AI ?????????????

AI ?????????????

???Shasha Zhao



Localizing content delivers strong business benefits. According to white paper released by Centific and Nimdzi Insights, companies that localize the user experience see a 100%–400% increase in sales, and by localizing into just 10 languages, a brand's message will effectively reach 90% of online customers. As brands appreciate the business benefits of localization, they are increasingly turning to artificial intelligence to make localization more effective. This is true especially for large, complex, multinational businesses that need to adapt multiple products and services across hundreds of geographic markets and cultures. In fact, we believe AI can unlock hyperlocal and hyper-personalized experiences that are culturally aware, as my colleague Ilia Shifrin blogged recently. Let's take a closer look at how AI localization is different from traditional forms of localization.

AI Localization Defined

With traditional localization, people with the help of software, adapt content for a local market, ranging from translating language to modifying an image to be more appropriate for a local culture. We often think of localization as the process of adapting a product or content to a specific market, making it local look and feel.

As we have discussed on our blog, AI localization as defined by Centific fully encompasses language and voice AI data processing. It is the process to localize vast quantities of data, such as text, emoji,

acronyms, and abbreviations often seen in the online chatting, voice, and more. An AI engine requires high-quality training data that reflects the way that people speak, sound, and express themselves.

With both traditional and AI localization, people are needed, and the goal is to make a product or experience relevant to a person. By means of AI localization, people train AI engines with vast pools of hyperlocal content and in-market user experiences-generated data.

Characteristics of Traditional Localization vs AI Localization

Language Style

With traditional localization, people need to share language that is more precise based on source. The content needs to be highly accurate. For example, repeated content or highly matched sentences in one product manual need to be translated in a consistent way. Translators need to work with more context, such as complete sentences, to understand what they are localizing. But with AI localization, each training element is relatively independent. Hence context does not need to be explained. A translator translates a single string of data without any particular context. This means that the content can come from a number of sources, however formal and informal -- emails, short messages, press releases, websites, casual conversations, and so on.

AI localization can also cover far more broad forms of expression, however formal and informal, including slang, local dialects, and even incomplete sentences or abnormal syntactic structure sentences. It requires the translation style to be more natural, colloquial and vivid, and informal. Translators should use their best judgment to imagine what the real scenarios are involved, and to translate such texts as appropriately as possible based on the logical inference. Taking one Chinese-to-English string as an example, a translator might work with a source such as “???????”. This sentence uses the spoken language style in Chinese. It is not clear in meaning without any contextual references. "??" is the combination of two pronouns "I" and "You" put together and placed at the end of the sentence, which sounds a little weird. We guide our translator to imagine the real scenario and translate it in one way. The suggested English translations should be “You text me just now, right?” In contrast to traditional localization, with AI localization, a similar source strings of training data can be translated in various ways to make language more diverse.

Process

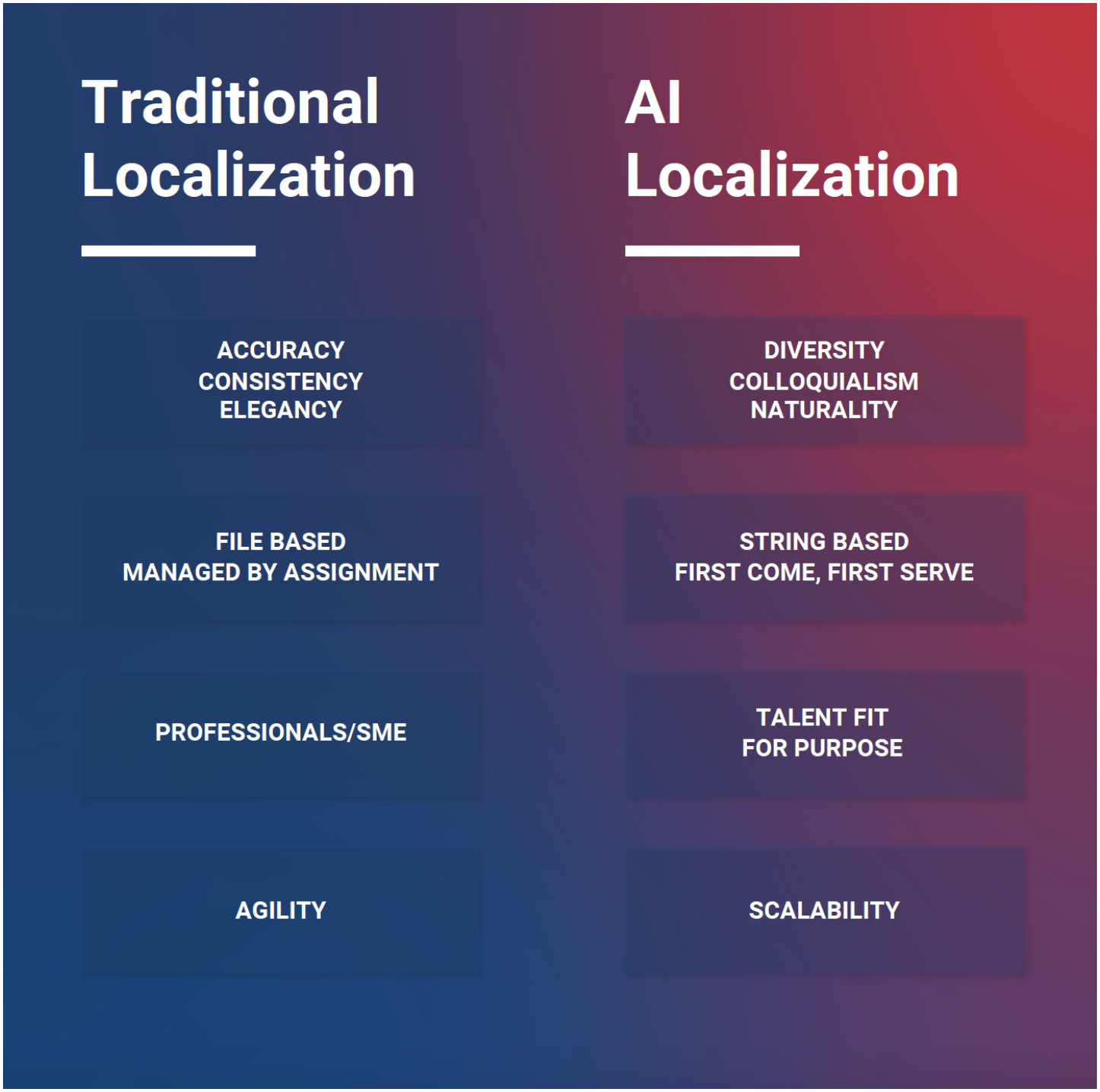
With traditional localization, people translate content following a sequential order of steps consisting of translation -> editing -> QA. The people involved may rely on either a TMS (translation management system) or a CAT (computer-assisted translation) tool. For traditional localization to work, a translator needs to examine a chunk of content and ensure that it is completely done before passing along a file to another person for review. You need a single person managing a single file to translate the content at one time. You cannot let multiple people work on the same source file because the production is organized under exclusive file ownership. In addition, consistency in style and terminology is a serious consideration.

With AI localization, due to the nature of data independence, we can perform string-based translation, which means multiple resources can process strings of data from one single file without the context needing to be understood.

Resources

Traditional Localization requires professional linguists. We prefer resources with abundant translation experience in certain areas like information technology, legal, marketing, medical and so on. They rely on CAT tools, translation rules, and general industry guidelines. The linguists may need to have in-depth knowledge of the content as well as the language – for example, a linguist translating a software user manual needs to understand the software as well as the language.

With AI Localization, humans in the loop do not need to have as much specific domain expertise, but it's more important that they possess a strong understanding of language, local cultures, and local forms of expression such as social media.



Trends: Where Traditional Localization and AI Localization Are Heading

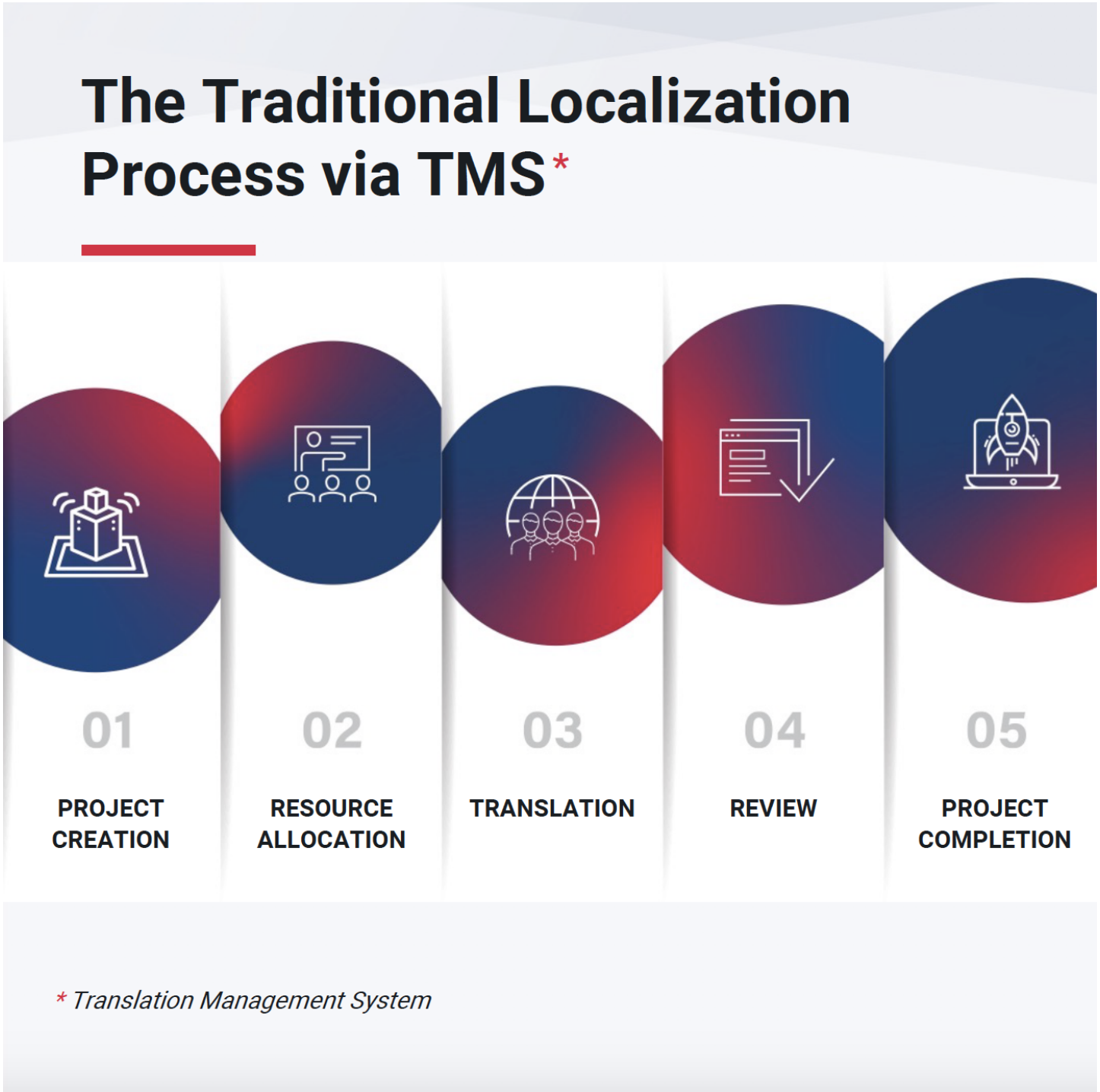
Businesses that rely on traditional localization are demanding shorter turnaround times. As a result,

traditional localization practices are becoming more agile. Translation turnaround times are shorter and shorter, and translation cadence is more and more frequent, but the volume of each translation drop may become smaller.

Meanwhile, for training purposes, an AI engine requires even more data at a far greater scale. For example, for one of our clients, Centific delivered 30 million words of translation within eight weeks. Consequently, more people are needed to train AI to deliver a better result.

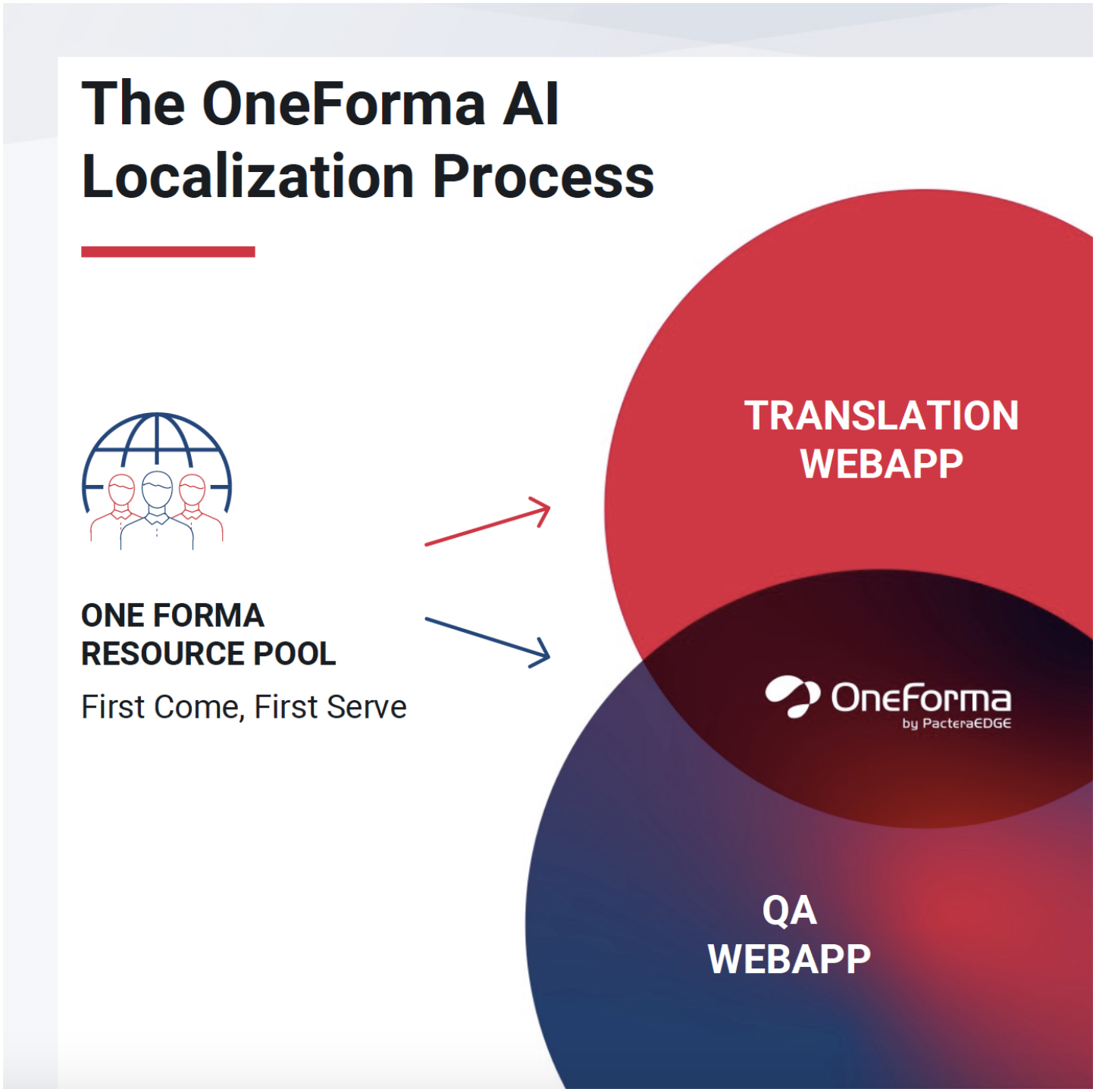
Process Evolution in AI Localization and Benefits

Let's see how AI Localization represents a change in the process of localizing data:



The difference between traditional localization process and AI localization is that the former has to

follow a sequential order of steps whereas in the latter the steps can be performed parallelly by multiple resources. The traditional localization process is linear, while the review step comes after the translation step is completed. AI localization enables multithreading wherein the translation resources share the content string with a quality assurance resources in a real-time manner.



Generally speaking, traditional localization needs a project manager to allocate resources to the task. For AI localization, Centific creatively applies the “first come, first served” approach. Pre-onboarded people with desired content and language expertise can easily access the Centific OneForma platform to translate as many strings of data as they are capable of whenever the project is available.

This process evolvement can provide high-quality work in collaborative fashion, extensively engage with a broader resource pool and dramatically increase the production throughput.

Why Centific

Centific has provided AI localization to multiple clients. We clearly understand clients' needs and expectations, such as making translation guidelines and quality measurements suitable and helpful for training an AI engine.

Centific also has a robust technology platform to support AI Localization, OneForma. OneForma is an enterprise-grade and fully customizable platform developed by Centific to manage machine learning and artificial intelligence training data needs such as recruiting/user management, workload distribution, and quality management. It facilitates our AI Localization process. The OneForma platform is customized to retain major advantages of a TMS such workflow automation, online project management, and embedded QA checks. Additionally, OneForma also achieves qualitative, advanced string-based translation work, which TMS doesn't support.

Thirdly, Centific has built up a fundamental resource management framework specially for AI localization, from initial screening, certification, and onboarding training to performance monitoring, feedback sharing, and resources ranking. Additionally, we have maintained a strong resource pool with the desired language expertise. We have more than 200,000 resources already onboarded in OneForma working on our daily projects.

Centific has a strong global team to support the scale and speed required to succeed with AI localization.

-
-
-
-
-