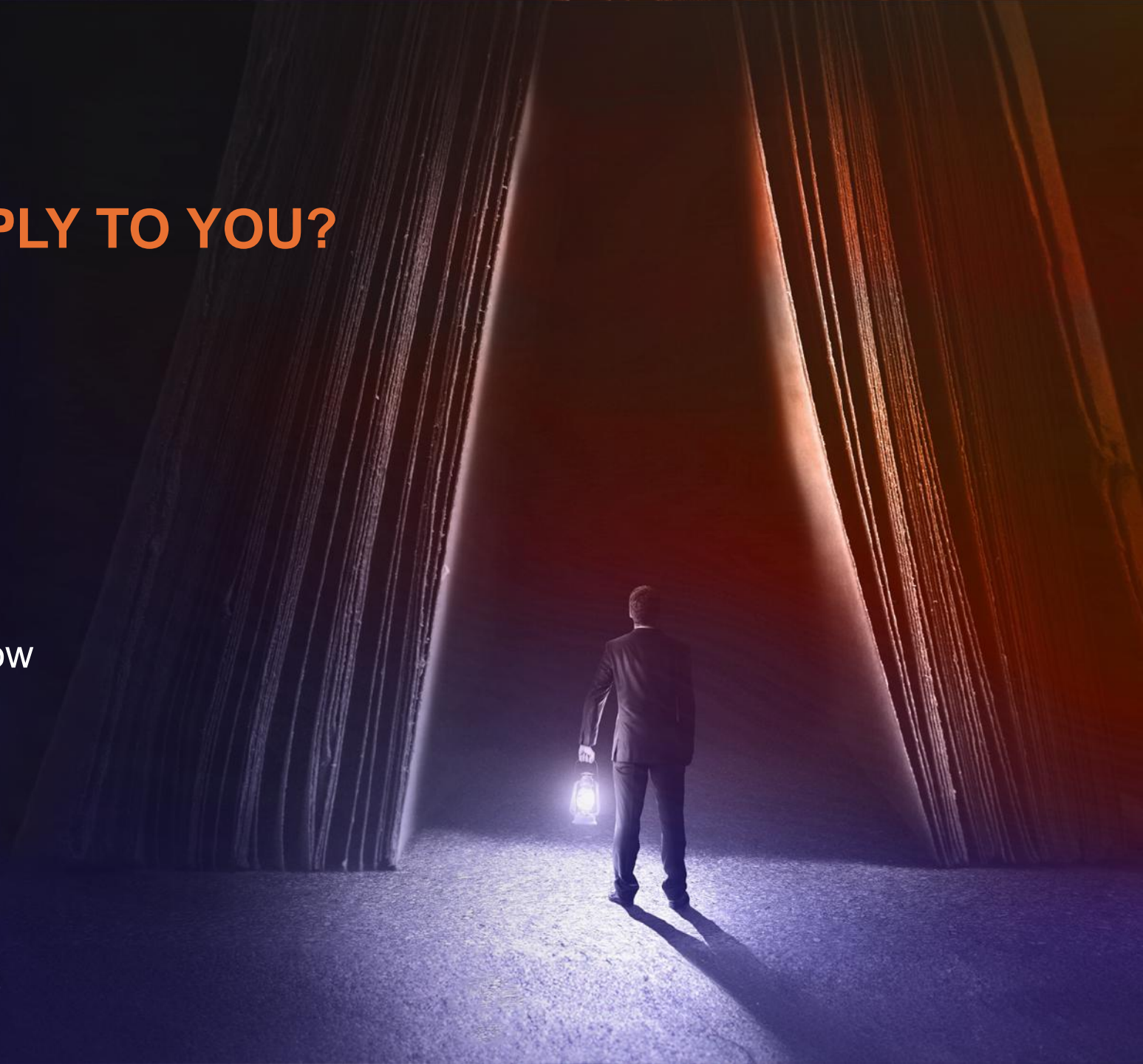
Facts & Figures

More than **80 percent** of AI projects fail (RAND)

42 percent of companies are abandoning them (S&P Global Markets)

However, with the right approach, you can achieve AI success.

WHAT OUR CUSTOMERS ARE EXPERIENCING

USE CASE 1

Solution

Customer with large trained LLM used for translation

Issues

- Very low quality
- High degree of hallucinations
- Misinterpretation or literal translations e.g. brand names translated
- Insertions of product numbers
- Not able to deal with measurements changes from imperial to metric.
- Lack of consistency with glossaries

USE CASE 2

Solution

Testing LLM usage for large scale translations or review of translations

Issues

- Effort to improve output very time consuming
- Lacking needed in-language prompting expertise
- Not able to deploy at scale – lack of system infrastructure

USE CASE 3

Solution

Trained ChatGPT for content creation and translation

Issues

- Quality issues
- Use cases for internal customers unclear
- Internal customers do not have process to evaluate output
- Lack of policy around usage
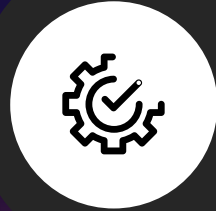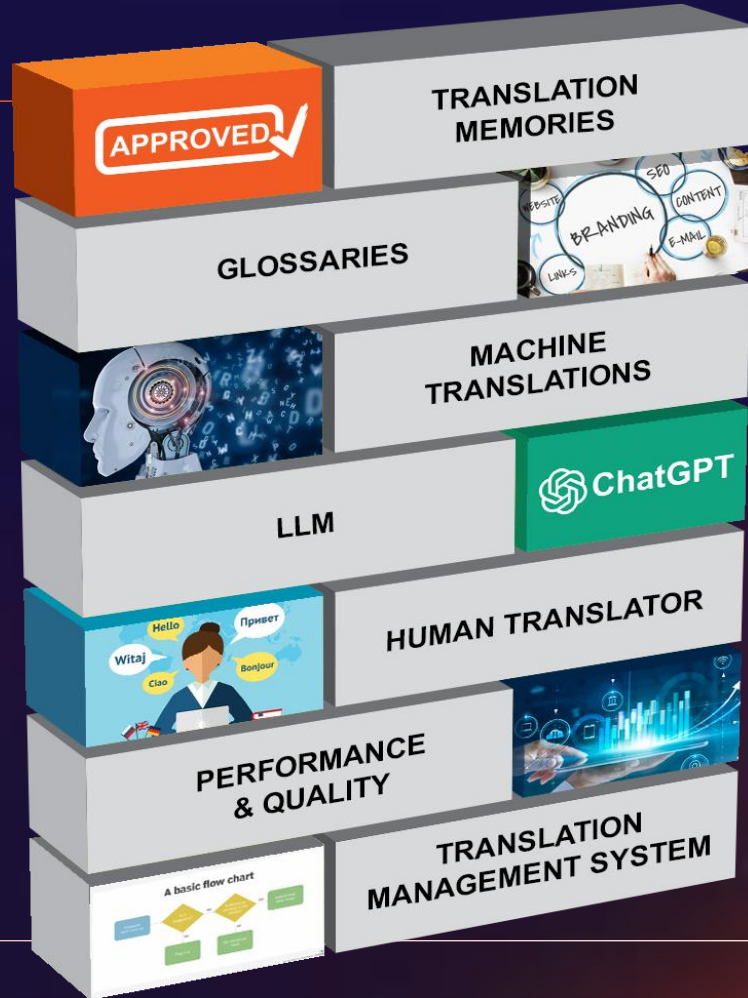- Lack of integration of linguistic assets – glossaries and TMs

WHY IS IT SO CHALLENGING TO GET IT RIGHT?

- IS CHATGPT GOOD ENOUGH?

- DO YOU NEED THE SAME QUALITY FOR ALL CONTENT?

LOCALIZATION **PROCESS**

TRANSLATION MEMORIES

APPROVED ✓

GLOSSARIES

MACHINE TRANSLATIONS

LLM ChatGPT

HUMAN TRANSLATOR

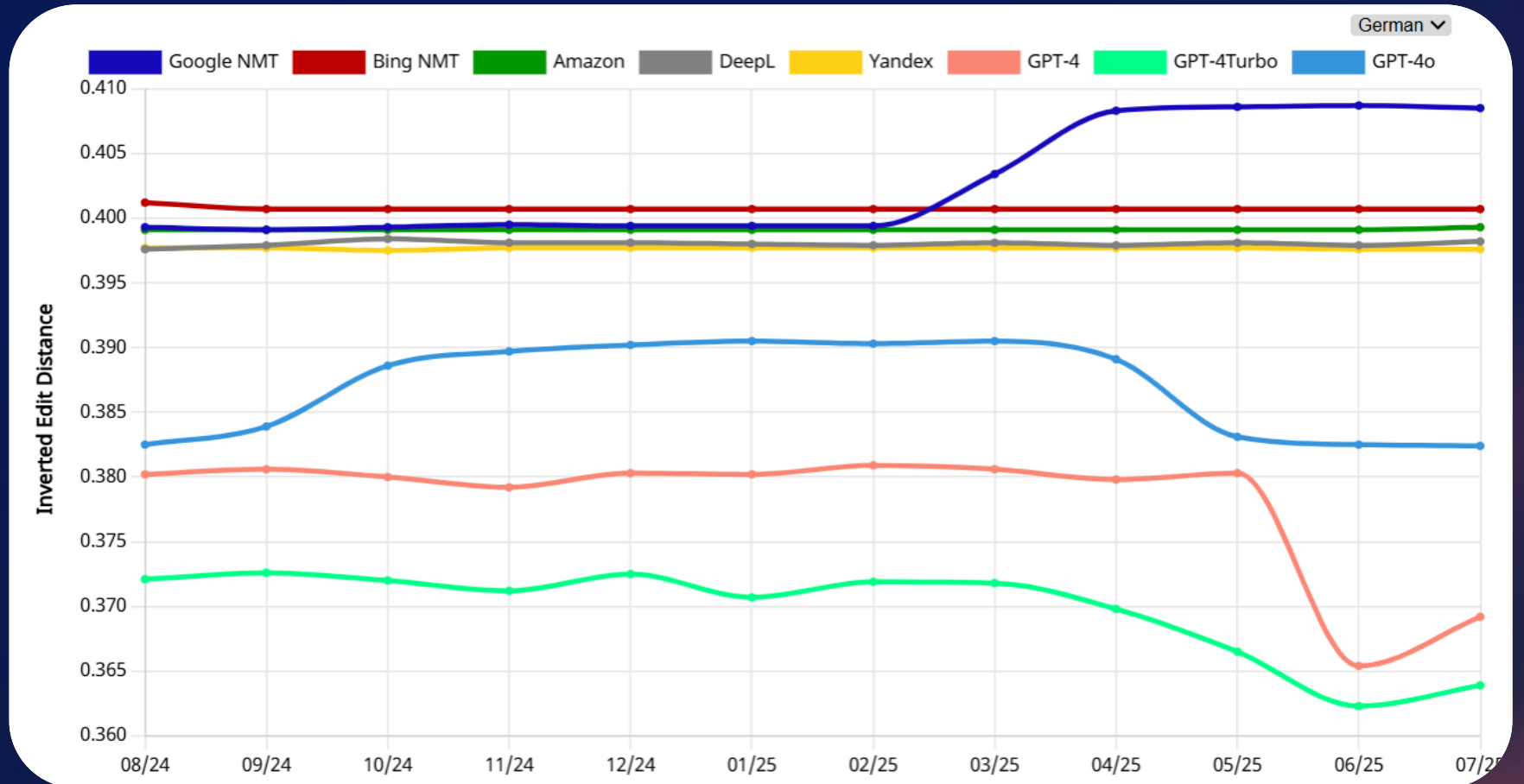PERFORMANCE & QUALITY

TRANSLATION MANAGEMENT SYSTEM

Proven building blocks to ensure an efficient localization process

Together these blocks orchestrate and deliver optimal results
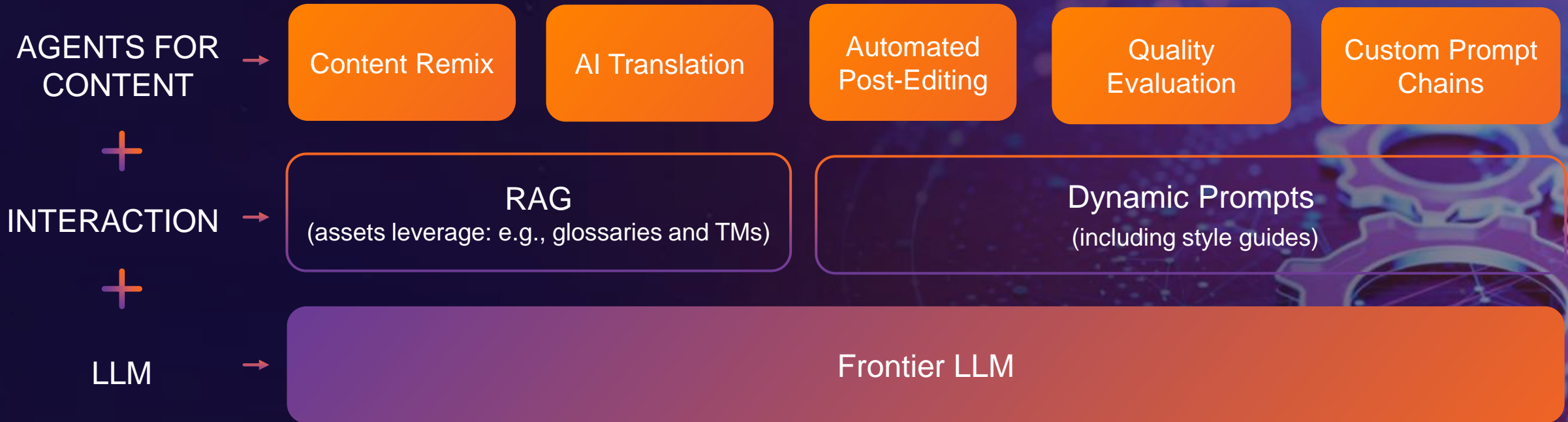
LIONBRIDGE MT TRACKER

For translation out of the box LLMs are continuing to perform worse than Neural Machine Translation systems

AI-FIRST AT LIONBRIDGE

LIONBRIDGE aurora AI™

An advanced platform combining market leading technology with your requirments and advanced agents to optimize outcomes

AGENTS FOR CONTENT →

| Content Remix | AI Translation | Automated Post-Editing | Quality Evaluation | Custom Prompt Chains |

+

INTERACTION →

RAG
(assets leverage: e.g., glossaries and TMs)

Dynamic Prompts
(including style guides)
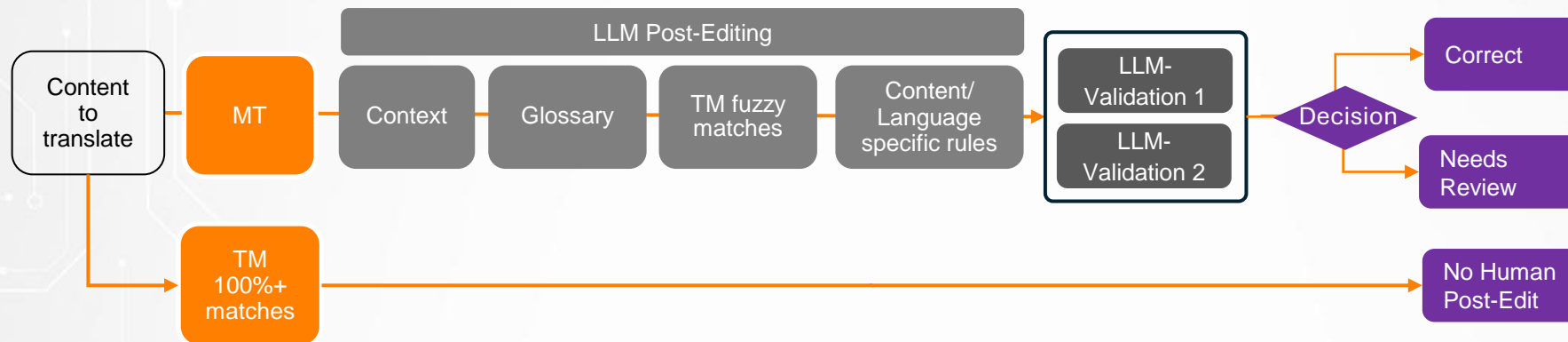
+

LLM →

Frontier LLM

REDUCING HUMAN EFFORT WITH
A CUSTOMIZABLE AI TRANSLATION PROCESS

Our Advanced Agentic Localization Method

- Using best MT-engine for translation and frontier LLMs for AI decisioning and editing to eliminate the need to train models
- LLM decisioning based on the content
- Integrating language assets (TM, glossaries, and style guides)
- Ability to customize prompts for your tone of voice and brand
- Fuzzy Match LLM Post-Editing
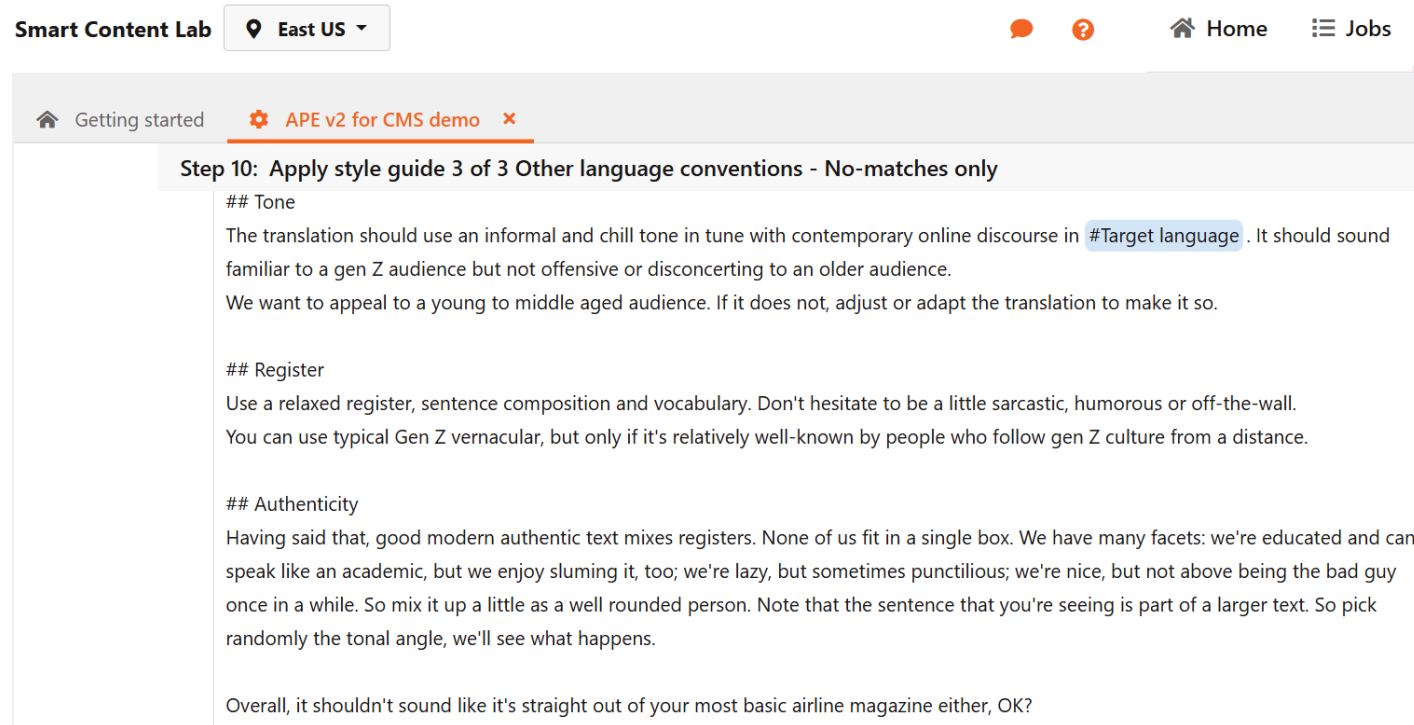- Advanced content validation based on human reasoning

LLM Post-Editing

Group 1: Prepare content and metadata (Steps 1–4)	⌄
Group 2: Core APE (Steps 5–8)	⌃
Step 5: Analyze and perform cleanup	⌄
Step 6: Evaluate translation options	⌄
Step 7: Analyze and fix glossary issues	⌄
Step 8: Apply style guide	⌄
Group 3: TM-match auto-editing (Steps 9–10)	⌄
Group 4: Validation and locking (Steps 11–15)	⌄

LLM Post-Editing

Content to translate → MT → Context → Glossary → TM fuzzy matches → Content/ Language specific rules → LLM-Validation 1 / LLM-Validation 2 → Decision → Correct / Needs Review

TM 100%+ matches → No Human Post-Edit

A CUSTOMIZABLE AI TRANSLATION PROCESS

PROMPTING FOR PARTICULAR CONTENT AND LANGUAGE RULES

example of a custom prompt to reflect stylistic preferences for all translations of that travel website

Smart Content Lab   📍 East US ▾   💬 ❓   🏠 Home   ☰ Jobs

🏠 Getting started   ⚙ APE v2 for CMS demo ✕

Step 10: Apply style guide 3 of 3 Other language conventions - No-matches only

Tone
The translation should use an informal and chill tone in tune with contemporary online discourse in #Target language . It should sound familiar to a gen Z audience but not offensive or disconcerting to an older audience.
We want to appeal to a young to middle aged audience. If it does not, adjust or adapt the translation to make it so.
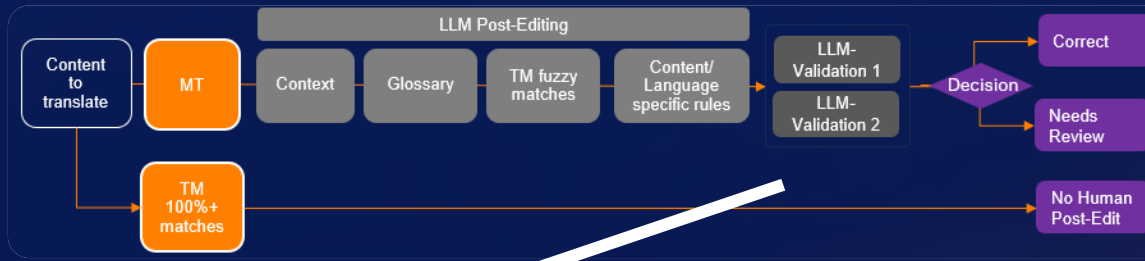
Register
Use a relaxed register, sentence composition and vocabulary. Don't hesitate to be a little sarcastic, humorous or off-the-wall.
You can use typical Gen Z vernacular, but only if it's relatively well-known by people who follow gen Z culture from a distance.

Authenticity
Having said that, good modern authentic text mixes registers. None of us fit in a single box. We have many facets: we're educated and can speak like an academic, but we enjoy sluming it, too; we're lazy, but sometimes punctilious; we're nice, but not above being the bad guy once in a while. So mix it up a little as a well rounded person. Note that the sentence that you're seeing is part of a larger text. So pick randomly the tonal angle, we'll see what happens.
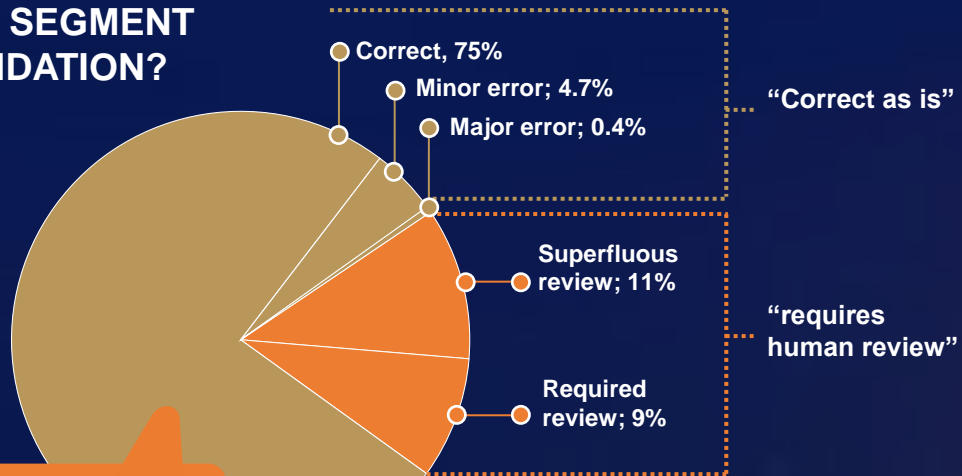
Overall, it shouldn't sound like it's straight out of your most basic airline magazine either, OK?

LIONBRIDGE
aurora™

RESULTS DASHBOARD

LLM Post-Editing

Content to translate → MT → Context → Glossary → TM fuzzy matches → Content/Language specific rules → LLM-Validation 1 / LLM-Validation 2 → Decision → Correct / Needs Review

TM 100%+ matches → No Human Post-Edit

HOW ACCURATE IS THE SEGMENT VALIDATION?

- Correct, 75%
- Minor error; 4.7%
- Major error; 0.4%

"Correct as is"

- Superfluous review; 11%
- Required review; 9%

"requires human review"

Evaluation of LLM decisions:
4.7% minor error
0.4 % major error

SAMPLE DATA OF LLM-ASSISTED POST-EDITING

Languages: German, French, Italian, Japanese, Polish, Simplified Chinese, Dutch, Spanish (Spain), Korean, Traditional Chinese, Portuguese (Brazil), Hungarian, Portuguese (Portugal)

Legend: Excellent / Very Good / Useable

"Very Good" results across all languages.*
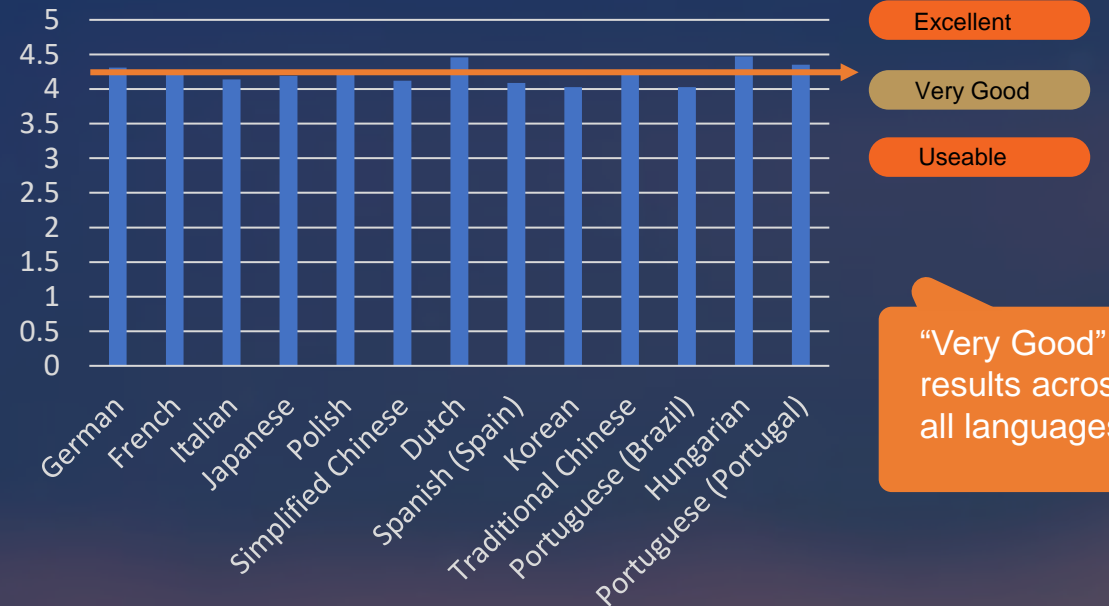
*source language English. Note that source languages German, French and Spanish perform at the same level.

LIONBRIDGE aurora AI

WHY IS IT SO CHALLENGING TO GET IT RIGHT?

- IS CHATGPT GOOD ENOUGH?

- DO YOU NEED THE SAME APPROACH AND QUALITY FOR ALL CONTENT?

IS THIS **GOOD ENOUGH?**

After 100 days in office, Germany's new chancellor has mixed record on key promises

Can a repressive pro-Kremlin president reverse Georgia's lean towards Europe?

US-China dispute over tiny AI chip highlights king-size geopolitical dilemma

Hero, television star, autocrat? Volodymyr Zelenskyy, a master of reinvention, is stuck in a dead end

excerpts of AI-generated translation of leading newspaper

MARKETING BROCHURE

BLOG POST

INTERNAL COMMUNICATION

ANNUAL REPORT

LANDING PAGE

LEGALLY BINDING CONTRACT

MEDICAL DEVICE INSTRUCTIONS FOR HOSPITAL BED

MEDICAL DEVICE INSTRUCTIONS FOR PACE MAKER

USER INTERFACE

SOFTWARE RELEASE NOTES

...

THE REACH FRAMEWORK

FOR CONTENT RELEVANCE

ROI

The investment in your global content matches expected returns and delivers business outcomes.

ENGAGEMENT

The content is specifically adapted to engage your global targets on specific goals.

AUDIENCE

The content resonates with each of your audiences in global markets on a cultural, community and demographic level.

CONTROL

Automated gathering of engagement and content quality data drives decision-making on content specifications.

HUMAN-IN-THE-LOOP

The AI-first process is controlled, curated and informed by human oversight and insights to optimize every step.

APPLYING THE REACH FRAMEWORK

What **type** of content is in scope? What is it **for**? **How much** of it is there?

R E A C H		
	ROI	How much business value is tied to this content?
	ROI	Is speed or quality the main driver of the value of the content?
	Engagement	How specialized is the content?
	Engagement	What kind of impact is expected of the content?
	Audience	What kind of stakeholders is the content for?
	Audience	How expert is the audience in the subject matter?
	Audience	What type of people are the target?
	Control	How much technical constraints need to be complied with?
	Control	How much quality control needs to be done?

GETTING IT RIGHT

- UNDERSTANDING YOUR LLM OUTPUT

- IDENTIFYING ITS STRENGTH AND WEAKNESES

- MAPPING ITS PERFORMANCE TO YOUR CONTENT

Evaluate 50,000 Words **— On Us**

Automated Assessment of your LLM translations of **up to 50,000 words.**
Discover why your LLM-generated translations are failing internal quality reviews — and how to improve them.
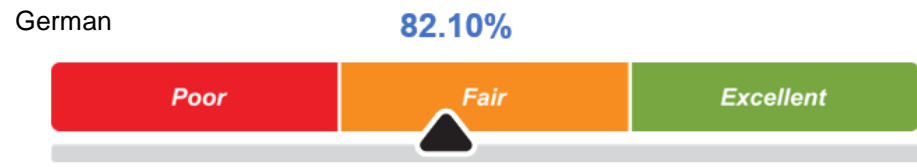
What we need:
- Bilingual source and target documents
- Style guides, glossaries, and other reference material
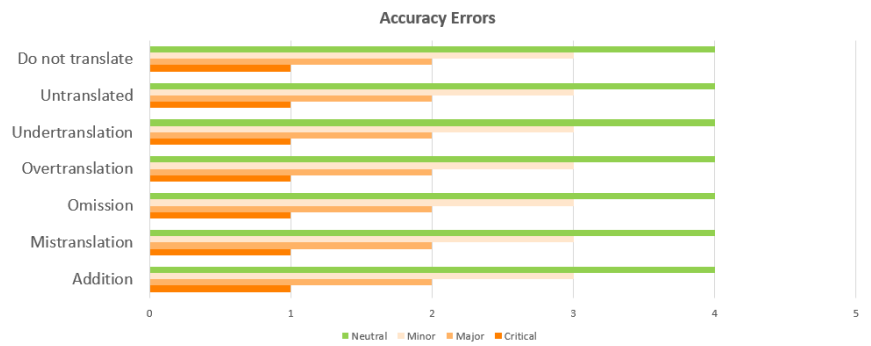- A description of the quality issues

What you'll get:
- **Objective quality score** — a data-driven evaluation of translation output
- **Comprehensive scorecard** — a breakdown of performance across key quality dimensions

Offer valid until Oct 31st, 2025

SCORECARD SAMPLE

LIONBRIDGE

BENCHMARKING TRANSLATION QUALITY: EVALUATION OF LLM-GENERATED CONTENT

LLM-Evaluation | A Lionbridge Solution

German — **82.10%**

| Poor | Fair | Excellent |

SCORECARD
ERROR CATEGORIZATION | ACCURACY

Accuracy Errors

- Do not translate
- Untranslated
- Undertranslation
- Overtranslation
- Omission
- Mistranslation
- Addition
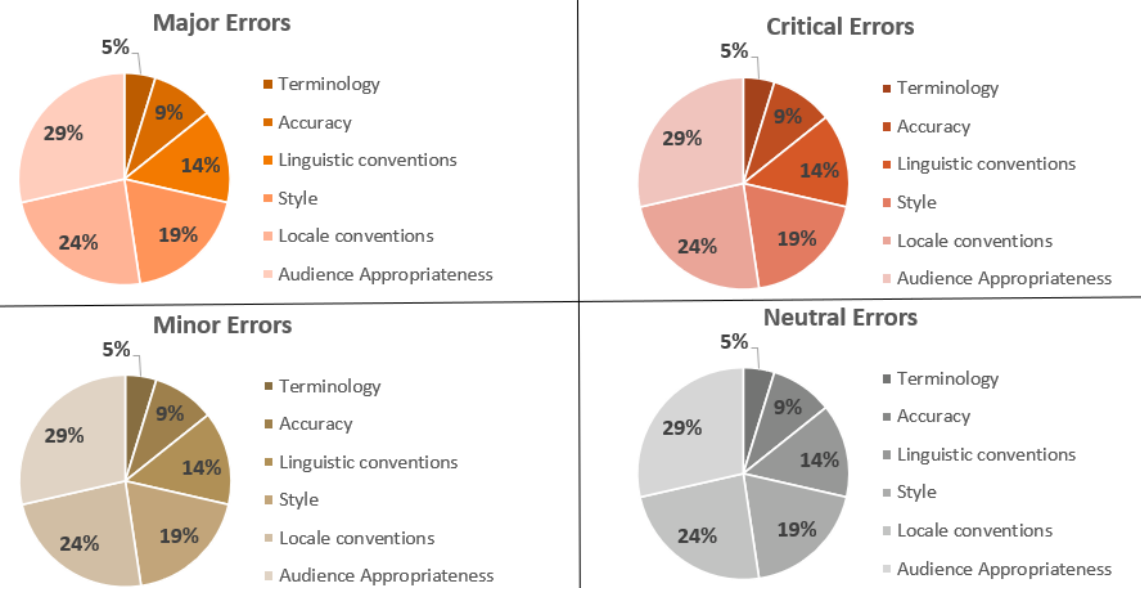
Legend: Neutral, Minor, Major, Critical

The error categories used **in Auto-LQA assessments are:**

- ✓ Terminology
- ✓ Accuracy
- ✓ Linguistic conventions
- ✓ Style
- ✓ Locale conventions
- ✓ Audience appropriateness

HIGH-LEVEL FINDINGS
QUALITY ASSESSMENT | SEVERITY LEVEL BY CATEGORY

Major Errors
5%, 9%, 14%, 19%, 24%, 29%
- Terminology
- Accuracy
- Linguistic conventions
- Style
- Locale conventions
- Audience Appropriateness

Critical Errors
5%, 9%, 14%, 19%, 24%, 29%
- Terminology
- Accuracy
- Linguistic conventions
- Style
- Locale conventions
- Audience Appropriateness

Minor Errors
5%, 9%, 14%, 19%, 24%, 29%
- Terminology
- Accuracy
- Linguistic conventions
- Style
- Locale conventions
- Audience Appropriateness

Neutral Errors
5%, 9%, 14%, 19%, 24%, 29%
- Terminology
- Accuracy
- Linguistic conventions
- Style
- Locale conventions
- Audience Appropriateness

LIONBRIDGE aurora AI™

LIONBRIDGE®

aurora™

THANK YOU

We break barriers and build bridges to new opportunities. Everywhere.