Using Large Language Models for Translation

SMARTLING

Contents

As a leader in AI translation, Smartling has translated millions of words using the world's leading Large Language Models (LLMs). Our AI Research and Development team is constantly experimenting, measuring, and deploying different LLMs to drive efficiency and improve quality. Now, we want to share some of those findings with you.

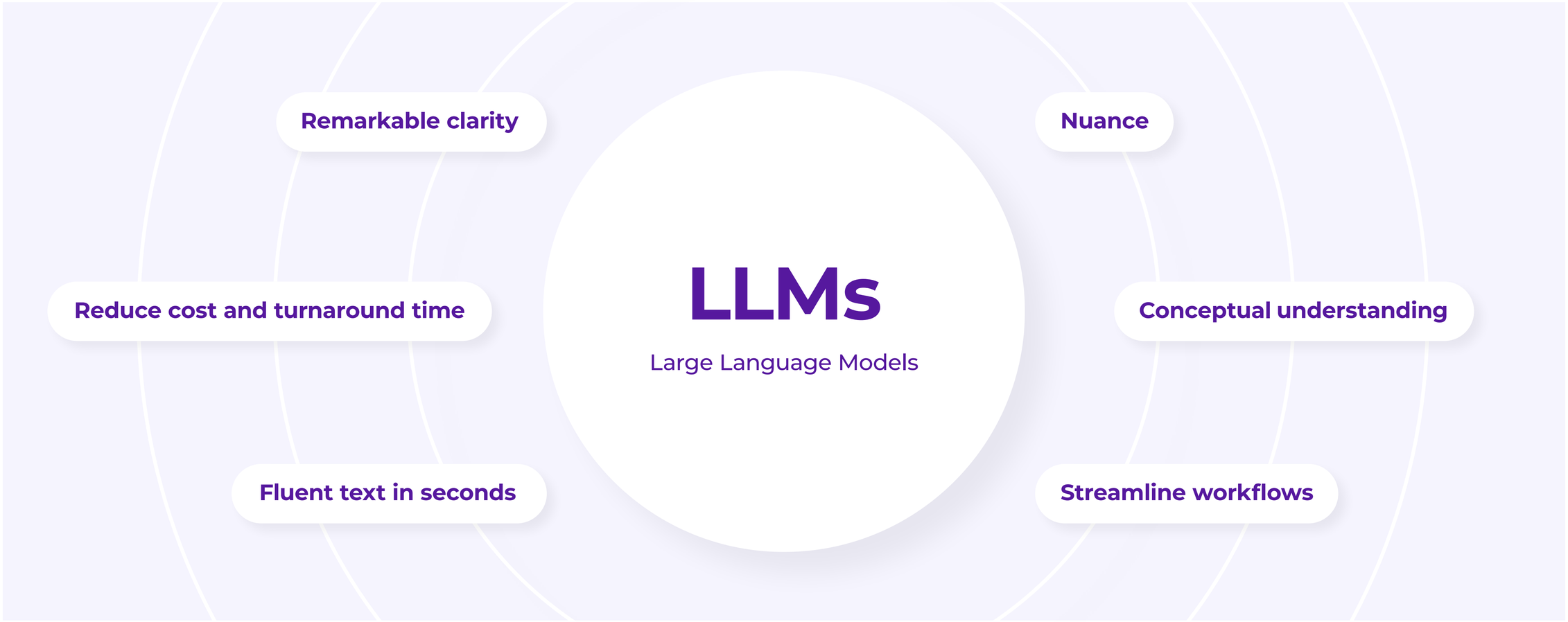
Large Language Models have caused a profound transformation in the translation industry. These deep learning models can interpret text with remarkable clarity, nuance, and increasing levels of conceptual understanding. Moreover, since LLMs produce conversationally fluent text in mere seconds, they enable companies to streamline workflows, slash turnaround times, and significantly reduce costs for translation services.

In essence, leveraging LLMs means businesses can achieve not only language translation but also cultural localization at scale and more cost-efficiently than ever before.

In essence, leveraging LLMs means businesses can achieve not only language translation but also cultural localization at scale and more cost-efficiently than ever before.

However, while LLMs have the potential to deliver enormous gains, hastily inserting them into a translation workflow can do more harm than good. Amid the excitement about how these AI models can help companies do more with less in the translation space, there is a great deal of confusion about what LLMs can and can't do and how best to leverage them.

This guide explains the benefits and limitations of LLMs. It then dives into how companies can incorporate LLMs alongside machine translation (MT) to produce content tailored to the intended audience that is linguistically accurate and culturally appropriate. Finally, we share some best practices for those looking to get the most out of this exciting technology.


Remarkable clarity
Nuance
Reduce cost and turnaround time
LLMs
Large Language Models
Conceptual understanding
Fluent text in seconds
Streamline workflows

Benefits and limitations of LLMs in translation

There are many benefits to using LLMs in translation that extend across different content types and domains. Further, LLMs are evolving rapidly, and their capabilities will undoubtedly continue to improve.

Nevertheless, this technology still has several limitations for translations because LLMs are trained for general language understanding (i.e. generating the most probable next token/ word) rather than specifically for the translation task. Therefore, overcoming these limitations to produce high-quality translations requires oversight.

Here are a few strengths and weaknesses of today's LLMs to consider when assessing how best to use them in translation.

BENEFIT #1

LLMs can help ensure preferred terminology is applied.

Context included in a prompt can facilitate this. When translating for the candy company Mars, for instance, the prompt to the LLM could include the fact that Mars is a candy company and should be localized accordingly.

Alternatively, when the LLM is used in conjunction with MT, it can correct mistranslations. To use the example above, the MT engine may have erroneously translated “Mars” as if it were the planet. If the LLM is provided with a glossary entry that specifies that “Mars” should not be translated, or that it should be translated as the proper name of the company, it can make that adjustment.

MARS

Translation method	Prompt
LLM only	Candy Company
MT only	Planet
MT and LLM	Candy Company
LLM and glossary	Mars should not be translated. It is the proper name of the company.

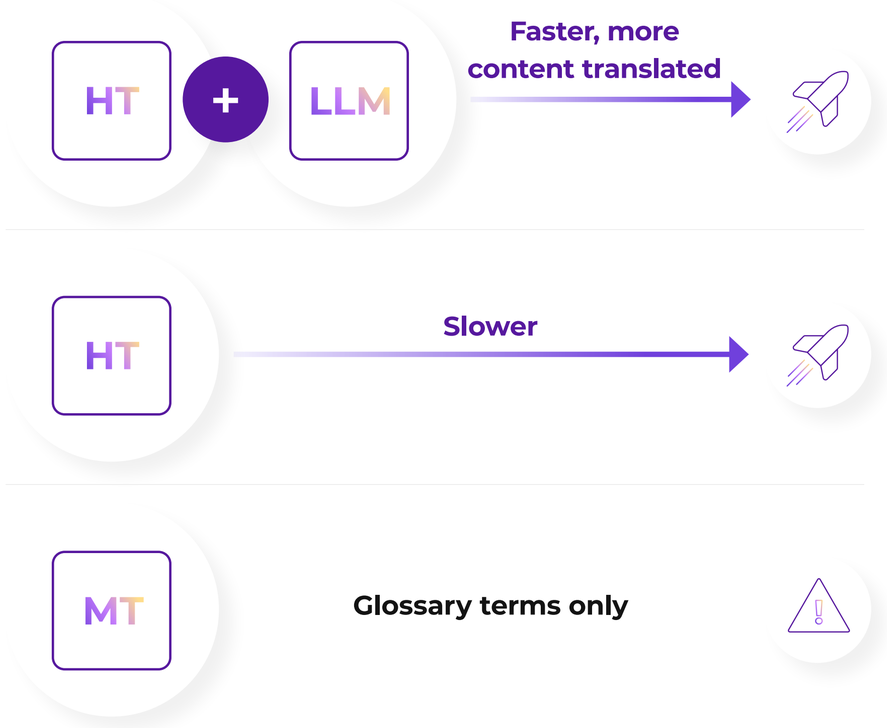
BENEFIT #2

LLMs can help align translations to a brand's unique voice.

Conventional MT engines are not adept at adjusting for stylistic preferences or levels of formality and incorporating context, humor, and idiomatic expressions. So previously, ensuring quality when translating creative content — such as marketing material or multimedia — required a human-only approach.

But LLMs excel at these areas, which means more content can go through AI-powered workflows with human oversight.

Translating brand voice workflows



BENEFIT #3

LLMs can apply appropriate industry-specific terminology.

LLMs have been trained to recognize subtleties and complexities in not only everyday language but also language used in technical content and specialized industries. For example, they can produce a solid first draft of a legal translation that experts can further refine and fact check.

BENEFIT #4

LLMs can help adapt content to fit cultural and regulatory landscapes.

With the right prompting, LLMs can adjust translated content to align with cultural norms, dialects (assuming such datasets are available), and regulations. An e-commerce company, for instance, might use an LLM to imbue product descriptions with local flair. Alternatively, a pharmaceutical company can use LLMs to ensure clinical trial data is translated and formatted according to local regulations.

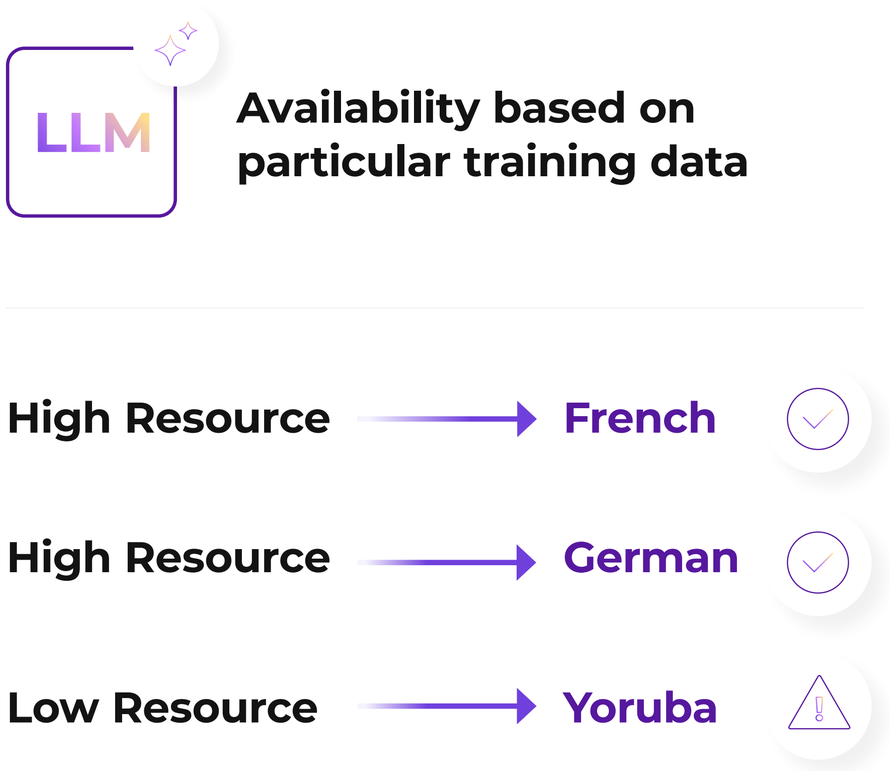
Although this doesn't eliminate the need for a human-driven review afterward for quality control, LLMs can do much of the initial heavy lifting.

LIMITATION #1

LLMs are not as advanced when applied to some languages.

LLMs are typically trained on vast quantities of text data. However, the quality of an LLM's output for a given language depends on how much of that data was written in that language. Extensive corpora are scarce or nonexistent for low-resource languages, which hampers the ability of LLMs to learn the nuances, idioms, and cultural aspects of these languages.

But even high-resource languages may run into trouble. For example, if an LLM's training data is only 10% Chinese, that LLM isn't likely to generate high-quality translations in Chinese. Or, if the training data only contains text from Ancient Chinese literature, it's unlikely to generate translations for e-commerce content that resonate with modern-day audiences.

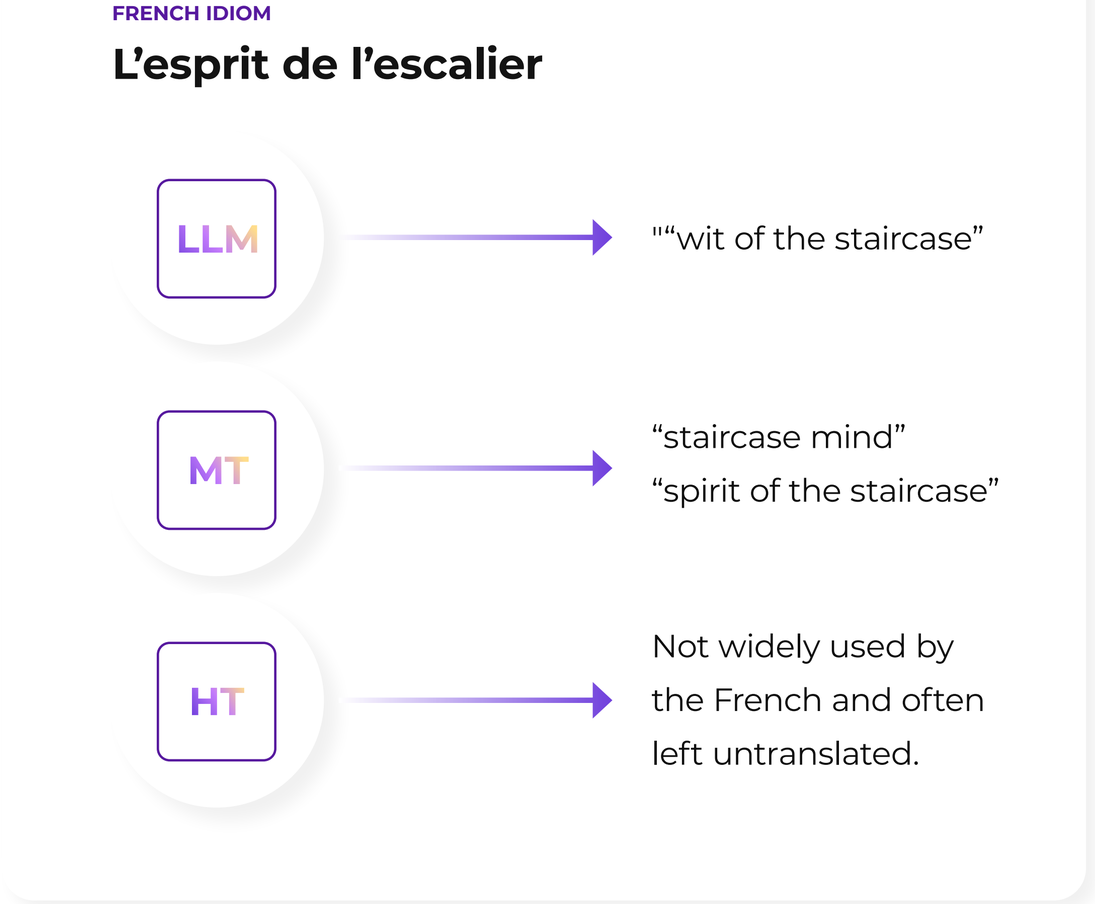


LIMITATION #2

LLM can exhibit US-centric biases.

These can be subtle. For instance, when Smartling's AI and Machine Translation team prompted an LLM to provide an example of a French idiom that is commonly mistranslated via MT, it offered "L'esprit de l'escalier." The LLM is likely to produce a smoother English translation — "wit of the staircase" — than an MT engine, which might render the idiom as "staircase mind" or "spirit of the staircase." But that misses the point.

"L'esprit de l'escalier" is not widely used by French people. It is mainly used in English — and often left untranslated. Businesses relying on LLMs for localizing content may encounter cultural oversights or misrepresentations like this one that could impact their brand's reputation and local consumer engagement.



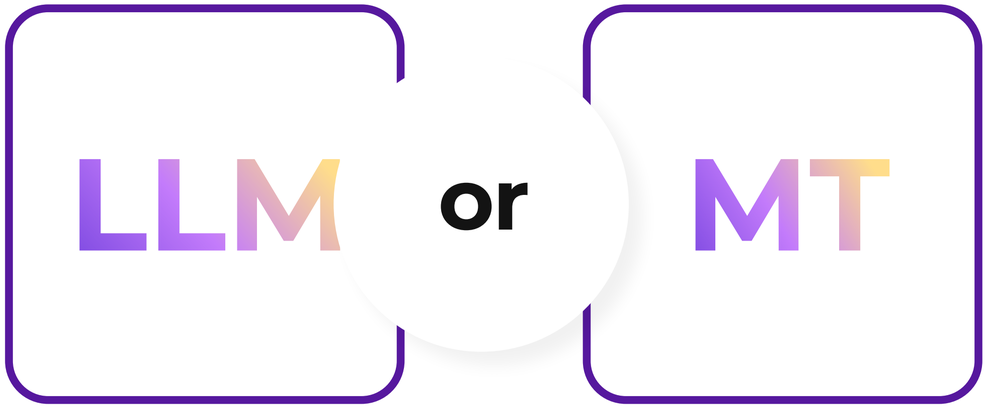
LIMITATION #3

LLMs can hallucinate.

LLMs generate text by predicting the next most probable word or phrase based on your instructions and the words or phrases that come before it. In essence, they provide an educated guess based on linguistic patterns. Therefore, LLMs are well suited to generate fluent but not always relevant or accurate translations.

LLMs or Machine Translation?

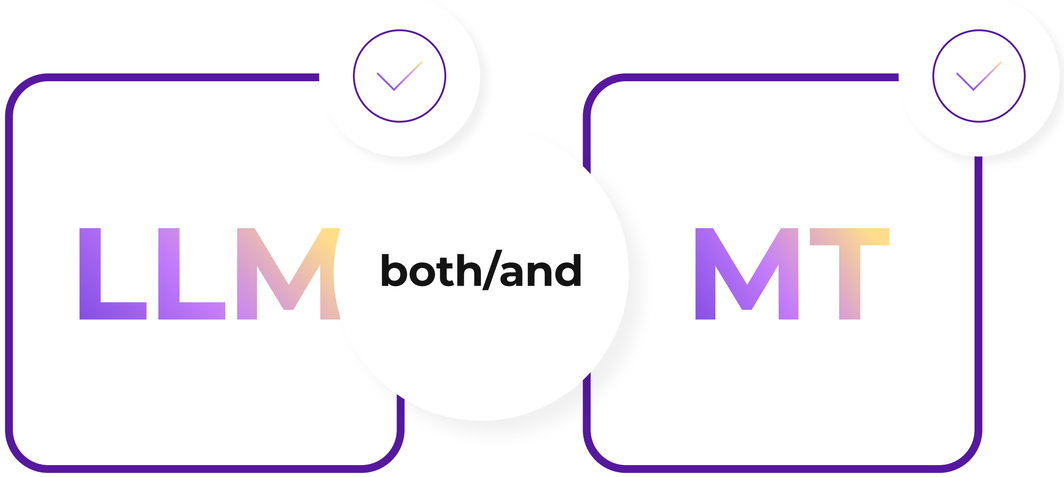
LLMs have an impressive ability to produce natural-sounding, culturally appropriate translations that consider different style preferences. They are also very good at choosing the correct domain-specific terminology and phrasing.



Examples include instances where sentences sound natural and are grammatically correct but may not be accurate.

As such, LLMs are far from a replacement for neural machine translation — which outperforms LLMs in producing more accurate translations.

So, it doesn't have to be "either/or." **Instead, it can be "both/and."**



LLMs and MT are complementary tools. MT systems such as Google Translate or DeepL can facilitate quick and relatively accurate translations. They are also more reliable and predictable in their output. As a result, they produce a good base translation. But where they struggle — producing contextually nuanced and colloquially fluent text — LLMs excel.

At Smartling, we've found that LLMs are best used as a pre- and post-processing tool — that is, before and after MT. They can minimize the risk of returning poor-quality MT output by correcting grammatical mistakes or typos in the source text. They can also compare the source text to previously translated strings and repair high fuzzy matches. Moreover, once the text has gone through MT, LLMs can further refine the translation.

Here are some examples of LLMs at work:

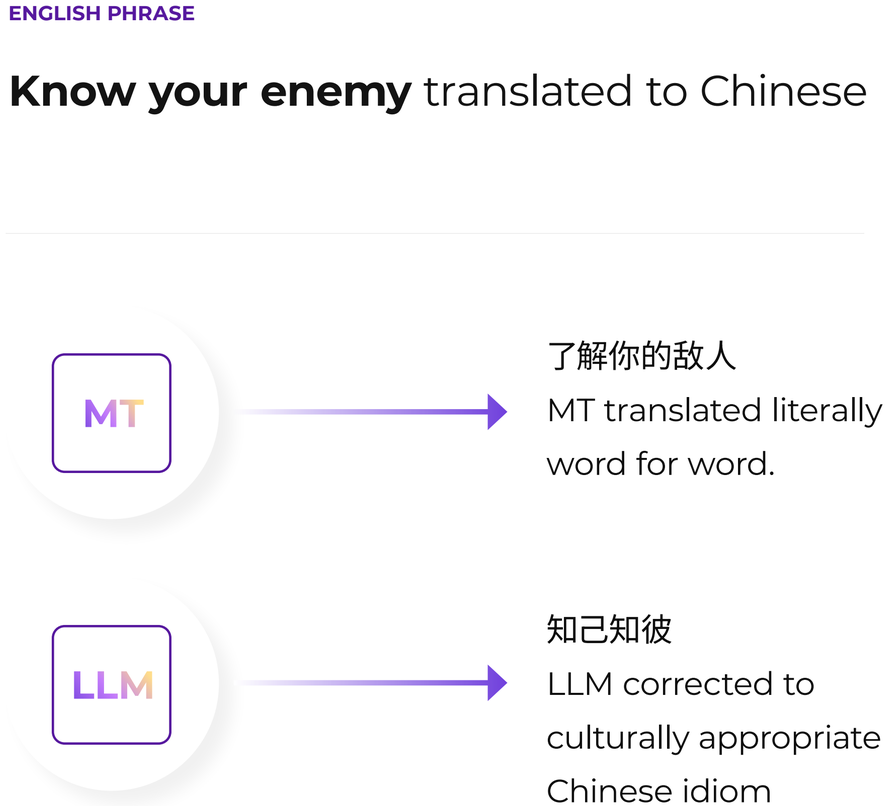
"At Smartling, we've found that LLMs are best used as a pre- and post-processing tool — that is, before and after MT."

EXAMPLE #1

Idiomatic expressions are translated with the appropriate nuances.

An English phrase like “know your enemy” might be translated literally by MT as “了解你的敌人.” An LLM could recognize this as non-fluent output and correct it to the more culturally appropriate Chinese idiom, “知己知彼,” which is used to describe knowing yourself and your enemy equally well in order to win.

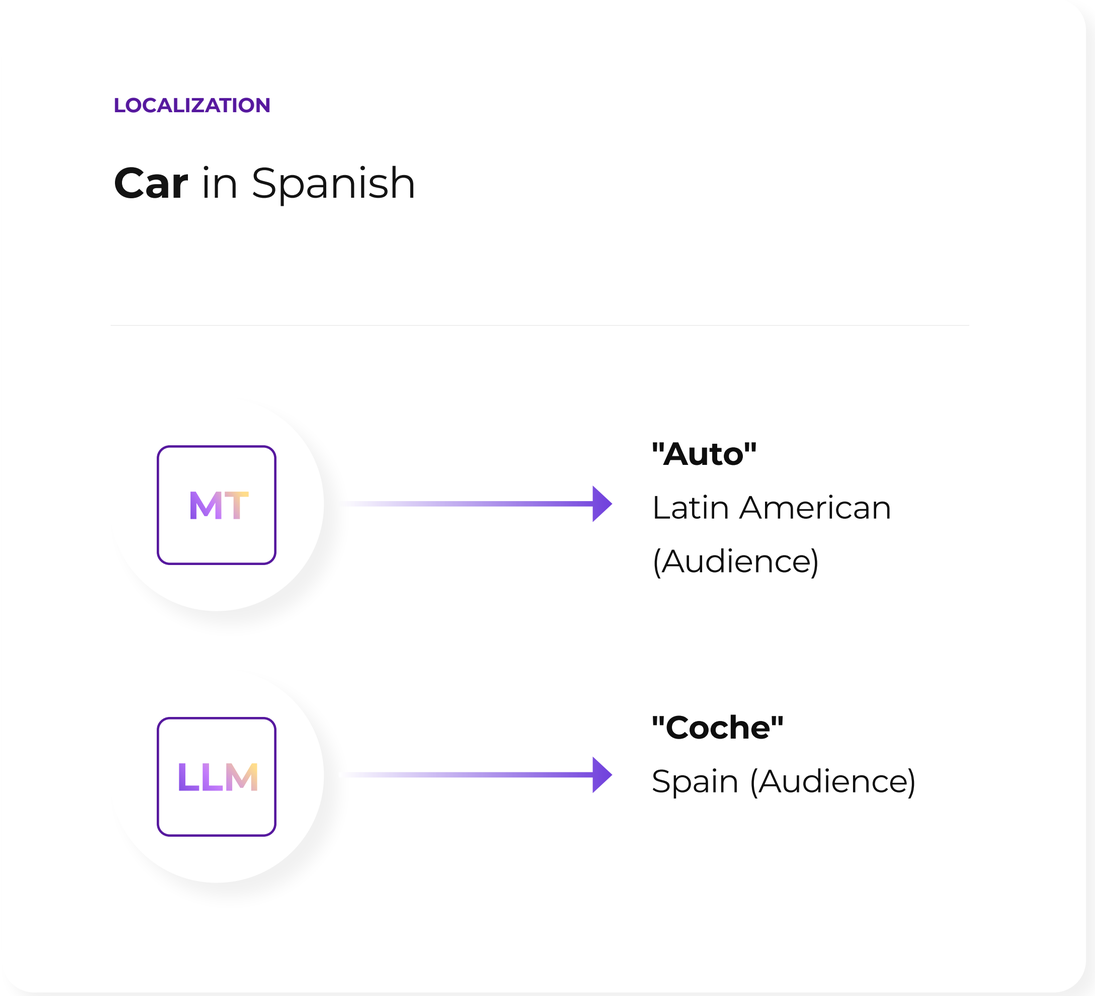
But translation needs don't usually stop there. So, when choosing a translation solution, you'll want to consider whether the solution's software will work with your entire tech stack. That will make it easy to send content for translation without leaving the applications you use daily.



EXAMPLE #2

The text is properly localized to the target audience's culture.

The word for "car" is different in Spanish for the audience in Latin America versus in Spain. MT engines are not currently equipped to make this adjustment. But an LLM could ensure "car" is localized appropriately to "auto" for an audience in Latin America and "coche" for an audience in Spain.



EXAMPLE #3

Translations adhere to brand-specific terminology.

A computer company who is translating their material from English into Spanish may choose to specify that “computer” always be translated as “ordenador” instead of ‘computadora.”

An LLM could be instructed to ensure adherence to the following glossary entry: {computer : ordenador}.

COMPUTER COMPANY INSTRUCTION

Computer always translated as **ordenador**

LLM GLOSSARY TASK

{computer : ordenador}

EXAMPLE #4

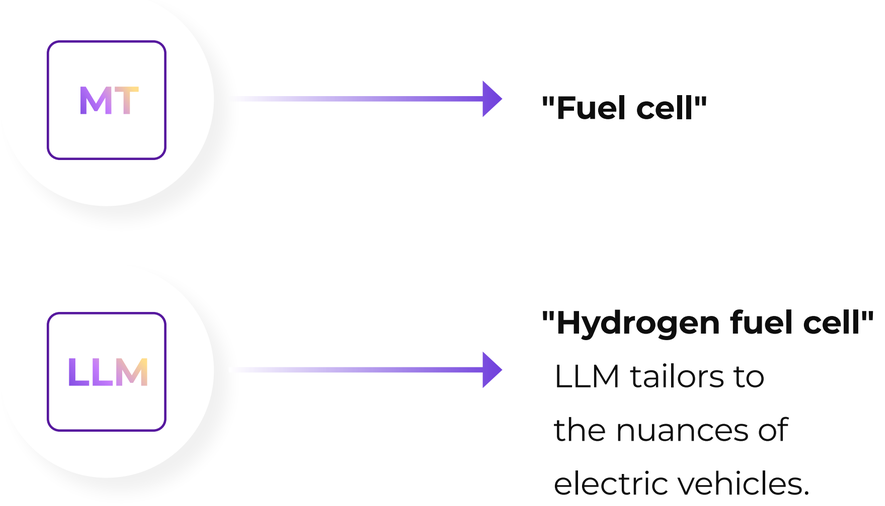
Translations are aligned with specific terminologies and the intended audience.

A standard MT may translate the German technical term “Brennstoffzelle” to “fuel cell.”

But for an automotive engineer versed in the nuances of electric vehicle technologies, an LLM can tailor this translation to “hydrogen fuel cell” if the context implies this specificity.

TRANSLATION

German technical term: **Brennstoffzelle**



EXAMPLE #5

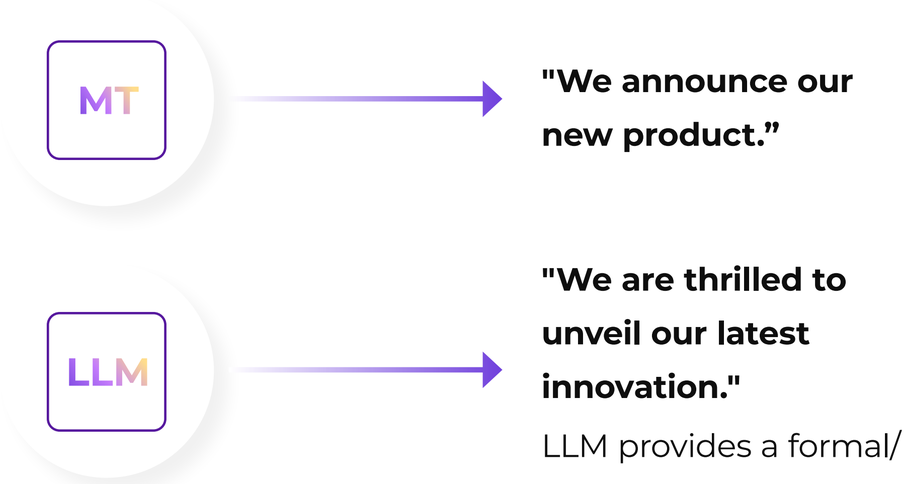
Translations match the formality or colloquialism necessary for the context.

A business document in Japanese may have a varying tone when translated for internal use versus a press release for public consumption.

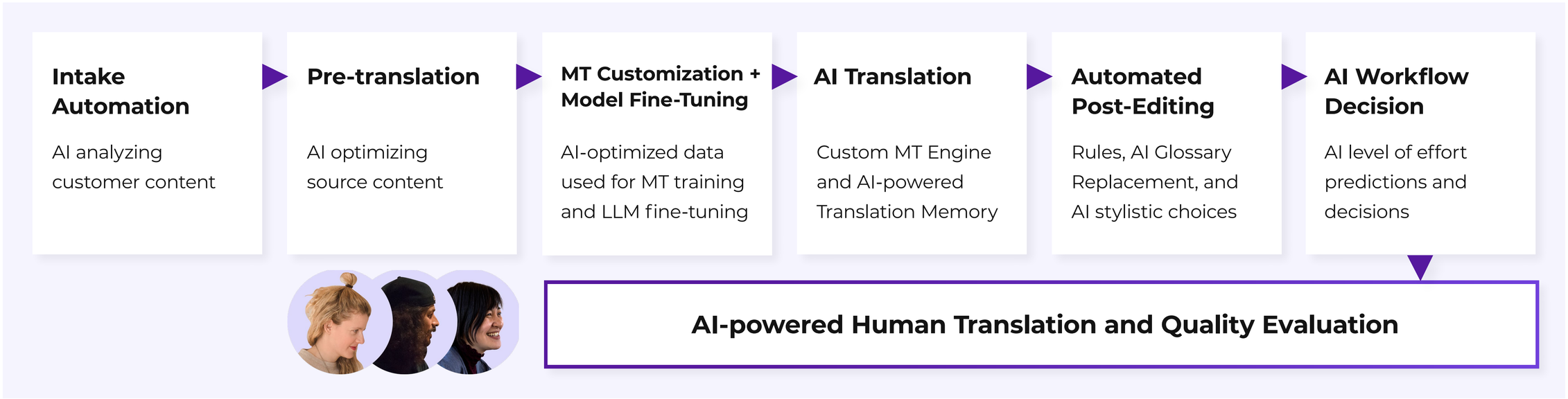
A traditional MT might deliver a dry, literal translation. An LLM, on the other hand, can imbue the end product with a more formal or promotional tone, as required, transforming “我々の新製品を発表します” from “We announce our new product” to the more engaging “We are thrilled to unveil our latest innovation.”

TRANSLATION

Japanese press release: 我々の新製品を発表します



Smartling Global Content Delivery with LanguageAI™



Integrating LLMs within a translation workflow might, therefore, look like the above graphic. At Smartling, we use MT engines powered with LLM pre- and post-editing functionalities to provide initial drafts that translators review and edit to perfect the final output. This results in a more efficient workflow and higher quality translations at a faster speed.

Best practices for using LLMs in translation

When done right, LLMs can do a lot of the heavy lifting in translation workflows. Although there are several opportunities to implement best practices, we've found that focusing your attention on two areas, in particular, will make a notable, positive impact on both translation quality and your bottom line.

Batching

Batching — processing a group of strings together — has significant advantages: shorter turnaround times, lower costs, and efficient resource utilization. It also enables the LLM to glean context from other strings within the same batch, potentially improving coherence and consistency. But there are drawbacks:

1. **Tracking translation quality may be more challenging.**
   One erroneous translation within a batch might be harder to isolate and rectify.
2. **Long texts may not fit easily into a single batch.**
   This may mean splitting the text into several batches, which can disrupt context.
3. **Un-batching can get complicated.**
   LLMs aren't foolproof. Bugs happen. If a string in the middle of the text wasn't returned, you may run into problems un-batching.

The opposite approach — translating string by string — means that each translation can be individually monitored, potentially resulting in higher-quality translations. This is especially beneficial for texts requiring high levels of accuracy, such as legal or technical documents. But this approach also has drawbacks:

1. **It can be slower and more costly.**
   It doesn't take advantage of the LLM's full capacity, leading underutilization of computational resources and longer processing times.
2. **A lack of context may hinder quality.**
   Translating in isolation means the LLM cannot draw upon surrounding content for context. This may result in a less nuanced, less interconnected translation.

LLMs and MT engines in action

Fortunately, there is a middle ground.

Using a rolling window approach when batching strings keeps costs down, provides sufficient context, and helps you avoid massive headaches if things go wrong. In addition, we've found that, when combined with a domain-specific prompt, this approach results in the highest quality translations overall.

What is a "rolling window approach"? Essentially, it's batching but with one major tweak. Instead of sending strings of a specific file for translation in static batches (e.g., 0–9, 10–19, 20–29, etc.), you would send strings in batches of a fixed size that "roll" over the file.

It works like this: Strings 0–9 are sent for translation. String 0 provides context for String 1, Strings 0 and 1 provide context for String 2, and so on, all the way to String 10, where Strings 0–9 are used as context.

Then, when it comes time to translate String 11, Strings 1–10 provide the context. When String 12 enters the window, Strings 2–11 are used as context. This continues until the entire file is translated.

Can you request translations into several languages at a time?

While it may be tempting to bundle your translation requests — i.e., requesting translation from one language to multiple languages in one prompt — we don't recommend it.

Our testing revealed that asking the LLM to translate one string into several languages at once resulted in translations that contained more erroneous word choices and, therefore, were unidiomatic. Instead, we've found that sending strings for translation one locale at a time produces the best results.

One to One or One to Many: Results

Language	Critical	Major	Minor
DE	0	0	85
One to One G2 - German	0	1	70
One to Many - German	0	0	85
ES	2	3	1
One to One G2 - Spanish	0	10	22
One to Many G2 - Spanish	0	15	15
ZH			
One to One G2 - Simplified Chinese	3	6	132
One to Many G2 - Simplified Chinese	7	31	217

A sample size of approximately 200 strings was used for testing. The translations generated were evaluated by professional linguists using Multidimensional Quality Metrics (MQM), an industry-leading framework for assessing translation quality. Note: Linguists can log multiple errors per string if they find multiple errors in one string.

Prompting

LLMs generate translations based on your instructions, or prompts. But how you structure the prompt matters. Here are six best practices to make sure you get the results you want.

1. **Start simple.**
   Talk to the LLM as you would a young child. Keep the prompt short, simple, and specific. You can iterate from there. Break large tasks into smaller subtasks. Include examples. Then, use delimiters to separate your instructions from the text you wish to translate.

2. **Give positively-stated commands.**
   Give direct commands — e.g., “translate” or “correct the grammar.” Avoid commands with “don’t.” It’s often best to tell the LLM what to do rather than what not to do.

3. **Be precise.**
   Generalizations leave room for interpretation. Take, for example, this instruction: “Translate the following text to Spanish using just a few sentences.” What does “a few” mean? To some, it might mean three to five sentences; to others, two or three. Precision is crucial.

4. **Think step-by-step.**
   If a task is best expressed as a sequence of steps, list all the steps clearly and succinctly. This will make it easier for the LLM to carry out the steps correctly and in the right order.

5. **Give the LLM context.**
   Provide details about your target language and audience, as well as any domain-specific terms or background information. This will help the LLM generate appropriate and relevant output.

6. **Test your prompts extensively.**
   Test how your prompt performs on various content domains and file types — HTML, JSON, SRT, etc. — before deploying it on a large scale. Understanding how the LLM will treat URLs, placeholders, tags, and more will help prevent issues during translation. Use the results to refine your prompt, then retest.

Conclusion

LLMs are still in their infancy. Therefore, today's LLMs should be considered a complement to — not a replacement for — existing technologies like MT.

Furthermore, they should not be looked upon as a substitute for human intervention. Rather, they are a very effective productivity tool that companies and professional translators alike can use to reduce time spent on translation while maintaining high quality standards.

They also have enormous potential. As this technology matures, so will its capabilities. That's why it's crucial to find a trusted translation partner who is both comfortable with and excited by LLMs.

At Smartling, we're committed to discovering innovative ways to help companies do more with less. We are investing heavily into researching and testing AI and machine learning technologies — including LLMs — so Smartling customers can get global-ready faster.

About Smartling

Smartling's LanguageAI™ platform is revolutionizing the translation and localization of the world's digital content. Named as the top translation management system by CSA Research and by users on G2, Smartling harnesses AI and machine learning to enable the automation of workflows, integrates seamlessly with existing tech stacks and measures and improves translation quality at enterprise scale for a fraction of the cost of traditional translation.

Smartling is the platform of choice for hundreds of B2B and B2C brands, including IHG Hotels & Resorts, State Farm, British Airways, and Lyft. Smartling is headquartered in New York with an office in Dublin.

SMARTLING

hi@smartling.com

1 (866) 707-6278

smartling.com