LIONBRIDGE

THE LIONBRIDGE 2023 MACHINE TRANSLATION REPORT

Lionbridge experts analyze trends in Machine Translation, provide insight into likely developments, and project how advancements will help businesses flourish.

EXECUTIVE SUMMARY

During the past few years, businesses have increasingly embraced Machine Translation (MT) technology. Even so, MT's quality stagnated in 2022. The Lionbridge Machine Translation Tracker — the industry's longest-standing measurement of the major MT engines — bears this out. It registered no significant quality improvements among the top five engines in 2022.

These results were a harbinger of a new Machine Translation technology paradigm emerging in 2023. As it takes hold, MT quality will improve significantly.

More and more companies have turned to Machine Translation technology to enhance customer experience — for good reasons. MT offers a fast and effective alternative to human translation and enables companies to have real-time conversations with their global customers.

Machine Translation has also helped companies improve their employees' productivity across languages and geographies, assisting brands in delivering better workforce experiences during the COVID-19 pandemic.

Business use cases for Machine Translation have grown over the last few years.

Here's how companies from a variety of sectors use Machine Translation for business gain:

E-commerce and Retail

Retailers use MT to quickly and efficiently translate product descriptions and other marketing collateral to bring their offerings to multiple markets. The technology has enabled these businesses to grow cross-border trading at exponential rates.

Travel and Hospitality

Travel-related businesses use MT to translate descriptions of destination sites, such as hotels, rental properties, restaurants, or other points of interest, and to translate customer reviews. The technology has enabled these businesses to enhance customer experience and connect with prospective customer bases faster.

Healthcare

Medical-related organizations use MT to translate medical research and clinical trial data. The technology has enabled these users to make information more readily available to the public and improve treatments and patient outcomes.

Legal and Financial Services

These service providers have been among the early adopters of MT, using it for eDiscovery processes and market research activities. MT has enabled them to process substantial amounts of multilingual content.

The Public Sector

Governmental agencies use MT to provide local municipal services in the languages of the local communities, thereby cost-effectively overcoming language barriers.

Global Companies of Any Sector

Companies with global workforces use MT internally to facilitate internal communications, whether the communications involve a chat or a simple document.

They also use it to permit support teams to extend their coverage and reduce incident resolution times. The technology has enabled global companies to bolster office productivity.

Every global company wishing to thrive in our interconnected economy must embrace and fully leverage Machine Translation.

Despite past advancements and growing use cases, MT still has limitations. Some of its longstanding quality issues include its inability to attain and consistently achieve the right formality level, tone, or handling of negation. These limitations deter growth. Research into and the use of Large Language Models (LLMs) hold promise to resolve these issues and unlock a new technological leap for Machine Translation.

Big Tech's investments in LLM technology — such as Microsoft's **$10B investment in OpenAI**, the company behind the ChatGPT, GPT-3, and GPT-4 models — is accelerating the development of this technology and advancing the Natural Language Processing (NLP) field. These advancements will inevitably disrupt the translation and localization industry and change how companies create and translate content.

The exponential advancements of NLP, specifically LLMs, will transform how content is created and localized. The upshot will be exponential gains in productivity and speed as human translators process much larger volumes of content.

Companies that master and leverage AI in their content engines will gain a significant competitive advantage in our increasingly digital economies.

What Does the Future Hold for Machine Translation?

Having tracked the major MT engines for many years and masterfully leveraging new technologies, Lionbridge is well-suited to analyze developments in 2023 and beyond. We anticipate MT's existing Neural Machine Translation (NMT) paradigm will end. A new paradigm will replace it, likely based on Large Language Models (LLMs) like ChatGPT. The release of GPT-4 and growth in LLMs are having significant business implications.

You can expect the following:

- A significant leap in MT quality, including workflow automations
- Increased content output
- A reduced supply of top-notch human translators
- Increased adoption of Machine Translation
- Machine Translation as a means for Customer Experience (CX) enhancement

Every global company wishing to thrive in our interconnected economy must embrace and fully leverage Machine Translation. Read on as we examine the technology's developments — or lack thereof — in 2022, what that has meant in 2023, and what it will mean for the years ahead.

What Does the Machine Translation Quality of 2022 Tell Us About How It's Evolving?

Machine Translation (MT) is having a moment. Companies increasingly realize the benefits of using MT technology. This realization is evident even in sectors like life sciences, which traditionally have been reluctant to embrace MT technology.

But as far as output quality goes, improvement was flat in 2022, which paves the way for new and exciting changes for MT.

Except for minor improvements by Microsoft at the end of 2022 and by Amazon, the major MT engines are performing similarly in quality. Further, these minor improvements are less significant than quality jumps in previous years.

At the end of the Statistical MT paradigm, we saw similar indicators: flat lines for quality improvement and a uniform MT output quality among all the top engines. It leads us to one conclusion — we are approaching the end of the current Neural Machine Translation (NMT) paradigm.

Large Language Models (LLMs) — with their massive amounts of content, including multimodality and multilingualism — will be a big part of what's next.

Companies that master and leverage AI in their content engines will gain a significant competitive advantage in our increasingly digital economies.

A MACHINE TRANSLATION PRIMER

To fully capitalize on Machine Translation and enjoy its profound benefits — for the first time, enabling companies to Locaⁱlize everything™ — it's necessary to have a fundamental understanding of the evolution of the technology.

What is MT? What has triggered widespread global adoption? What are its major strengths and pitfalls to avoid when using it? And what is the backdrop against which it has been evolving?

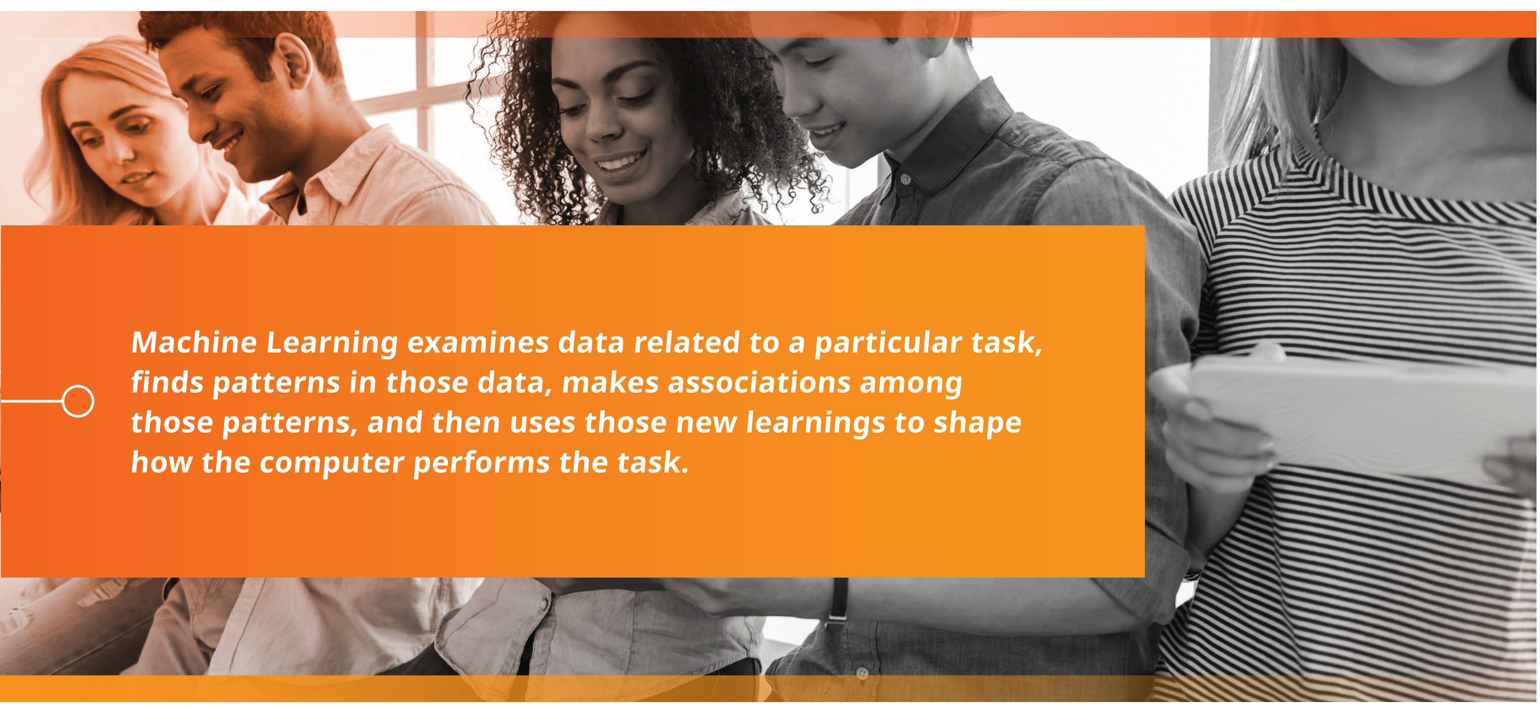
Artificial Intelligence

In the most basic sense, Machine Translation uses Artificial Intelligence (AI), or the "intelligence" machines demonstrate, to perform tasks that usually require inherently human thinking, such as learning and problem-solving. In this case, AI is used to perform translations. In recent years, AI has benefited from increasing computer power. More powerful computers yield more intensive processing during a task at hand and more advanced Machine Learning, which is how computers gain the knowledge required for AI applications.

Machine Learning

Machine Learning (ML) is a branch of computer science that uses massive amounts of data to teach computers how to perform tasks. Machine Learning examines data related to a particular task, finds patterns in those data, makes associations among those patterns, and then uses those new learnings to shape how the computer performs the task.

If, after this analysis, the computer gets better at performing the task, then Machine Learning has occurred. Because we have vast language and localization data, people are using Machine Learning to improve computer performance in everything from weather forecasting to automatic stock selection to Machine Translation.



Machine Learning examines data related to a particular task, finds patterns in those data, makes associations among those patterns, and then uses those new learnings to shape how the computer performs the task.

What Is Machine Translation Technology?

Machine Translation is automated translation. When you present source material to a computer in one language, it gives it back to you in another language. MT technology is imperfect, but it's one of the most powerful tools for producing usable translations more efficiently.

Over the last several decades, MT has improved the quality of its output and the breadth of languages it supports. From simple word replacement systems in the very early days of MT to the explicitly coded grammar and lexicons of rules-based MT, to the number-crunching paradigm of Statistical MT, to the Deep Learning and neural networks of Neural MT, the development of Machine Translation has mirrored our increasingly sophisticated use of computers.

How Has Machine Translation Evolved?

Rule-based Machine Translation

Early Machine Translation engines used rule-based methods. As the name suggests, Rule-based Machine Translation (RBMT) relies on language rules developed by humans and from dictionaries to execute translations. These systems are word-based.

Although this method can produce accurate translations, Rule-based Machine Translation systems work best with simplified languages that use simplified grammar and a smaller dictionary.

The reduced complexity allows the Machine Translation engine to perform better. Rule-based methods are less successful when translations are needed for specific domains, when idioms are used, or when the text has ambiguity. Since the introduction of RBMT, language technology has evolved considerably.

MT technology is imperfect, but it's one of the most powerful tools for producing usable translations more efficiently.

The learning algorithm then builds a language model that calculates the likelihood that given words and phrases appear next to one another in the target language.

Statistical Models

Statistical Machine Translation (SMT) relies on a large number of translation candidates for a given source sentence, then selects the best one based on the likelihood of words and phrases appearing together in the target language.

SMT learns about translation through the lens of "n-grams" — small groupings of words that appear together in the source and target language.
The SMT system is given training material — that is, many examples of sentences in the source language and their translations into the target language.

The learning algorithm divides source sentences and target sentences into n-grams. It determines which target language n-grams are likely to appear in a translation when a certain source language n-gram appears in a sentence.

The learning algorithm then builds a language model that calculates the likelihood that given words and phrases appear next to one another in the target language. When it's time to translate new material, the SMT system breaks the new source sentence down into n-grams, finds the highly associated target language n-grams, and generates candidate sentences.

The final translation is that sentence whose target language n-grams correlate most highly with the source sentence's n-grams and whose target language words are most likely to appear together in the target language. SMT works surprisingly well, especially since there is nothing linguistic about an SMT system; indeed, the system only considers n-grams, never a complete sentence.

Hybrid Machine Translation

Companies then began experimenting with Hybrid Machine Translation (HMT), which combined the output of Statistical Machine Translation and Rule-based Machine Translation systems.
These advancements popularized Machine Translation technology and helped adoption on a global scale.
Another technological leap would come from a newer approach to MT: Neural Machine Translation.

Neural Machine Translation

Neural Machine Translation (NMT) overcomes the most significant shortcoming of SMT: its reliance on n-gram analysis. NMT empowers the machine — the system receives the training material, just as it would with SMT, but there's a fundamental difference. Once the system receives the material, it decides on its own how to learn everything possible about that data.

NMT systems build information vectors for each source sentence, associating information about each word with surrounding words. Some systems develop hundreds of pieces of information per word, creating a deep sense of accuracy.

Through deep learning, NMT systems capture a massive amount of information about each word and source sentence. Then, it uses an attention model to hone the critical features it has learned by analyzing these massive data streams, which are important for the translation process.

The result is translations that show marked improvements in fluency, which means that computer-generated translations are starting to sound more and more natural. Neural Machine Translation has addressed some of Machine Translation's longstanding shortcomings, including the poor readability of automated translations and its incompatibility with certain languages.

This development resulted in Machine Translation being good enough for comprehending or gisting large volumes of documents and for regular, non-mission-critical business documents. Of the three approaches, NMT is the clear winner. **It has been a game-changer as it has increased the use of MT to accelerate production processes.**

Visualizing How Neural Machine Translation Works

While NMT trains on Translation Memories (TMs) as Statistical Machine Translation does, it uses deep learning — and possibly a higher volume of training data — to build an artificial neural network.

You can describe Neural Machine Translation as "raising" a neural system. Think of it like playing the piano: When you make a mistake, you back up, try again, and repeat the process until you succeed. Neural MT systems try to find their way through neural networks similarly.

When you present training material to the deep learning algorithms, you don't necessarily tell them how to proceed. You let the system find patterns, such as contextual clues around the source sentence. The specifics of the process, however, remain mysterious in many ways.

Much of the processing is done in "hidden layers" of complex data, meaning it's hard to see how the neural network makes its decisions, which makes it difficult to build a concrete mental picture of the NMT process. **We can only present the training material, let the algorithms do their thing, and tweak the training material if the translations aren't accurate.**

Once the system receives the material, it decides itself how to learn everything possible about that data.

What Is the History of Machine Translation?

Machine Translation has come a long way since its inception in the 1950s.
Here are some of its major milestones.

1954 Georgetown researchers perform the first-ever public demonstration of an early MT system.

1962 The Association for Machine Translation and Computational Linguistics is formed in the U.S.

1964 National Academy of Sciences forms a committee (ALPAC) to study MT.

1970 The French Textile Institute begins translating abstracts using an MT system.

1978 Systran begins to translate technical manuals.

1989 Trados is the first to develop and market Translation Memory technology.

1991 The first commercial MT system between Russian, English, and German-Ukrainian is developed at Kharkov State University.

1996 Systran and Babelfish offer free translations of small texts on the web.

2002 Lionbridge executes its first commercial MT project using its rule-based MT engine.

Mid-2000s Statistical MT systems launch to the public. Google Translate launched in 2006, and Microsoft Live Translator launched in 2007.

2012 Google announces that Google Translate translates enough text to fill 1 million books daily.

2016 Both Google and Microsoft enable Neural Machine Translation (NMT), slashing word order mistakes and significantly improving lexicon and grammar.

2020 As of October, Google Neural Machine Translation (GNMT) supports 109 languages.

2022 ChatGPT, a Large Language Model (LLM) that can generate human-like text based on context, goes mainstream in November with significant implications for Machine Translation.

2023 A major MT paradigm shift is anticipated as a type of LLM evolves and disrupts MT.

What Are Machine Translation’s Benefits?

Machine Translation is highly appealing for two primary reasons:

- It’s incredibly fast, providing reams of translations within seconds
- It’s highly cost-effective, enabling companies to translate more than they ever thought possible

Speed + cost-savings = more localized content and enhanced customer experience

Automated translations enable companies to Localize Everything and meet the ever-growing demand for more content to be pushed out to the market faster, all within the same or even smaller budget.

Why Else Should Companies Implement Machine Translation?

In addition to saving time and reducing localization costs, using MT for non-critical content enables companies to:

- Localize more material, thereby extending their reach based on the number of languages they translate and the number of markets they enter
- Improve their terminology and brand voice consistency
- Localize real-time customer interactions and thereby enhance customer satisfaction
- Increase the productivity of global teams by removing communication barriers

Machine Translation also helps companies allocate more of their resources and budget to content creation since they require less of their budget for translation.

This budget reallocation can help companies increase their content velocity and attract more readers to fuel their growth. This gain is especially true for digital-first businesses.

On average, Lionbridge customers achieve the following results when using Machine Translation as part of their workflow.

Greater Translation Efficiency

Up to 40% cost savings.

Faster Turnaround Times

Up to 60% shorter delivery times.

Enhanced Customer Experience

The ability to speak to customers in their native language everywhere.

What Is Machine Translation's Primary Shortcoming?

Despite MT's ever-growing sophistication and improvements to its quality output, engines still have a long way to go before they can match humans' understanding of nuance, tone, sarcasm, humor, meaning — the list goes on.

If you rely solely on Machine Translations, its robotic tone could end up compromising your message. That, in turn, could suggest to customers that your service is low-quality or unreliable.

Translations of public-facing content must read naturally and fit seamlessly into the cultural context of your target market to be effective. To produce high-quality, localized content, it's often worth paying extra for native speakers to review MT output. That's where Machine Translation Post-Editing (MTPE) comes in.

How Does Machine Translation Post-Editing Provide the Best of MT and Human Translation?

Machine Translation Post-Editing (MTPE) is a hybrid of MT and traditional human translation. As its name suggests, a human provides post-editing following Machine Translation. First, the software produces an initial translation of the material. Then, a human translator edits the text, checking for accuracy, clarity, flow, and local resonance. This approach offers the unparalleled speed of MT with the attention and sensitivity of human translators.

While machines make short work of bulk translations, human translators infuse the copy with the translation quality only a human can provide. This hybrid approach is a good option when you translate customer-facing content and want a tailored method that maximizes quality and minimizes turnaround times. This option is one of the quickest and most affordable ways to get high-quality translation results.

What Other Risks Are Associated With Machine Translation?

Though MT has made significant advances, the technology has yet to be perfected. As such, you are taking some risks when deploying the technology beyond getting a robotic output — especially if you take no action to mitigate MT's shortcomings.

What Exactly Can Go Wrong?

Here's what can happen when using Machine Translation:

- Your copy may not be inclusive or reflect your intended level of language formality due to the engine's inability to know when to use neutral words or match the appropriate formality level to your target audience.
- Your copy may have errors — to varying degrees — in the translation output.

While these issues can cause negative consequences, errors have the potential to be especially harmful to your company.

MT Risk: Errors

The Type of Errors Resulting From MT

MT engines produce two major types of errors:

- **Standard Errors**
  These are target language mistakes that pertain to the content's linguistic features. Standard errors include grammar, spelling, or punctuation mistakes. While native speakers are likely to notice these slip-ups, they rarely cause disastrous consequences for a company.
- **Catastrophic Errors**
  MT catastrophic errors transcend linguistic features and inaccurately convey the intent of the source text. Misinformation or misunderstandings that result from these errors can potentially cause companies reputational, financial, or legal repercussions and lead to adverse public safety or health consequences.

Why MT Engines Make Catastrophic Errors

Think of a catastrophic error as an MT engine malfunction. It can occur if the engine doesn't understand the context of the text, such as when one word has two meanings or if there is a typo in the source text. These errors can happen if the engine is not trained well or a flawed glossary is used, which then causes the same mistakes to appear repeatedly. Catastrophic errors occur because engines are imperfect despite their sophistication. Machines cannot exercise judgment the way people can.

Catastrophic errors typically show up in three primary contexts:

- **Mistranslation of Key Entities**
  This error is the mistranslation of proper names (individuals or organizations), important numbers, or units of measurement.
- **Negation and Opposite Meaning**
  This error involves mistakes in the target language that result in the opposite meaning of the original text's intended meaning.
- **Hallucinations**
  This error occurs when MT introduces content that is not present in the source. Engines may generate offensive, profane, aggressive, or highly sensitive words under certain circumstances. When this type of catastrophic error happens, there is a problem with the MT engine software itself.

The public witnessed a real-world example of a catastrophic error involving a proper name on a Spanish governmental agency website. In that instance, the department head's name, Dolores del Campo, was omitted from the ministry's official site. Instead, the literal translation — It is pain of field — appeared in place of the name.

How to Minimize The Risk of Catastrophic Errors

Only when computer scientists improve existing MT technology to prevent these errors from occurring in the first place can we use automated technology to identify potential issues, revise problematic sentences, and promote accuracy during the translation process. For instance, Lionbridge administers specific automated quality checks in translated texts via its **Smaïrt MT™** offering — and in conjunction with its cutting-edge **Smaïrt Content™** language AI — to detect errors while maintaining MT speed and minimizing the need for post-editing by human translators.

These automated methods detect:

- **Incorrect translations of proper names,** including the names of individuals or organizations, by identifying entities in the source text that can be both a named entity and a common word
- **Offensive, profane, or highly sensitive words** by combining a supervised Machine Learning (ML) algorithm and a list of offensive terms
- **Opposite meanings** between original and translated texts by identifying negative particles (sentences that contain the word not or its shortened form n't) in the original text or the translated text, but not in both
- **Hallucinations in the translation** — words that appear in the translation but not the source — via dictionaries or a list of offensive words to address insulting hallucinations

Automated quality checks do not guarantee the elimination of catastrophic errors. These checks can miss errors, causing a false negative result. Nonetheless, they are highly effective at helping us find problems.

Using this approach, we can focus professional translators on the flagged sentences and avoid reworking the whole document. When we can alert professional translators to where problems are most likely to be found, we enhance the efficiency of the localization process.

Catastrophic errors occur because engines are imperfect despite their sophistication. Machines cannot exercise judgment the way people can.

MT Risk: Mismatch in Language Formality

What is Language Formality, And Why Does it Matter?

Language formality is the degree of formality with which we express ourselves. We typically wouldn't speak to our supervisor like we would talk to a close friend, even if we related the same story to each person. We may choose to use different vocabulary and grammar.

We're more likely to use **formal language** in a business context or a more formal situation. Formal language connotes politeness or respect. We typically use informal language during relaxed situations and for people we know well.

Speakers from certain cultures may perceive a wrong formality level as rude. For instance, you wouldn't speak to an elder like you would to a child in Korean. There are different verb endings for each of the **seven speech levels in Korean**, five of which are categorized as formal and the remaining two informal. Misusing formality levels can be embarrassing.

Additionally, there are many cases in which formal language may be wrong, not because it is offensive but because it is totally out of place. Errors like these can alienate your audience. For example, translations for computer games and student programs require a casual tone.

When companies take the time and try to adapt the appropriate style to the target during translations, they send a strong message that they care about their customers. Such actions may be even more appealing to potential prospects. Effectively connecting with your audience helps you to succeed.

Speakers from certain cultures may perceive a wrong formality level as rude.

Why it's Difficult For Engines to Get Language Formality Right

Machine Translation and formal vs. informal language can be problematic as engines can produce incorrect and inconsistent formality. Why? MT models typically return a single translation for each input segment. When the input segment is ambiguous, the model must choose a translation among several valid options, regardless of the target audience.

Letting the model choose between different valid options may result in inconsistent translations or translations with an incorrect formality level.

It is especially challenging to get the correct output when the source language has fewer formality levels than the target language. For instance, languages like French have well-defined formal modes — tu vs. vous — while English does not.

Lionbridge's enterprise-grade Machine Translation solution, Smairt MT, allows linguistic rules to be applied to the target text to produce Machine Translations with the desired style or formality.

Addressing Language Formality When Using MT

Companies can control language formality through rules-based techniques that use rules to replace the undesired style with a correct translation and non-rules-based techniques that involve the use of custom MT models.

While most commercial MT systems do not support language formality or gender parameters, companies are making progress. DeepL (API) and Amazon (console and SDK) currently offer features that control formality.

There are three options to choose from: default, formal and informal. The default option does not change the formality of the Neural Machine Translation (NMT) output. The formal/informal function allows the user to choose between a formal or informal tone of voice. Specifically, the function sets the pronouns and related words used in the translation.

Lionbridge's enterprise-grade Machine Translation solution, **Smaᴵrt MT**, allows linguistic rules to be applied to the target text to produce Machine Translations with the desired style or formality. It can even promote neutral language and prevent **bias in localization** — the difference between referring to someone as a salesperson versus a salesman.

Our specialists maintain an updated database of rules that are fed back into the analysis of MT outputs to control the outcome. Provided that there is sufficient material, we combine a rule-based strategy with custom MT models to achieve an optimal result.

While it's challenging to get language formality right when using MT, it's possible. Later in the whitepaper, we'll explore exciting technological developments that may solve language formality-related issues.

2022 MACHINE TRANSLATION YEAR-IN-REVIEW

What Are the 2022 MT Key Trends?

Lionbridge MT experts found 2022 to be notable for both what did — and did not — transpire. Having observed so many MT-related technological advancements during the past few years, our team anticipated more of the same. But MT did not make significant strides, as our **Machine Translation Tracker** revealed.

With rare exceptions, the major engines made little-to-no improvements during the year. This trend has implications for the future. But first, let's take a closer look at the 2022 results.

How Did the Top MT Engines Perform in 2022?

When a company wants to start using MT or improve the way it currently uses MT, it is critical to identify which MT engines will work best based on their specific needs. As we delve deeper into how the major MT engines performed in 2022, one thing becomes very clear: One engine can't do it all.

Comparison of MT Engine Performance Based on Language

A company working with Spanish content benefited from selecting DeepL for its automated translations; it had better alternative options when translating Japanese. That's because each engine's performance varies based on the language it handles.

We calculated how well three major engines handled automated translations from English into numerous languages. We determined the quality by calculating the average edit distance — the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation.

The lower the number, the more effective the automated translation is. As shown in Figure 1, paying attention to these results is worth a company's while.

According to our analysis, in certain situations:

- DeepL translated Spanish better than Google and Microsoft
- Google translated Japanese better than DeepL
- Microsoft translated Polish better than DeepL
- The three engines performed similarly for Italian, Turkish, and Hebrew

These results demonstrate the complexity and challenges inherent in Machine Translation, which involves navigating the nuances and complexities of different languages, cultures, and domains.

It is not surprising to see variations in performance across various MT engines, as no single algorithm or approach can work perfectly for all languages and content types.

Why Does MT Perform Differently Depending on the Language?

Possible reasons for these performance differences include the quality and quantity of training data available for each language. Machine Translation models rely on a large corpus of high-quality bilingual data to learn how to translate a language accurately. The MT engine may struggle to produce accurate translations if there is a shortage of such data for a particular language.

Additionally, each MT engine's technical and linguistic features vary, with each engine employing its own set of algorithms, architectures, and approaches to Machine Translation. Specific engines may perform better on certain language pairs or content types, depending on the features and capabilities of their models. Moreover, the level of optimization and customization of the MT engine may also play a role. Some MT engines may be optimized and customized specifically for particular languages or domains, resulting in more accurate and effective translations for those use cases.

Overall, these factors contribute to the complexity and variability of Machine Translation, highlighting the importance of understanding the strengths and limitations of different MT engines for various languages and use cases.

Choosing the best MT engine for a specific use case requires expertise in Machine Translation. It is not a simple task. Careful evaluation and ongoing optimization are critical to ensure the best possible results.

The Average Performance of MT Engines Per Language

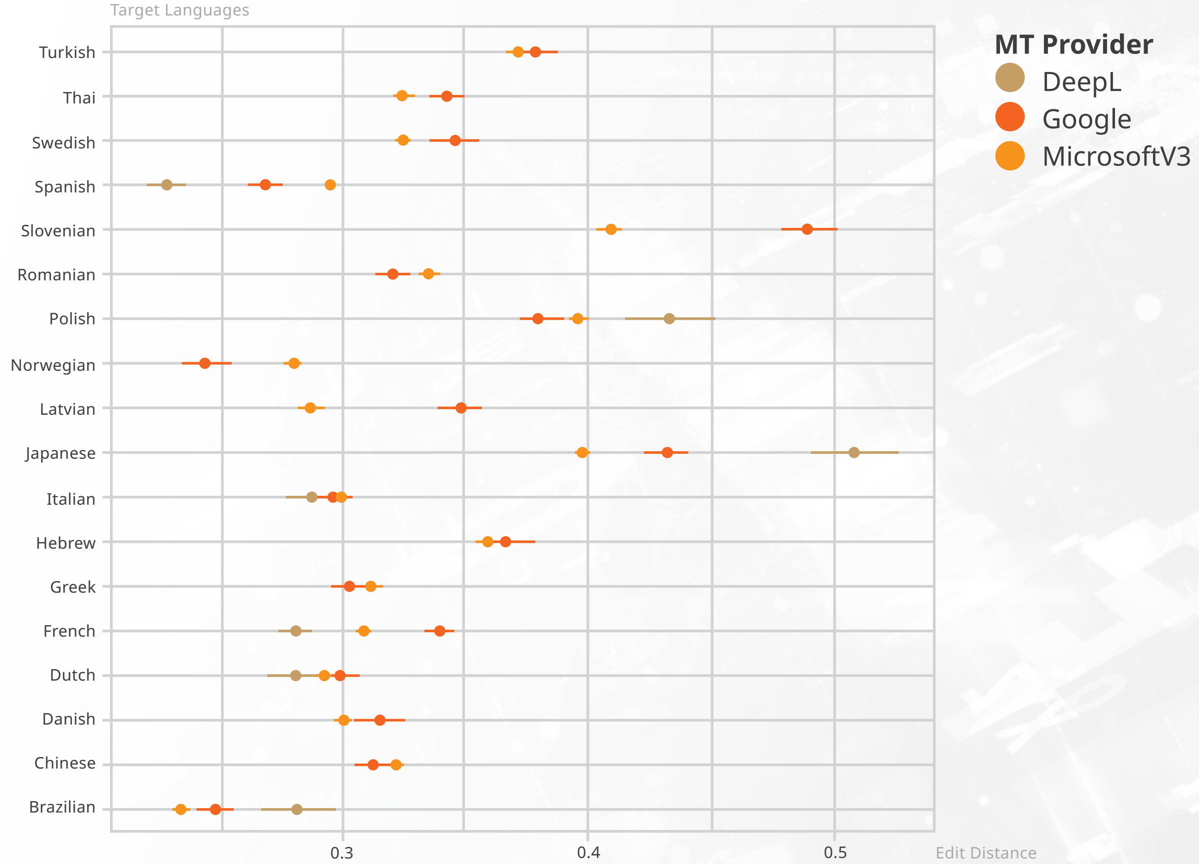


Figure 1. The average edit distance per language and MT provider

Automated Translation Performance Based on Domain

The more creative your content, the more difficult it will be for engines to translate it effectively.

Translation quality based on domain makes that point. We tracked the average edit distance — the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation — for the main domains to determine how well the engines handle various types of content.

The lower the number, the better the translation quality is.

According to our analysis, as shown in Figure 2,

- Engines found Media and Marketing content — known for its originality — the most challenging type of content to translate
- Content related to Textile and Fashion — which often has imaginative and nuanced descriptions — was the next most challenging type of content for machines to process
- Machines handled clear-cut content in the Automotive and Machinery sector best

We expected results like these. Generic MT engines shine when dealing with straightforward content that is relatively simple and easy to understand and has a clear structure and vocabulary that is not highly specialized or technical.

The Quality of MT Output Per Domain

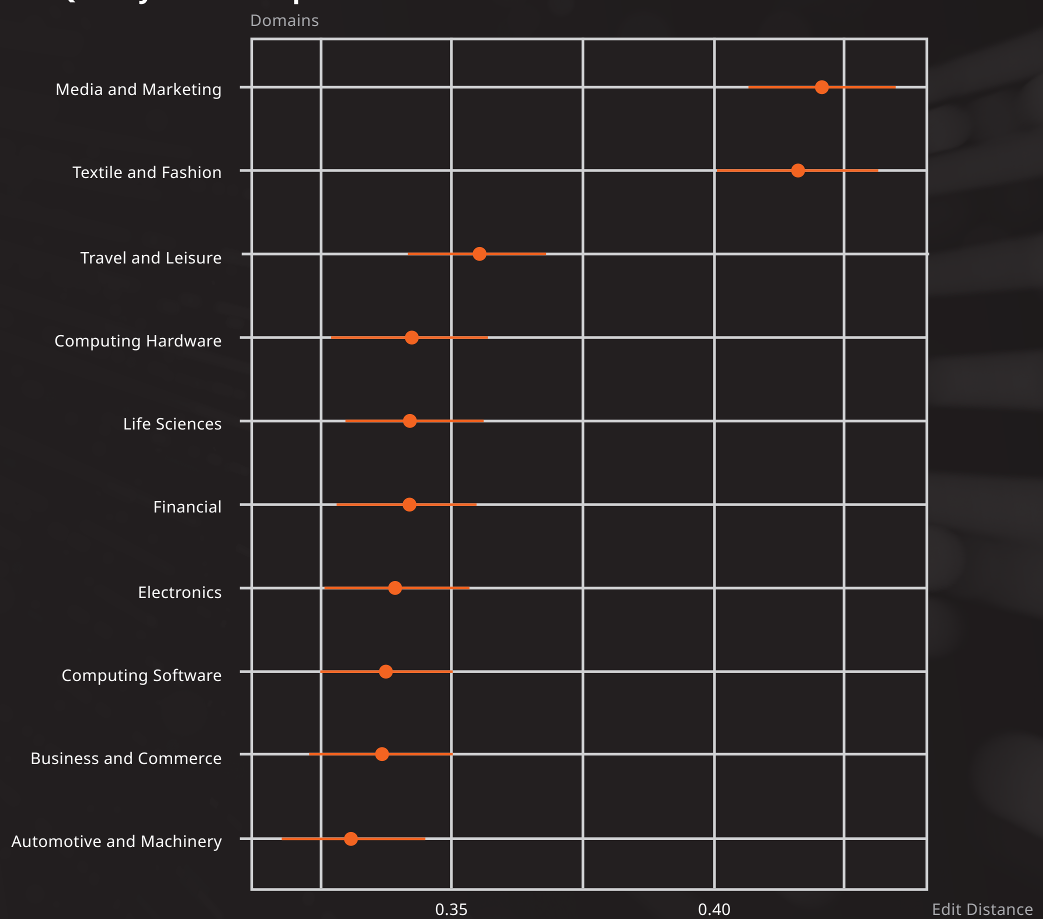


Figure 2. The average edit distance per domain.

Comparison of MT Engine Performance Based on Domain and MT Engine

A company translating Media and Marketing content benefited from selecting DeepL as its MT provider; a company translating Textile and Fashion-related content had better alternative options. Based on available data, we considered how well four major engines handled content for various sectors. We determined the quality by calculating the average edit distance — the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation. The lower the number, the more effective the automated translation is. As shown in Figure 3, engines don't perform similarly across sectors and content types in certain situations.

According to our analysis, in these specific examples:

- **Media and Marketing:** DeepL performed better than Google and Microsoft
- **Life Sciences:** DeepL performed better than Microsoft and Google
- **Financial:** Microsoft and DeepL performed better than Google
- **Automotive and Machinery:** Google performed slightly better than Microsoft and DeepL

However, note that for us to recommend the right system, we need to analyze your specific content. That way, we know what you need and which engine will address your requirements best.

The above results suggest the performance of MT engines can depend not only on language pairs as we commented on, but also on the specific domain or industry terminology and phrasing used.

Therefore, when selecting an MT engine, it is essential to consider the language pairs being translated and the specific domain and industry context. This approach requires expertise and knowledge of the language and the particular domain and can result in better translation quality and accuracy for the given content type and industry.

The Average Performance of MT Engines Per Domain

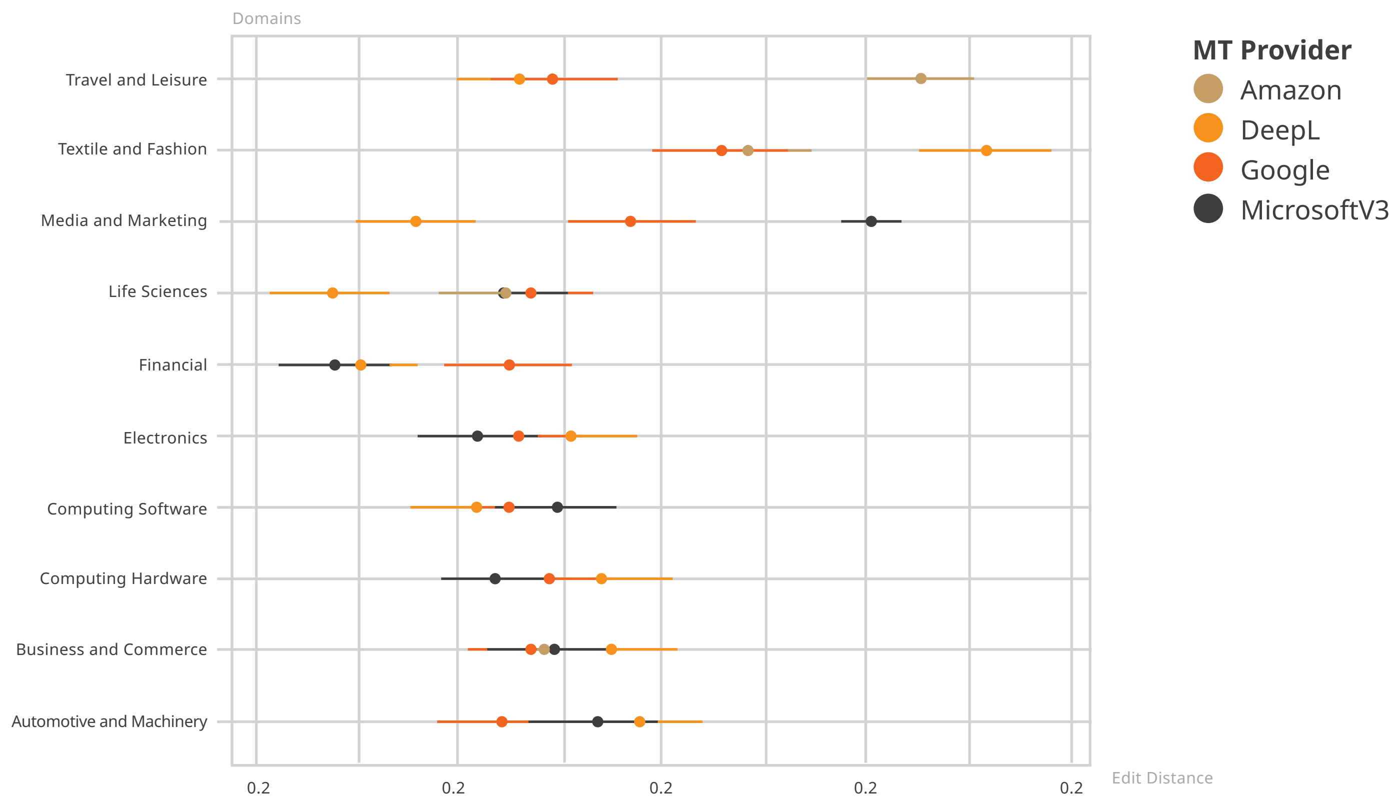


Figure 3. The average edit distance per domain and MT provider

Comparison of Machine Translation Engine Quality

How do the main engines perform against one another overall? In figure 4, we compared the output quality between the five major engines from May 2018 through December 2022 for German, Spanish, Russian, and Chinese using the inverse edit distance.

The edit distance measures the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation. The inverse edit distance means the higher the resulting number, the better the quality.

For figure 5, we used trend lines for each engine, which provides interesting insights. Figure 4 and figure 5, taken together, demonstrate engine quality among the top MT providers converging.

According to our analysis:

- Microsoft Bing aggressively catches up with the leaders
- Amazon and Google trend lines are almost perfectly parallel
- As a technology, Machine Translation did not significantly improve during 2022

By the beginning of 2023, the difference between the major engines is minimal. Suppose the Neural Machine Translation (NMT) paradigm continues to be dominant and MT providers continue to invest at a similar rate. In that case, we anticipate that MT engine performance will ultimately converge within the year, though there may be some differences between language pairs and domains.

By the beginning of 2023, the difference between the major engines is minimal.

Performance of MT Engines

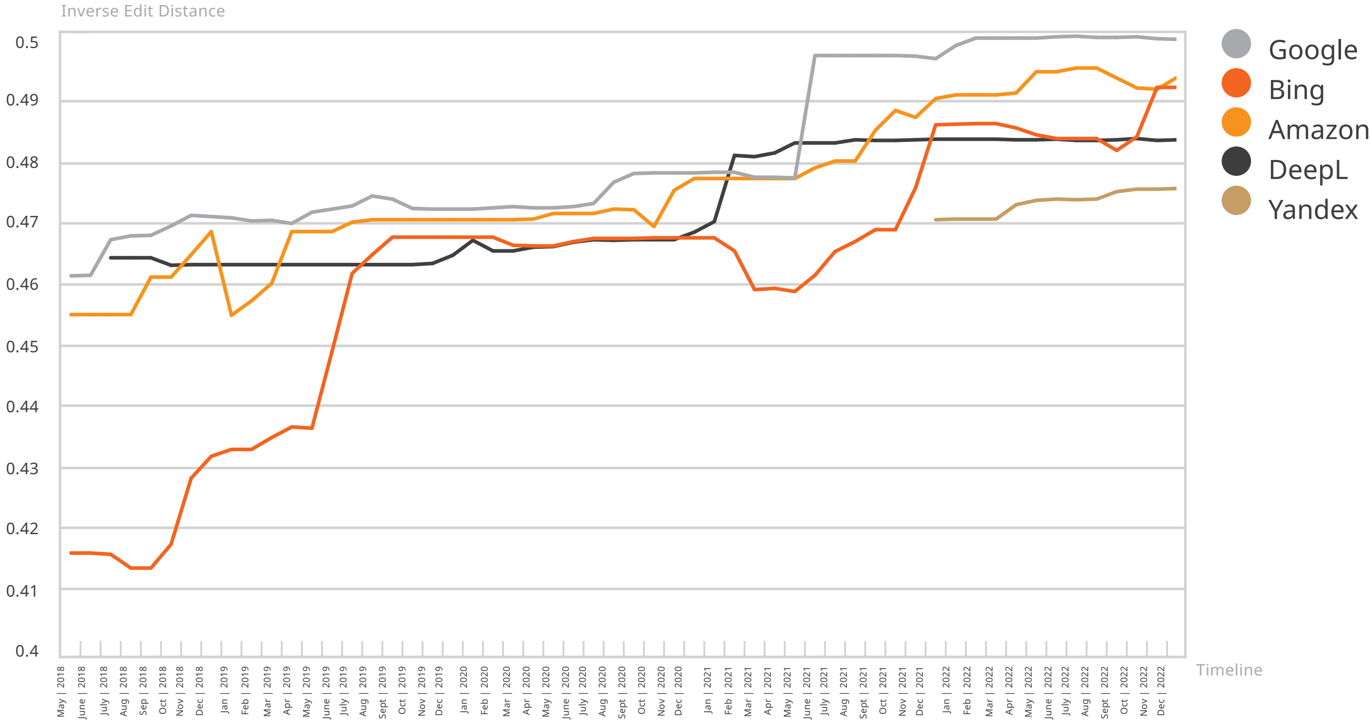


Figure 4. A comparison of overall MT quality based on the inverse edit distance

Performance of MT Engines: Trendlines

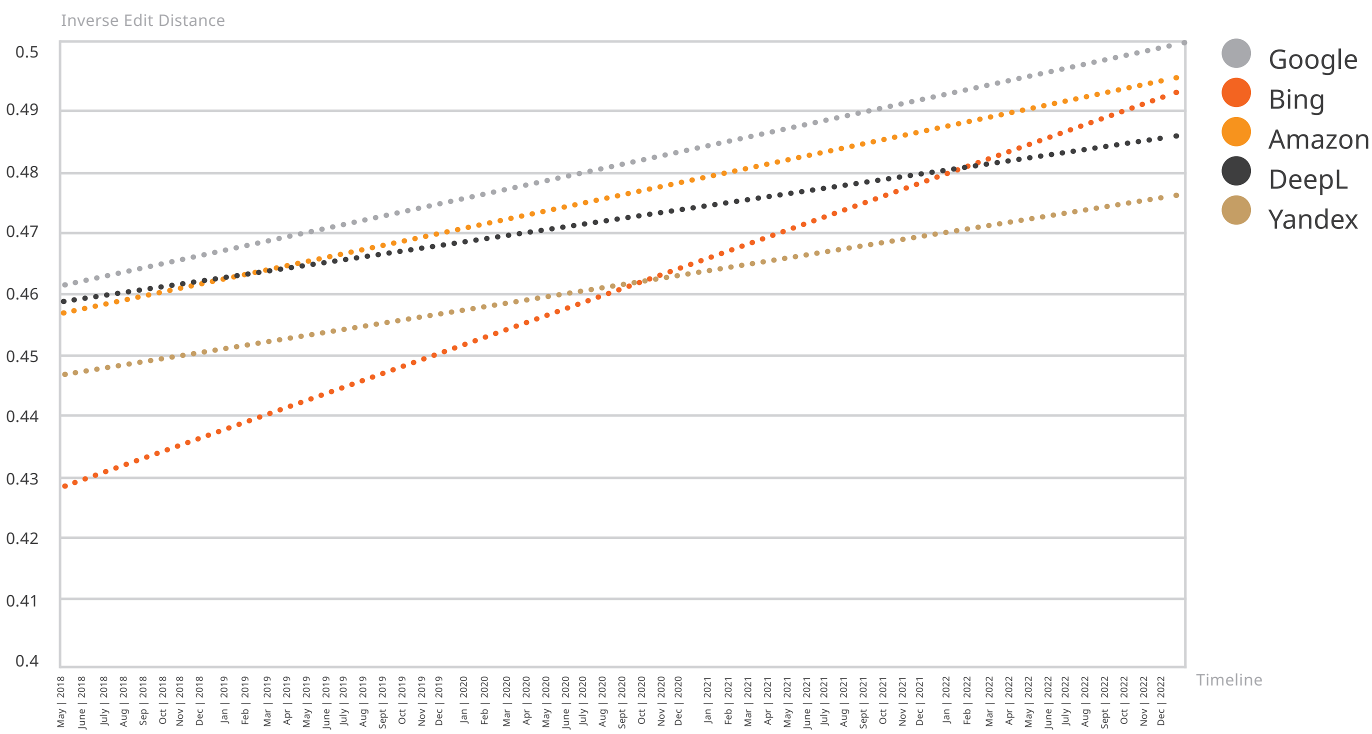


Figure 5. Trendlines of the five major MT engines' performances

Comparison of Machine Translation Engine Quality per Language

How did the main engines perform against one another in 2022, specifically for German, Spanish, Russian, and Chinese? We measured quality based on the inverse edit distance.

The edit distance measures the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation.

The inverse edit distance means the higher the resulting number, the better the quality.

As shown in Figure 6:

- There were minimal MT improvements overall, as reflected by the scale used to measure the inverse edit distance
- Microsoft Bing made minor improvements in German, Spanish, and Chinese during October/November
- 2022 proved to be a flat year

We can conclude that Neural Machine Translation has hit a plateau. A new iteration will be necessary for MT to make significant quality gains.

Performance of Machine Translation Engines per Select Languages via Inverse Edit Distance

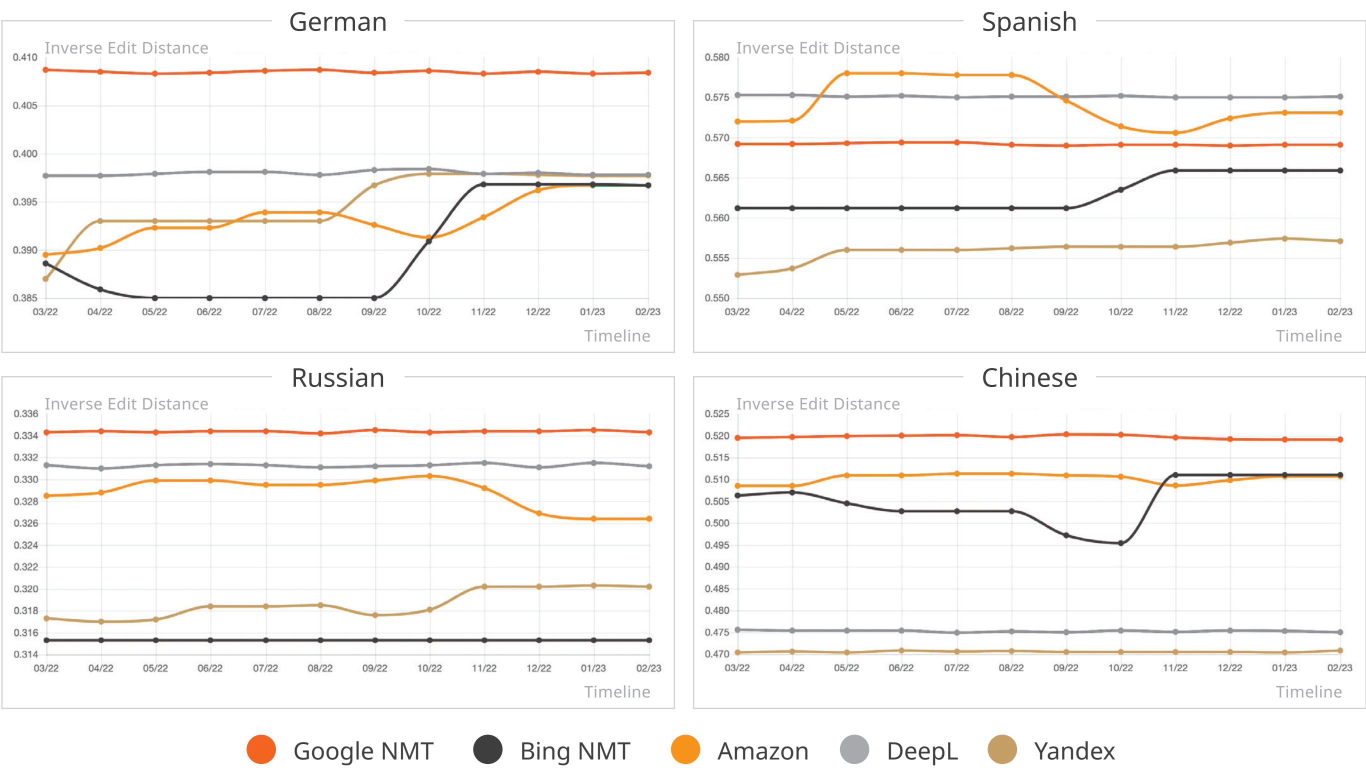


Figure 6. A comparison of MT quality per language based on the inverse edit distance

We can conclude that Neural Machine Translation has hit a plateau.

Comparison of Machine Translation Engine Quality per Domain

How did the main engines perform against one another in 2022 for specific domains? We measured quality based on the inverse edit distance, as shown in Figure 7.

The edit distance measures the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation. The inverse edit distance means the higher the resulting number, the better the quality.

According to our analysis:

- **Machine Translation** performs better for procedural content than more creative content because procedural content is typically more straightforward and easier for machines to process. As such, it was no surprise that MT quality was better for the Automotive and Computing Software sectors than for the Retail, Marketing, and Travel & Tourism industries since they all have more complex content. These results are similar to and amplify the results we found when measuring the automated translation performance based on domain and the average edit distance. (See Figure 2.)
- **Google MT** and DeepL had stable performances over the year compared to Amazon and Bing.
- **Amazon** did not improve significantly during the period we analyzed. Several peaks and valleys brought it back to its starting point by the end of the time we measured, except for two domains: Legal & Law and Media, Advertising & Marketing. For these two sectors, we observed a positive evolution that, in the case of Legal & Law, led Amazon to outperform its competitors slightly and for Media, Adverting, & Marketing, enabled them to co-lead the ranking with Google.
- **Bing** clearly led two domains intrinsically related to its parent company, Microsoft: Computing Software and Financials. Over the year, this MT engine demonstrated significant improvement, rising from the third position to claim the top spot in the ranking. No other MT engine has shown a comparable level of advancement in the ranking within any domain.
- **Yandex** is the only MT provider analyzed that did not lead any domain. It ultimately surpassed and performed better than DeepL by the end of 2022 in only two domains: Automotive and Computing Software.

Performance of Machine Translation Engines per Select Domains

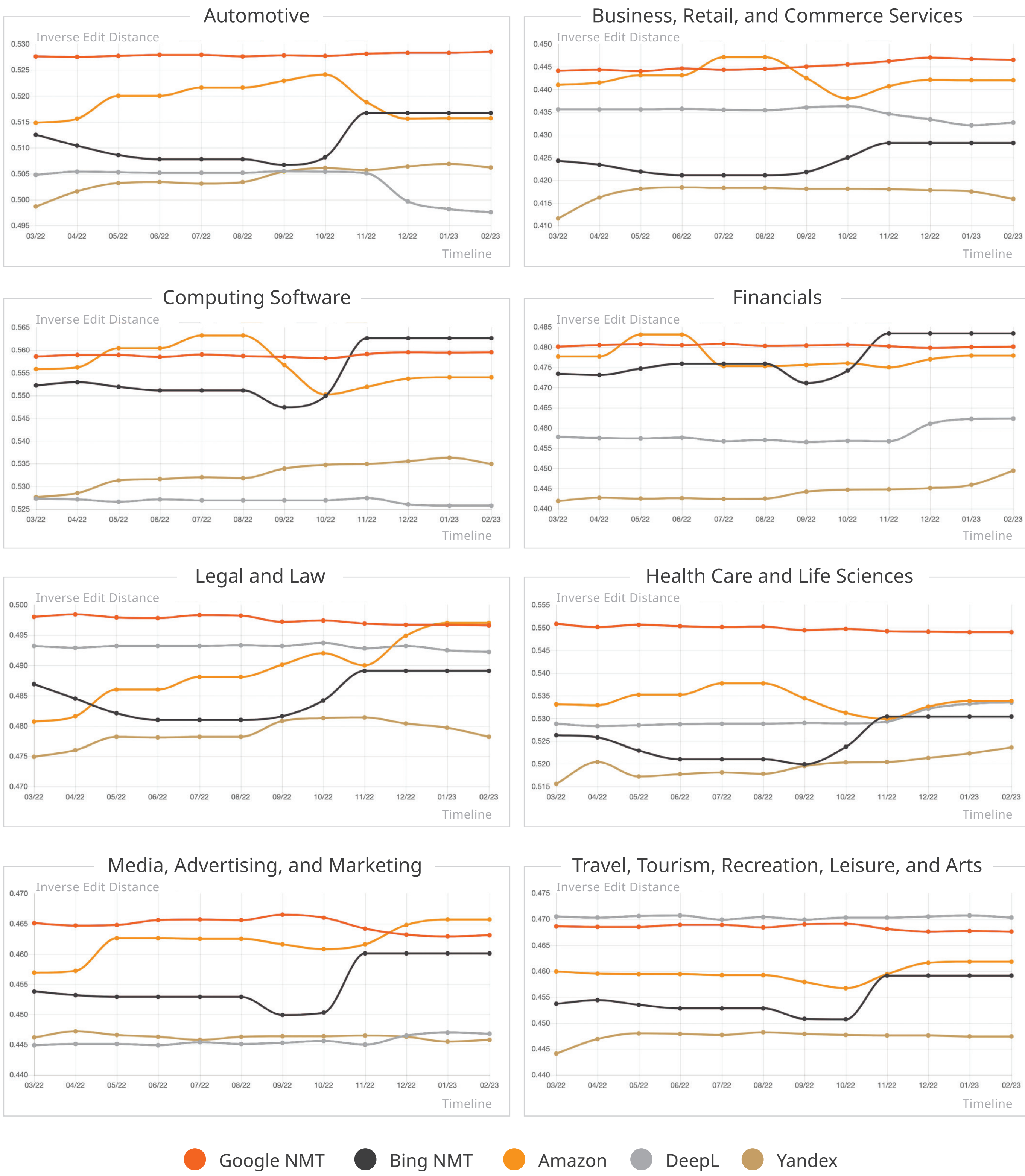


Figure 7. A comparison of MT quality per domain based on the inverse edit distance

How Can I Get the Most Out of Machine Translation?

To get the most out of MT technology, consider taking the following steps:

- Identify the ease with which MT engines translate specific languages — the **machine translatability** or m-translatability of languages — to help inform your MT strategy and decide which markets to pursue. We've done the legwork for you.
- Use terminology effectively to improve output quality for any domain, even the ones that MT often bungles.
- Consider when to conduct MT customization vs. MT training. Each method can improve MT output, but they are not interchangeable.

Machine Translatability

Identifying how challenging it is for engines to handle specific language pairs will help you allocate your budget when planning translation costs across languages. You will better understand which language pairs will require more effort to translate. Having insight into language complexity can help support your business decisions and help you answer the following questions:

- Should a larger share of the budget be used to post-edit more complex language pairs?
- Will light post-editing or focused post-editing, which targets only critical areas of the content for post-editing, be sufficient for some languages when dealing with a tight budget? For which languages should I use these post-editing methods?
- Should my company add language ranking to business and cultural factors when considering how to best allocate its budget, particularly for low-budget projects? If a culture accepts a lower quality level, should my company translate into a language with a low m-translatability ranking?

Ranking languages by translatability is not a straightforward process; however, we can use different metrics for evaluation. Edit Distance, the number of changes a post-editor makes to ensure the final text has a human quality, can provide a sense of MT complexity and translatability for each language pair. To help you compare languages, we've ranked the machine translatability or m-translatability of the top 28 target languages from English.

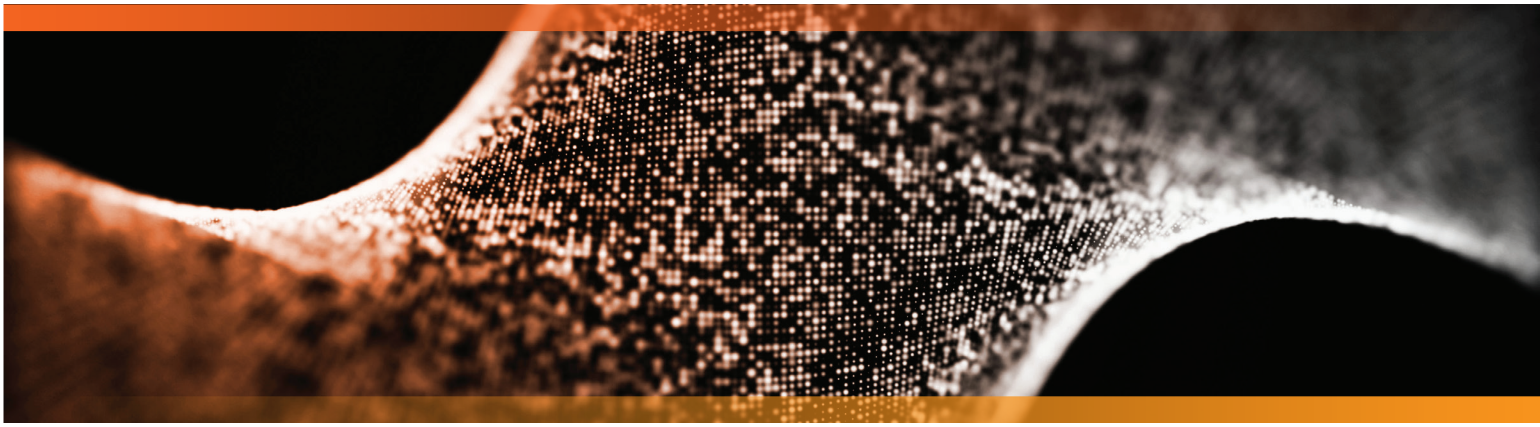
As shown in Table 1, most Romance languages, such as Portuguese, Spanish, French, and Italian, require fewer changes to reach high-quality levels when translated from English. We identified these target languages as the easiest for machines to handle, and they took the first four spots in our m-translatability ranking.

Hungarian and Finnish — two Uralic languages — are more complex languages; they placed last in our ranking, taking the 27th and 28th spots. Estonian, another language in the same family, is also among the more complex languages. Based on millions of sentences processed by Lionbridge, these results underscore the importance of language families in MT results.

While language comparison has limitations, the ranking can provide interesting insights to manage multilingual projects better.

Table 1. Language M-translatability Ranking

Rank	Language (From English)
1	Portuguese
2	Spanish
3	French
4	Italian
5	Chinese (Simplified)
6	Dutch
7	Danish
8	Japanese
9	Greek
10	Romanian
11	Thai
12	Norwegian
13	German
14	Swedish
15	Turkish
16	Slovak
17	Hebrew
18	Latvian
19	Polish
20	Chinese (Traditional)
21	Lithuanian
22	Czech
23	Arabic
24	Estonian
25	Korean
26	Russian
27	Hungarian
28	Finnish



The combination of trained MT engines, glossary customization, and the identification of preprocessing and post-processing rules ensure MT output contains proper terminology and is similar in style to the customer's documentation.

Terminology To Improve Domain Performance

As noted, generic MT engines can put out erroneous translations; they can especially cause undesired results for specific domains from a terminological point of view. The impact can be particularly harmful to the medical and legal fields. The effective use of terminology can enable you to improve the quality of MT and achieve accurate, consistent translations no matter what your subject matter is.

It's imperative to train customized MT systems with domain-specific bilingual texts that include specialized terminology. Still, when engines are trained with specialized texts, accurate translations cannot be guaranteed if the terminology is not used consistently. Research in this area proposes to inject linguistic information into Neural Machine Translation (NMT) systems. Implementing manual or semi-automatic annotation depends on available resources, such as glossaries, and constraints, such as time, cost, and availability of human annotators.

Lionbridge's **Smairt MT** allows the application of linguistic rules to the source and target text and the enforcement of terminology based on Do Not Translate (DNT) and glossary lists added to a specific profile to address **Machine Translation terminology**. We help our customers create and maintain glossaries, regularly refined to include new, relevant terms and retire obsolete terminology. When glossaries are created once in Smairt MT, they can be used for all the MT engines, saving time and money.

Using glossaries for MT projects is more complex than it may seem. Glossaries, if used inappropriately, can negatively affect the overall quality of Machine Translation. The best way to follow terminology in MT is through MT training. The combination of trained MT engines, glossary customization, and the identification of preprocessing and post-processing rules ensure MT output contains proper terminology and is similar in style to the customer's documentation.

MT Customization vs. MT Training

MT customization and MT training can help you get more out of your MT output, but you must be intentional about when to apply these methods. Table 2 provides an overview of **Machine Translation customization vs. Machine Translation training** and offers some considerations when evaluating each method.

Table 2. Machine Translation Customization vs. Machine Translation Training

	MT Customization	MT Training
What it is and how it works	An adaptation of a pre-existing Machine Translation engine with a glossary and Do Not Translate (DNT) list to improve the accuracy of machine-generated translations	The building and training of an MT engine by using extensive bilingual data from corpora and Translation Memories (TMs) to improve the accuracy of machine-generated translations
What it does	Improves MT's suggestions for more accurate output and reduces the need for post-editing	Improves MT's suggestions for more accurate output and reduces the need for post-editing
Specific benefits	Enables companies to adhere to their brand name and terminology and achieve regional variations	Enables companies to attain a specific brand voice, tone, and style and achieve regional variations
The risks of using it	The MT could make poor suggestions and negatively impact overall quality when executed improperly	MT training may fail to impact output if there is not enough quality data to train the engine; the MT could generate poor suggestions and negatively impact overall quality if inexperienced authors overuse terminology
When to use it	Ideal for technological and detail-oriented content and any content that requires: • Accurate translations of terminology • Regional variation, but you lack sufficient data for MT training	Ideal for highly specialized content, marketing and creative content, and any content that requires: • A specific brand voice, tone, or style • Regional variation, and you have enough data for MT training
Success factors	An experienced MT expert who can successfully manage input and output normalization rules, glossaries, and DNT	A minimum of 15K unique segments to adequately train the engine
Cost considerations	There is a one-time cost to update the profile that goes into the MT engine and some ongoing costs to maintain a glossary over time; costs are relatively inexpensive when factoring in the potential benefits and are typically lower than MT training costs	There are costs associated with the first training and potential costs for additional training, which may be considered over time if the MT performance monitoring indicates room for improvement; MT training can be worth the investment in certain cases when factoring in the potential benefits

CONCLUSIONS AND THE OUTLOOK FOR MACHINE TRANSLATION

What can we conclude about the state of Machine Translation from the 2022 data and surprising results that mainly showed stagnant quality performances for the year? The technology is mature and will continue to attain widespread adoption as it has unequivocally proven its value as a business-grade technology.

People recognize the technology's usefulness for almost any translation case — with or without human intervention and hybrid approaches. Indeed, according to Global Market Insights, the translation market size is projected to grow at a Compound Annual Growth Rate (CAGR) of 30% from 2022 to 2030. Companies will increasingly embrace MT — including those businesses in traditionally MT-resistant domains, such as games and life sciences. The ability to fully capitalize on the technology — in conjunction with the use of AI-driven technology that automates workflows and translator selection — will position companies to increase their content velocity, produce captivating multilingual content that is always on brand, grow their markets, and thrive in what has become a brutally competitive digital market.

What is the Future of Machine Translation?

2022 Machine Translation results made us question the current Neural Machine Translation paradigm.

- Is the NMT paradigm reaching a plateau?
- Is a new paradigm shift needed, given the engines' inability to make significant strides?
- What could be next?

We're betting that Large Language Models (LLMs) — with their massive amounts of content, including multimodality and multilingualism — will have something to do with a future paradigm.

Why do we think this? Because of the results of our ground-breaking analysis that compared **ChatGPT's translation performance** with the performance of MT engines.

OpenAI's ChatGPT produced inferior results than designated MT engines — but not by much. Its performance was nothing short of remarkable. GPT-4 even surpassed one major Neural Machine Translation engine in one instance and one language pair. These results undoubtedly have implications for the future of Machine Translation.

Why Is a New Machine Translation Paradigm Likely Underway?

Current MT engine trends give us a sense of déjà vu. During the end of the Statistical Machine Translation era, which NMT replaced, there was virtually no change in MT quality output. In addition, the quality output of different MT engines converged. These things are happening now.

While NMT may not be replaced imminently if we believe in exponential growth and accelerating returns theories, consider Rule-based MT's 30-year run and Statistical MT's decade-long prominence, and note that NMT is now in its sixth year, a new paradigm shift is near.

What Could Be the Next Machine Translation Paradigm?

Important advances to LLMs during 2022 have primed the technology for 2023. LLMs are generic models that are trained to do many things. However, we saw some dedicated — or fine-tuned — LLMs make essential advancements in some specific areas by the end of 2022. These developments position the technology to perform translations with some additional training.

For instance, take ChatGPT. OpenAI fine-tuned this model to conduct question-and-answer dialogues while still being able to do anything generic LLMs do. The company made even more improvements with its GPT-4 model. Expect to see more LLMs fine-tune for translation as time goes on.

In What Way Would Large Language Models Need to Be Fine-Tuned To Handle Translation?

It would be more probable to use LLMs to execute translation if the machines are trained with a more balanced language corpus. GPT-3's training corpus was 93% English; only 7% constituted corpora from all the other languages. We guess that GPT-4 has more training data for more languages, which could explain its improved quality output. A more language-balanced corpus may be the basis for building a fine-tuned model on top of LLMs specialized in translation.

Another interesting aspect of this hypothetical new MT paradigm based on LLMs is the multimodality trend. We may train LLMs using linguistic and other training data, such as images and video. This type of training may provide additional world knowledge for a better translation.

Would Large Language Models Be a Good Alternative to a Neural Machine Translation Paradigm?

To assess the promise of LLMs to replace the NMT paradigm, we compared the translation quality of ChatGPT and GPT-4 to the performance of the five major MT engines we use in our MT Quality Tracking. As shown in Figure 8, ChatGPT performed almost as well as the specialized engines. And as shown in Figure 9 below, GPT-4 outperformed Yandex in the English-to-Chinese language pair.

How Did We Assess the Quality of ChatGPT vs. Generic MT Engines?

We calculated the quality level of the engines based on the inverse edit distance using multiple references. The edit distance measures the number of edits a human must make to the MT output for the resulting translation to be as good as a human translation. For our calculation, we compared the raw MT output against 10 different human translations — multiple references — instead of just one human translation. The inverse edit distance means the higher the resulting number, the better the quality.

A more language-balanced corpus may be the basis for building a fine-tuned model on top of LLMs specialized in translation.

Multiple References Evaluation via Inverse Edit Distance

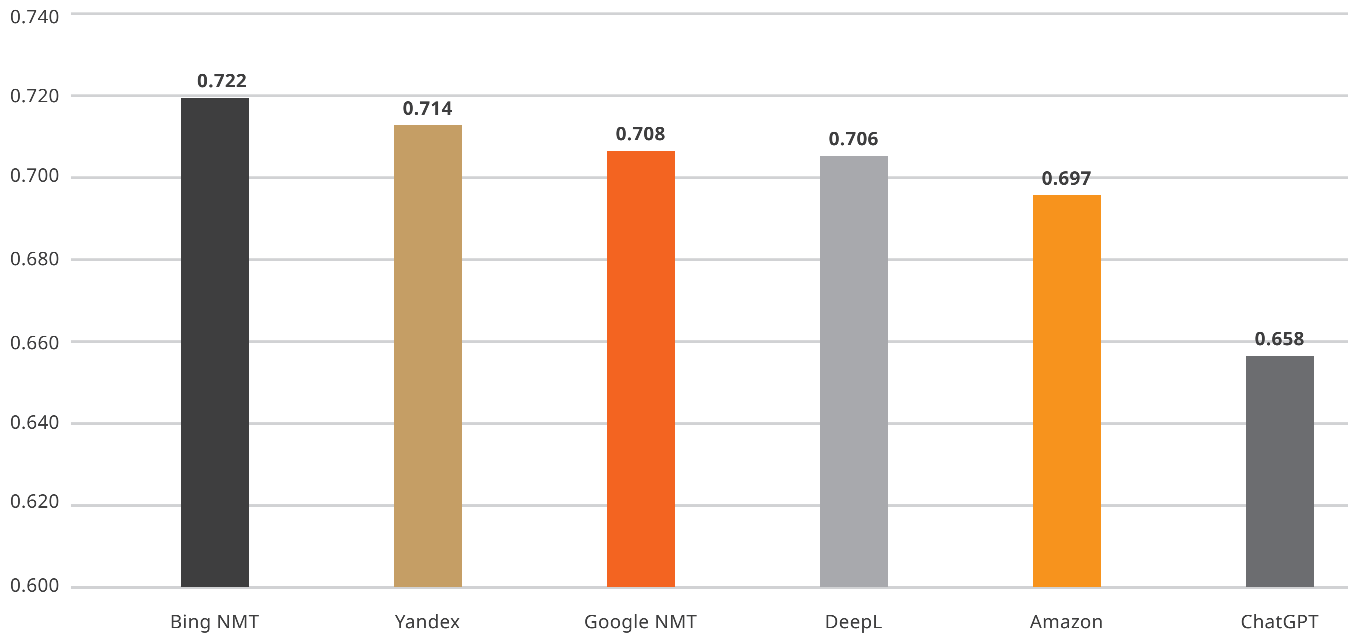


Figure 8. Comparison of automated translation quality between ChatGPT and the major Machine Translation engines based on the inverse edit distance using multiple references for the English-to-Spanish language pair.

English-to-Chinese Translation Quality

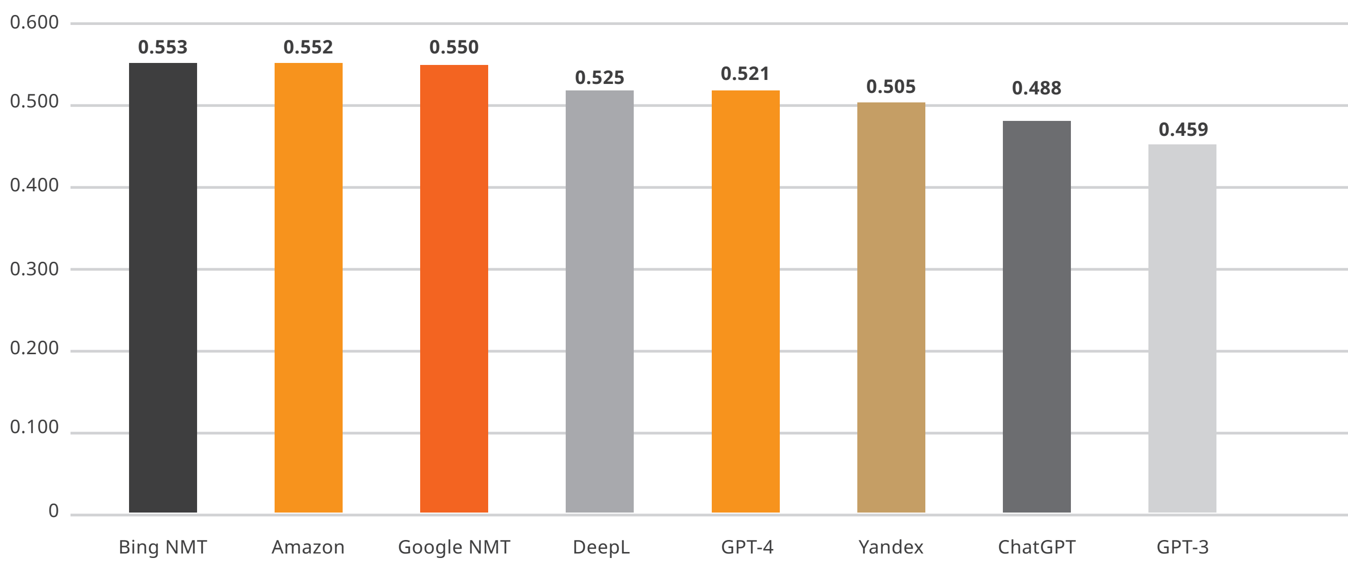


Figure 9. Comparison of automated translation quality between GPT models and the five major Neural MT engines based on the inverse edit distance using multiple references for the English-to-Chinese language pair.

Why Are the LLM Translation Results Noteworthy?

The results of our comparative analysis are remarkable because the generic model has been trained to do many different Natural Language Processing (NLP) tasks as opposed to the single NLP task of translation that MT engines have been trained to do. And even though GPT has not been specifically trained to execute translations, its quality is exceptional.

How Might Machine Translation Evolve as a Result of Large Language Models?

Given the growth of LLMs — based on the public's attention and the significant investments tech companies are making in this technology — we may soon see whether MT will start adopting a new LLM paradigm.

MT may use LLMs as a base but then fine-tune the technology specifically for Machine Translation. It would be like what OpenAI and other LLM companies are doing to improve their generic models for specific use cases, such as making it possible for the machines to communicate with humans in a conversational manner. Specialization adds accuracy to the performed tasks.

What Does the Future Hold for Large Language Models in General?

The great thing about Large Language "Generic" Models is that they can do many different things and offer outstanding quality in most of their tasks. For example, DeepMind's GATO, another general intelligence model, has been tested in more than 600 tasks, with State-of-the-Art (SOTA) results in 400 of them.

Two development lines will continue to exist — generic models, such as GPT and GATO, and specialized models for specific purposes based on those generic models. The generic models are important for advancing Artificial Generic Intelligence (AGI) and possibly advancing even more impressive developments in the longer term. Specialized models will have practical uses in the short run for specific areas.

One of the remarkable things about LLMs is that both lines can progress and work in parallel.

What Are the Implications of a Paradigm Shift in Machine Translation?

As the current Neural Machine Translation technology paradigm reaches its limit and a new dominant Machine Translation technology paradigm emerges — likely based on LLMs — we anticipate some changes to the MT space. Most of the effects will benefit companies, though we expect additional challenges for companies seeking to execute human translations.

Here's what to expect:

Enhanced Quality

There will be a leap in Machine Translation quality as technological advancements resolve longstanding issues, such as language formality and other quality issues pertaining to tone. LLMs may even solve MT engines' biggest problem: their lack of world knowledge. This achievement may be made possible through their multimodality training.

Technologists not only train modern LLMs with vast amounts of text, but they also use images and video. This type of training enables the LLMs to have a more linked knowledge that helps the machines to interpret the texts' meaning.

Increased Content Output and a Reduced Supply of Top-Notch Translators

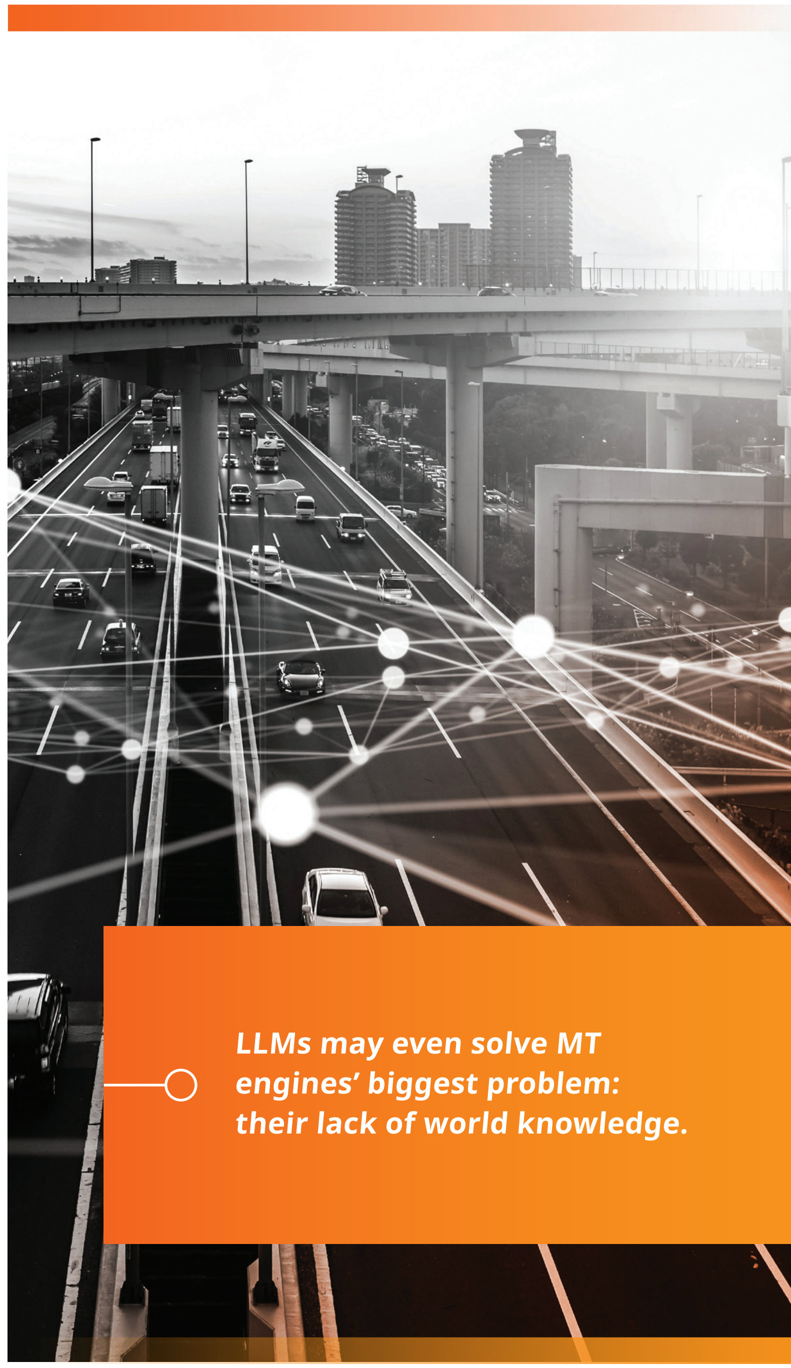
Companies will be able to create more content faster, and content creation will outpace the growth of the pool of translators capable of translating this content. Even with improved MT and increased translator productivity, the translation community will continue to struggle to satisfy translation demands.

Increased Adoption of Machine Translation

As the new technology paradigm becomes available and the quality of Machine Translation improves, the demand for translation services will continue to grow, increasing adoption in more situations and use cases.

The Use of Machine Translation to Enhance Customer Experiences

With improved MT quality and the need for more personalized and tailored customer experiences, companies will use MT more frequently to improve their global customers' digital experiences and create stronger relationships.



LLMs may even solve MT engines' biggest problem: their lack of world knowledge.

What's the Bottom Line?

Technology companies are demonstrating immense interest in LLM technology. **Microsoft invested $10B** in OpenAI. Nvidia, Google, and other companies are also investing heavily in LLM and AI technology.
To fully understand **GPT and localization** and the real potential of Large Language Models to generate value for the language industry, we need to:

- Conduct real-world tests at scale to evaluate the error rate for each type of localization and editing task
- Analyze the detailed macro and micro user journeys occurring within the localization value chains and identify where they are likely to be disrupted with this type of text automation
- Understand how to prompt and provide relevant context to GPT and other LLMs at scale, and document pitfalls and best practices
- Develop new automation and human-in-the-loop editing workflows, inventing what post-editing and QA will mean tomorrow with such an AI in the loop
- Design new automation and User Experience (UX) interaction contexts for both localization agents and customers for each possible improvement opportunity
- Ensure that the economics of licenses, deployment costs, and maintenance makes sense for our business

We are intrigued by what the future holds. We will continue to evaluate LLMs, help you stay current on this exciting evolution, and enable you to gain from it.

Technology companies are demonstrating immense interest in LLM technology.

LIONBRIDGE MACHINE TRANSLATION EXPERTS

Rafa Moral

Vice President, Innovation

Rafa oversees R&D activities related to language and translation, including Machine Translation initiatives, Content Profiling and Analysis, Terminology Mining, and Linguistic Quality Assurance and Control.

Yolanda Martin

MT Specialist

Yolanda is responsible for the creation of customized translation models, as well as quality analysis and the development of strategies to fine-tune them. In parallel, she collaborates with the R&D department to develop new linguistic tools and resources.

Thomas McCarthy

MT Business Analyst

Thomas ensures Lionbridge customers and stakeholders obtain maximum benefits from MT-related technologies, services, and consultancy.

Lionbridge experts make it easy for customers to implement Machine Translation effectively and achieve business gains. Lionbridge is closely evaluating emerging technologies to enable customers to benefit further as the Machine Translation space undergoes rapid change.

To learn more about how Lionbridge can help you fully capitalize on automated translations, contact our team today.

ABOUT LIONBRIDGE

Lionbridge partners with brands to break barriers and build bridges all over the world. For over 25 years, we have helped companies connect with their global customers and employees by delivering translation and localization solutions in 350+ languages. Through our world-class platform, we orchestrate a network of passionate experts across the globe who partner with brands to create culturally rich experiences. Relentless in our love of linguistics, we use the best of human and machine intelligence to forge understanding that resonates with our customers' clients. Based in Waltham, Massachusetts, Lionbridge maintains solution centers in 24 countries.

LEARN MORE AT
LIONBRIDGE.COM

LIONBRIDGE