Assessment of Online Translation Tools for Indigenous Languages in the Ocean Domain

Final Report | May 28, 2023

Authors: Maria van Veldhuizen, Sigmund Kluckner, Jay Pearlman[1]

The authors' acknowledgements and thanks go to Michael Køie Poulsen, Françoise Pearlman, Stein Sandven, and Siverth Amondsen for their help and advice, and to the CAPARDUS Project[2] for commissioning this study.

[1] mariavveldhuizen@gmail.com, sigmund@kluckner.eu, jay.pearlman@fourbridges.org

[2] The CAPARDUS project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 869673.

Introduction and scope

Several knowledge communities focusing on Arctic environment, maritime and ocean sciences have been creating, publishing, and sharing a multitude of information artifacts (reports, best practices, standards, white papers, etc.). This information could be of immense help to indigenous peoples living in the Arctic and in coastal areas around the world, but most of it is currently available only in major world languages; primarily English and French.

This project reviews and assesses the potential for emerging translation tools and technologies to provide an automated way to facilitate broader access to these information artifacts. In addition to providing an overview of these tools, the project also assesses the general capabilities of the most widely used tools to produce accurate, usable translations, by working with native speakers to check word usage, grammar, overall understandability, etc. The project has started with translations from English to major languages to assess the overall capabilities and reliability of the different tools and is working towards including indigenous languages depending on availability of tools.

This report offers an explorative study of translation tools and a preliminary assessment of 1) their general quality and relevance for the ocean domain; and 2) their capabilities to provide translations to indigenous, rare, and/or low-resource languages. The assessment is non-exhaustive and provides a snapshot view of translation technologies in May 2023; it will likely become outdated quickly due to the speed of the development of new Artificial Intelligence (AI) technologies, especially since the advent of generative language models like BERT, GPT and LLaMA.

The work has been carried out in April and May 2023, in the lead-up to the CAPARDUS project meeting in May 2023.

The objectives of the CAPARDUS project are to establish a comprehensive framework for Arctic standards, enhance coordination of Arctic Practices, synthesize project results into requirements and recommendations, and develop engagement in Arctic standardization through workshops and research schools. The project reviewed existing standards, engaged with Arctic communities to co-develop the Comprehensive Framework (from a requirements perspective), and identify gaps and deficiencies of standards. The project also aimed to coordinate with other groups working on standardization and create a framework for an Arctic Practices System to make Arctic Practices more readily accessible. The project identified guidelines for the development of best practices and Arctic standards related to research, data, and technologies supporting safe and sustainable operations in the Arctic. The project addressed EU's Arctic policy, the Sustainable Development Goals (SDGs), the Paris Agreement, and provide guidelines for new economic opportunities in the Arctic.

Background

Introduction to Artificial Intelligence (AI) Tools

AI tools are machines or software that can perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, problem solving, and even exercising creativity. AI tools have evolved over time with advances in computing power, algorithms, and data availability. One of the main branches of AI is machine learning, which is based on algorithms that can learn from data and improve over time without explicit programming. Machine learning can be further divided into subfields such as deep learning, natural language processing, computer vision, and reinforcement learning.

AI tools are already in widespread use in various domains and for various purposes, such as voice assistants and smart speakers, chatbots, facial recognition, medical diagnosis, gaming, marketing, and automated language translation. The most recent breakthroughs in AI were the release of large language models (LLMs) and their "creative" outputs (most notably GPT-3 and 4) or image generation through textual input (DALL-E, MidJourney).

Computer-Supported and Automated Translation

Machine translation (MT) tools are software programs that can translate text or speech from one natural language to another. Machine translation tools can be customized for specific domains or purposes, such as legal, medical, or literary translation. Though machine translation tools have improved greatly over the past several years, they are not perfect and for most purposes still require human intervention or post-editing to improve their quality and accuracy. Where they are used without human intervention, such as on social media platforms, incorrect translations have occasionally caused problems.

Machine translation methods can be classified into four main types: rule-based, statistical, hybrid, and neural:

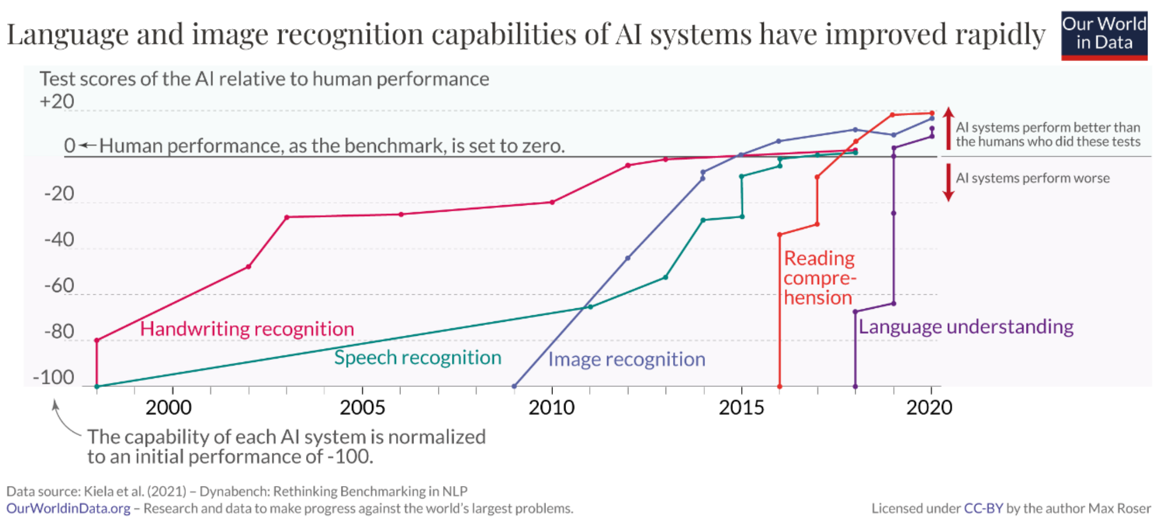
- Rule-based machine translation (RBMT) is the earliest form of machine translation that relies on linguistic rules and dictionaries to translate between languages. RBMT has the advantage of being consistent and predictable, but it also has the disadvantage of being rigid and unable to handle idioms, slang, or ambiguity.
- Statistical machine translation (SMT) is a form of machine translation that uses large corpora of parallel texts to learn probabilistic models of translation. SMT has the advantage of being adaptable and scalable, but it also has the disadvantage of being noisy and unreliable.
- Hybrid machine translation (HMT) is a form of machine translation that combines RBMT and SMT to leverage their strengths and overcome their weaknesses. HMT has the advantage of being flexible and accurate, but it also has the disadvantage of being complex and costly.
- Neural machine translation (NMT) is a form of machine translation that uses AI in the form of deep neural networks to encode and decode languages in a vector space. NMT has the advantage of being fluent and natural, but it also has the disadvantage of being opaque and data-hungry.

With increasing computing power, NMT has become the dominant machine translation method. It is the focus of this report, because it performs better than other methods and is still rapidly improving. NMT has some advantages over other types of machine translation, such as:

- It can capture complex linguistic phenomena, such as dependencies, word order, morphology, and syntax.
- It can generate fluent and natural translations, avoiding the word-by-word or phrase-by-phrase translations of other methods.
- It can learn from large amounts of parallel and monolingual data, improving its generalization and robustness.
- It can work with languages where the rule structure is nontraditional, as the model captures the statistical relationships between words and phrases in the source and target languages.

NMT models are trained using "parallel corpora", which are collections of texts available online in two or more languages, aligned at the sentence level. Often used collections include Wikipedia, which provides translations of the same articles in many languages; document repositories of international bodies such as the EU, which publish reports in several languages; and publicly available translations of books such as the Christian Bible. The need for sufficient training data explains why, while NMT models perform well when translating between major languages, they struggle with rare and low-resource languages.[3] However, innovations are emerging to overcome these challenges, for example by Meta, which has developed a single AI model, NLLB-200[4], that can translate across 200 different languages, including low-resource languages.

Machine translation systems have existed for a long time – Google Translate, for example, was launched in 2006. Their capabilities have developed rapidly, especially in the last few years. As the below graph shows, in 2017, no AI model could outperform an average human in reading comprehension tests, and language understanding capabilities had not been developed at all. Yet by 2020, AI models were beating humans at both and close to outperforming humans in translation.[5] Since then, the launch of publicly accessible generative AI models[6] such as OpenAI's ChatGPT, Microsoft's Bing AI and Google's Bard have given the development of these tools a further boost, as interest in them has grown enormously. Another factor in the rapid development of AI's language capabilities is the growth of computing power available to train AI models. For a long time, training computation for AI models followed Moore's law, doubling every 20 months. Since 2010, it has sped up and now doubles every six months. In addition, cloud computing has replaced super computers, making AI much more accessible.



Background on the Arctic Practices System

The Arctic Practices System (APS) is an online tool whose structure and framework were defined as part of the CAPARDUS project. The APS will be a platform for the development and dissemination of best practices in the Arctic region. The system will be designed to support the implementation of the Comprehensive Framework for Arctic Standards.

---

[3] Rare languages are languages with few speakers; low-resource languages are languages for which the corpus of text available online, based on which translation tools can be trained, is small. Most rare languages are also low-resource, but many low-resource languages cannot be considered rare – e.g., Lingala is a low-resource language spoken by 45 million people in Central Africa.

[4] No Language Left Behind: Driving inclusion through the power of AI translation: https://ai.facebook.com/research/no-language-left-behind/

[5] "Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals": https://www.nature.com/articles/s41467-020-18073-9

[6] Generative AI models can be defined as a type of software that is capable of generating new, original content based on patterns and examples they have learned from training data.

APS could be built on top of and leveraging existing features of the Ocean Best Practice System (OBPS[7]). The OBPS was created to serve the need for a collection of practices that was cross-discipline, end-to-end from observations to applications and sustained collections based on open (FAIR) principles. The OBPS is a global system comprising technological solutions and community approaches to enhance management of methods as well as support the development of ocean best practices. It is a permanent global repository of ocean research, operations, data/information management and applications methodologies (also known as "Best Practices").

Languages Considered in this Project

Major Languages

The following major languages have been included in the assessment:

- English as the main source language
- German
- French
- Danish
- Dutch

These languages were selected based on the availability of native speakers within the project team.

Indigenous/Rare Languages

The project team also evaluated automated translation possibilities for a number of indigenous and/or rare languages that are of relevance to the APS, including Pacific and Arctic languages (see overview in Table below). As the list of translation tools shows (see Annex), very few of these are currently supported, and only by very few tools.

[7] https://www.oceanbestpractices.org/

Table 1: List of indigenous languages considered in this project

Language Name	Language Family (Top Level)	Language Family (Sub-family)	Scripts (character-set used)	Number of speakers	Spoken in which Countries	Spoken in which Region	Number of known Dialects	ISO 639-3 Abbrev / link	Glottolog link/abbrev
Fijian	**Austronesian**	**Viwa-Lomaiviti-East Viti Levu**	**Latin alphabet**	**650,000**	**Fiji**	**Papunesia**	**4**	**fij**	**fiji 1243**
Pijin (Solomons Pidgin)	**English Creole**	**Melanesian Pidgin**	**Latin alphabet**	**307,000**	**Solomon Islands**	**Melanesia**		**pis**	**piji1239**
Saami	**Uralic**		**Latin alphabet with some special characters**	**30,000-40,000**	**Finland, Sweden, Russia**	**Northern Scandinavia**	**11**	**N/A**	**saam1281**
Greenlandic Inuit	**Eskimo-Aleut**		**Latin alphabet with some special characters**	**57,000**	**Greenland**	**North America**	**2**	**N/A**	**gree1280**
Kalaallisut	Eskimo-Aleut	**Greenlandic Inuit**	Latin alphabet with some special characters		Greenland	North America		kal	kala1399
Tunumiisiut	Eskimo-Aleut	**Greenlandic Inuit**	Latin alphabet with some special characters		Greenland	North America		N/A	tunu 1234
Canadian Inuit (Inuktitut)	**Eskimo-Aleut**		**Inuktitut alphabet**	**40,000**	**Canada**	**North America**	**2**	**N/A**	
Eastern Canadian Inuktitut	Eskimo-Aleut	**Inuktitut**	Inuktitut alphabet		Canada	North America		ike	east2534
Western Canadian Inuktitut	Eskimo-Aleut	**Inuktitut**	Inuktitut alphabet		Canada	North America		ikt	west2618
Alaskan Inuit (Iñupiaq)	**Eskimo-Aleut**		**Latin alphabet with some special characters**	**3,000**	**United States (Alaska)**	**North America**	**2**	**ipk**	**inup1234**

Methodology

The current study undertook two separate but linked assessments: one to create a general overview of tools available and of their capabilities, to show the range of actors engaged in translation and what they offer; and one translation quality assessment, focusing on a few major tools and languages, to assess the overall technical capacity of translation tools to produce fluent, accurate translations of scientific/academic texts.

Tools Assessment

In order to understand the ecosystem of machine translation tools, and particularly tools that are built on neural machine translation models, an initial desk study has been conducted using search engines and mentions on translation news websites (e.g., slator.com) and online communities (e.g., machinetranslate.org). The results were used as the basis for further analysis, the outcome of which is presented in an overview of tools.

The overview categorizes translation systems according to their main functionalities, the translation types they offer, their general cost and the basic options for technical interoperability they provide. Additionally, each system was checked for the support of the languages relevant to this project, based on self-reported information by the service providers on the systems' websites or product sheets.

Translation Quality Assessment

Text samples in English were extracted from resources available in the Ocean Best Practices System, consolidated in one document and translated to the selected major languages (Danish, Dutch, French, and German) using six of the most readily available and widely used translation tools that support full document translation: Alibaba Translate, DeepL, Google Translate, Reverso Translation, Systran Translate PRO, and Yandex Translate (see next section). The samples were selected to test the ability of the translation tools to process a range of types of texts relevant to the APS and its users:

- Scientific language describing methodologies for experiments, including numbers and symbols; to test the tools' ability to correctly translate names of chemical substances and descriptions of scientific processes, and to render symbols, including sub/superscript;
- Academic language, including acronyms and references to international organizations, structures and processes; to test the tools' ability to translate long, complex sentences and to use the official translations of acronyms and names of international organizations, treaties etc.;
- Language referring to indigenous rights and issues; to test the tools' reliability in using respectful language surrounding indigenous issues;
- Practical instructions in bullet point format; to test the tools' ability to translate short sentences without a clear context, as well as their reliability in correctly conveying crucial information;
- Tables and diagrams; to test the tools' ability to render these correctly and to translate terms with no immediate context.

Native speakers were asked to review each translation, compare it to the original, flag any issues and fill in a brief survey. The survey included specific questions (e.g., 'Did tool X correctly render the symbols in sentence Y?') as well as an overall evaluation of accuracy and readability of each translation. The goal was not to provide quantitative or statistically significant evidence of differences in translation quality, but rather to identify recurring issues with translations, highlight specific elements or sentence structures that tools struggle with, assess consistency of quality across different languages, and evaluate the overall readiness of the technology for use with the APS.

Outcome 1: Assessment of Translation Tools

This section outlines the outcomes of the research and assessment of the translation tools. The full list of tools assessed can be found in Annex 2.

Categorization of Translation Tools

The assessment gathered information about over 35 online tools and services related to automatic translation. These tools can be broadly split into the following categories (many companies offer more than one of these within their product range):

- User facing translation: these provide an input field directly in a web browser (most well-known example: Google Translate)
- Translation via application programming interfaces (API, most often offered by Public Cloud service providers like Microsoft, AWS or Alibaba)[8]: these services are located in the "backend", hosted on a virtual server within the broader cloud infrastructure of the service provider. Translation requests are sent via programmable calls through the APIs, including the source text, the code for the target language, and additional parameters as needed. The service then translates the text and returns the translated output through the API to the customer's system.
- Computer-Assisted Translation tool (mostly used for localization of web pages): these tools are mostly used to support translation operations, both for professional translators as well as for the automatic translation of websites (localization of content).
- Open-source translation software and models: these tools are released with an open-source license, which generally allows users to download, install and run these tools and models on their own infrastructure.

Features of Translation Tools

Depending on the aim of the tool and the sophistication of the technologies it uses, companies offer several layers of translation features. The most important are the following:

- Plain text translation: simple translation of words, sentences, or paragraphs. Users can copy/paste or type into a form, and the text is translated; or it is sent via an API call and the translated text is returned to the caller.
- Document translation: the translation tool takes in a whole document (usually Word or PDF) and translates all the text within it while aiming to keep the formatting intact (headings, paragraphs, emphasis, …), and then makes the translation available for a user to download. An API call would submit the file and return the translated document.
- Image translation: accepting an image (.jpg) as input, the tool tries to extract the text included in the image (via Optical Character Recognition OCR), translates it and aims to put the translated text back in its place, while the rest of the image stays untouched as far as possible.
- Website translation: when a user enters a website address, the translation tool receives the content and format of the original page, translates the visual text, and presents the user with a page that looks similar to the original website, but with translated text.
- Media translation: some platforms also support the translation of videos and audio (N.B.: these features have not been tested).

[8] An API is a set of protocols, routines, and tools for building software and applications. It specifies how software components should interact with each other and provides a way for different applications to communicate with each other. APIs allow developers to access the functionality of another piece of software, such as a web service, and use that functionality in their own applications.

Many service providers offer their translation features to users (i.e., on a website or an app), through an API, or they offer a combination of user-facing and API access. While user-facing services are more geared towards a one-off interaction for a translation, APIs are geared towards the automation of translations. These can be especially useful if a company or organization wants to publish documents, blog posts or otherwise in more than one language, as they allow them to automate the copy/paste and/or upload steps in the translation process.

Language support

During the assessment, we identified the languages supported by each tool. From the over 40 tools, we found that most of the tools support translation between the major languages. Apart from the very targeted tool Nutserut, which translates between Greenlandic and Danish, all platforms supported translations to and from English, Spanish, French and German; most also include Danish, Dutch, Swedish, Norwegian, and Finnish. It is also worth mentioning that not all tools offer support for all their claimed languages across all different features, e.g., Alibaba does not support Danish for document translation, but for plain text only.

Only a handful of tools supported one or more of the indigenous languages in the scope of this project, with Fijian being the most frequently supported. The Greenlandic government has developed the above-mentioned Greenlandic-Danish translation tool, but its performance is not comparable to others as it relies on rule-based translation. The only other language we found and were able to test in a service is Inuktitut.

The reasons for the lack of support for these languages are manifold:

- Limited training data: Machine translation models require large amounts of high-quality training data to achieve accurate translations. Indigenous and rare languages often have limited available data for training these models, making it challenging to develop robust translation tools.
- Limited economic incentive: Developing machine translation tools requires significant investment in research, development, and maintenance. The economic incentive to develop these tools for indigenous and rare languages may be limited due to their small number of speakers and fewer potential users compared to major languages.
- Lack of clear use cases: Endangered languages, including many indigenous languages, face the threat of extinction due to the dominance of major languages. Consequently, there may be a lack of clear use cases or demand for machine translation tools for these languages. The focus may instead be on language preservation and revitalization efforts to ensure the survival of these endangered languages.

These challenges highlight the need for increased investigation into use cases for translation into indigenous languages, to make the case for investment in their support by translation tools. These use cases could include, but should not be limited to, the translation of ocean-related information to those languages.

Custom training of models

Some companies – predominantly the big Cloud service providers – also offer the development, training, and hosting of custom models. This practice can support organizations and companies in training the systems to understand domain-specific jargon or proprietary names (product names, codenames for projects, etc.). Custom-trained models can generate translations that are more effective for reaching a specific target audience, and as such be better understandable within the context domain. A disadvantage of this method is the need for huge corpora of domain-specific texts in language pairs, which may not exist or may be harder to come by than publicly available data that is less context relevant.

Outcome 2: Report on translator output usefulness/quality

Tests for Translations into Major Languages

The translation tests revealed clear differences in quality between the different translation tools, with Alibaba Translate and Yandex Translate judged as very poor to poor across all languages for both accuracy and style, Reverso Translation judged as poor to average, Systran Translate PRO as average to very good, and Google and DeepL as good to very good (see the tables below). It is interesting to note that while Systran performed as well as DeepL and Google for German and Dutch, it performed much worse for French, which may indicate that it works better for some language families than others. However, further testing would be required to confirm this.

Table 2: Reviewers' ratings of the accuracy of the translations provided by each tool by language (DA: Danish, DE: German, FR: French, NL: Dutch)

Translation tool	VERY POOR	POOR	AVERAGE	GOOD	VERY GOOD
Alibaba Translate		DE | FR			
DeepL					DE | NL | FR | DA
Google Translate				NL | FR | DA	DE
Reverso Translation		DE	NL | FR | DA		
Systran Translate PRO			FR	DE	NL
Yandex Translate	NL	DE | FR | DA			

Table 3: Reviewers' ratings of the readability/style of the translations provided by each tool by language (DA: Danish, DE: German, FR: French, NL: Dutch)

Translation tool	VERY POOR	POOR	AVERAGE	GOOD	VERY GOOD
Alibaba Translate	DE | FR				
DeepL				NL	DE | FR | DA
Google Translate				DE | FR | DA	NL
Reverso Translation		DE | NL	FR | DA		
Systran Translate PRO			FR		DE | NL
Yandex Translate	DE | NL	FR | DA			

Full evaluation outcomes are available in the annex.

DeepL, Google Translate and Systran Translate PRO produced translations that were readable and close to fully accurate, and that would be unlikely to cause major problems if used, with caution, by native speakers. They also are remarkably good at recognizing context, for example correctly translating 'platform' as 'web platform' (and not, like Yandex, as 'train station platform'), and 'free travel zone' as 'borderless zone' rather than 'no-cost travel zone'. When a text was translated from English to a target language and then back to English using Google Translate and DeepL, not only did very little noise creep in, but the result often improved on the original in terms of word choice, syntax and style, particularly the use of commas.

At the same time, all tools, including top-performing Google and DeepL, also made random errors of varying severity that, while unlikely to lead to catastrophe, could cause confusion and/or offense. For this reason, the authors conclude that machine translation could be an immensely valuable tool to knowledge repositories such as the APS, but should only be used without human intervention in limited settings, with clear disclaimers. If budget is available for some form of human intervention (either a quick vetting to verify key details or terms, or a full post-machine translation edit), machine translation tools are able to largely replace human translators at least for academic and technical texts – especially since, as should be kept in mind, human translations also normally require editing and verification.

Key issues that the tools struggled with include:

- **Consistency:** often, the same terms are translated in various different ways throughout a document. This is most frequently the case with terms that have translations but are commonly left in English, such as 'best practices.' In addition, when translation tools are forced to assign gendered articles to unknown terms or acronyms, such as 'the APS', they are often inconsistent in the use of those articles. Some tools are more consistent than others; for example, Google Translate performs very well in this aspect, whereas DeepL has difficulties. These inconsistencies can be irritating but are mostly harmless; one exception was DeepL's single translation of 'Inuit' as 'Eskimo' in Dutch.
- **Symbols and sub/superscript:** special symbols and sub- or superscript as used in chemical formulas are often rendered incorrectly. Google in particular often leaves entire sentences in sub/superscript.
- **Diagrams with labels:** These form a particular challenge as they contain very little context for the terms contained in them. None of the tools was able to provide a useful translation of a diagram included in the document. Perhaps in recognition of this flaw, Google Translate disregards any text within text boxes or diagrams, leaving it in English. Yandex produces corrupted Word files if a diagram is included.

Often, the tools were wrong in predictable ways; there are specific sentence structures in English that machine translation tools struggle with and that in the future, authors could avoid if they knew their texts were to be machine-translated. For example, the sentence "Repository content can be annotated through defined metadata fields, enhancing structured archiving and retrieval" contains an implied 'thereby', which the machine translation tools often missed, resulting in a confusing sentence. To support accurate machine translation, this thereby could be added: "Repository content can be annotated through defined metadata fields, *thereby* enhancing structured archiving and retrieval". In fact, writing with translation in mind may generally help authors produce texts that are easier to understand for laypersons and non-native speakers, and that contain less ambiguity.

Test of Translations for Indigenous Arctic Languages

Due to the lack of support of most languages of the arctic region, it was only possible to run limited tests for Inuktitut and Greenlandic/Kalaallisut.

Inuktitut

The research on translation tools revealed that only Microsoft offered translation to and from Inuktitut, via its Azure cloud service as well as the user-facing Bing Translator. For our tests with Bing Translator, we used various sentences of the Report "People of the Ice Bridge", which was published in three languages (Inuktitut, Kalaallisut, English) and thus provided a baseline for comparison of Inuktitut to English translations.

The translations, as can be seen in the examples below, might not be usable by themselves, but can paint a coarse picture of what is being said and what the contents of the original text allude to. As Microsoft's Bing translation tool performs well in major languages (it was excluded from the quality assessment as it does not support document translation, but smaller tests were conducted), this demonstrates that there are large differences in translation quality even between different languages offered by the same tool; it follows that even if more top-performing providers start supporting indigenous/rare languages, this does not guarantee the translations will be of good quality.

Table 4: Exemplary test translations from Inuktitut to English, including the original English from the same source document

Original from "People of the Ice Bridge"[9]	**Machine Translation from Inuktitut (MS Bing Translator)**	**Human-created English translation from the report**
ᐱᑭᐊᓚᓲᖅᓱᐊᖅ ᑲᒪᓇᖅᑑᑦᑐᓂ ᑲᔾᔮᓇᖅᑐᖅ ᓄᓇᖓ. ᐅᑭᐅᖅᑕᖅᑐᒥ ᓇᔪᕈᒥᓇᖅᑑᕗᖅ ᓴᐳᔾᔨᓲᖑᓂᖓᓄᑦ ᓯᓚ, ᐃᒪᖅ, ᐊᒻᒪᓗ ᓯᑯ ᑲᑎᓲᖑᑦᑐᑎᑦ, ᐳᐃᔨᖃᐅᕈᓘᔭᖅᑐᓂ ᓂᕿᖃᕐᕕᐅᓯᒪᓪᓗᓂ ᐃᓄᓐᓄᑦ ᐃᓕᖅᑯᓯᕆᔭᐅᓯᒪᓪᓗᓂ ᑕᒪᓐᓇ ᑕᐅᓴᓄᑦ ᐊᕐᕋᒍᓄ. ᐅᑭᐅᖅᑕᖅᑐᒦᑐᖅ ᐃᓄᓐᓄᑦ ᑕᐅᕙᓂᕐᒥᐅᑕᓅᖓᔪᖅ ᐊᑐᓲᖑᑦᑐᑎᓪᓗ ᐱᑭᐊᓚᓲᖅᓱᐊᕐᒥ ᑲᒥᓯᓇᐅᓪᓗᑕ ᐱᒋᐊᕈᑎᒋᓚᐅᖅᓯᒪᔭᕗᑦ ᑕᒪᓐᓇ ᐃᖅᑲᓇᐃᔮᖅ.	Pikilassuaq is a surprisingly beautiful landscape. The Arctic is a place where it protects the weather, water, and ice, which has been a source of seafood that has been a source of country food for thousands of years. It's a northern inuit person who lives in that area and they use the Bkikishua Commissioners and we started this position.	The Pikialasorsuaq is an extraordinary place. It is an Arctic haven where sky, ocean, and ice meet, teaming with marine life that has nourished Inuit subsistence culture for thousands of years. It is for the people who live and use the Pikialasorsuaq that we as Commissioners initiated this work.

[9] PEOPLE OF THE ICE BRIDGE: THE FUTURE OF THE PIKIALASORSUAQ / ᐃᓄᐃᑦ ᓄᖕᓂᕐᒥᐅᑦ: ᐱᑭᐊᓚᓲᖅᓱᐊᑉ ᓯᕗᓂᒃᓴᖓ / SIKUKKUT IKAARTUT KINGUAAVI: SIUNISSAMILU PIKIALASORSUAQ, Report of the Pikialasorsuaq Commission November 2017

Original from "People of the Ice Bridge"[9]	**Machine Translation from Inuktitut (MS Bing Translator)**	**Human-created English translation from the report**
ᐃᓚᖓ ᖃᐅᔨᓴᐅᑐᐊᖅᓯᒪᔪᖅ ᑲᒥᓴᓐᑯᓐᓄᑦ ᑐᓴᕋᐊᕐᕕᖃᖃᑦᑕᑎᓪᓗᒋᑦ ᐋᖅᑮᔪᖅᑲᓇᐊᖅᑲᖅᒪᑦ ᐃᓄᓐᓄᑦ ᓯᕗᒃᑲᖅᑖᓇᔭᐅᔪᒥᒃ ᐊᐅᓚᑦᑎᔨᐅᓗᔭᖅᑐᑦ ᐱᑯᐊᓚᓱᖅᓱᐊᖅᒧᑦ ᐊᒻᒪᓗ ᖃᓂᒋᔭᖕᒋᓐᓃᑦᑐᓄᑦ. ᐱᔾᔪᑎᒋᓚᔭᖅᑕᖓ ᐊᐅᓚᑦᑎᔩᑦ ᐸᐸᑦᑎᓇᓱᐊᖅᓗᑎᑦ ᐆᒪᔪᖅᓂᑦ ᐊᑐᖅᑕᐅᓲᑦ ᑕᐃᑲᓂ ᖃᓂᒋᔭᖕᒐᓃᑦᑐᓪᓗ ᐱᑯᐊᓚᓱᖅᓱᐊᖅᒥ, ᐊᒻᒪᓗ ᐋᓐᓂᐊᖅᑕᐃᓕᒪᔾᔪᑎᒋᓲᖕᒋᑦ ᓄᓇᓖᑦ ᑕᒪᒃᑯᓂᖓ ᓄᓇᒦᙶᖅᑐᓂᑦ ᐃᒪᒦᙶᖅᑐᓂᓪᓗ. ᐅᖃᐅᔾᔨᕆᐊᖅᑏᑦ ᑲᑎᒪᔨᕋᓛᖕᒋᑦ ᐃᑲᔪᖅᑐᐃᓗᔭᖅᑐᑦ ᐊᐅᓚᑦᑎᔨᒃᑯᖕᒋᓐᓂᑦ.	One of the key findings of the Commission's consultations is the need to establish a public leadership role in the management of the Pacific and the surrounding area. The purpose of the management is to protect wildlife use in and around the vicinity of the Pikilasuk, as well as the health of communities from the natural and water resources. The advisory committee would support the co-management.	One of the key findings of the Commission in the course of their consultations was the need to establish an Inuit- led management regime for the Pikialasorsuaq and surrounding regions. The purpose of the management regime is to promote the conservation of living resources within and adjacent to the Pikialasorsuaq, and the related health of communities that depend on these resources. Advisory committees will support the management regime.
ᑕᐃᔅᓯᒪᓂᓕ, ᑭᓱᑐᐃᓐᓇᐃᑦ ᐊᓯᔾᔨᖅᓂᐊᖅᑐᖅᔫᔾᔭᓚᐅᖅᓯᒪᙱᑦᑐᑦ. ᑭᓱᐊᓂᓕ ᐅᓪᓗᒥ ᓯᓚ ᐊᓯᔾᔨᖅᑐᖅ ᐊᒻᒪᓗ ᖃᐅᔨᒪᔪᖓ ᑲᓛᓛᑦ ᓄᓈᓐᓂᒥᐅᑦ ᐃᓅᖅᑲᑎᕗᑦ ᐃᓄᑋHᐅᐃᑦ ᑕᒪᐅᖓ ᑎᑭᑉᐸᒍᓐᓃᖅᑐᑦ. (ᓛᐅᓇ ᐊᐅᓪᓚᓗᒃ, ᐊᐅᔪᐃᑦᑐᖅ).	At that time, things seemed to change. But today the weather is changing and I know that Greenland is not the people who come here anymore. (Larry Audlaluk, Grise Fiord).	Back then, it seems like nothing would change. But today there is climate change and I am aware of that since the Greenlandic people who are our fellow-Inuit Inughuit do not travel here anymore. [Larry Audlaluk, Aujuittuq (Grise Fiord)].

Greenlandic

Similar to the methodology followed for Inuktitut, we used the "Local documentation and management of living resources: User Guide"[10] to compare automatic translations from Greenlandic/Kalaallisut into English with human-written English.

The only machine translation tool that offered this language pair was Alibaba translate, which performed poorly (analogous to its performance with major languages). See the table below for two examples.

[10] Local documentation and management of living resources: User Guide / Pisuussutinik uumassusilinnik aqutsineq nalunaarsuinerlu: Atuisunut ilitsersuusiaq, Piniakkanik Sumiiffinni Nalunaarsuineq Greenland Self-Rule, January 2014

Table 5: Exemplary test translations from Greenlandic to English, including the human-created English version of the same paragraphs

Original Greenlandic	English Machine Translation	Human-created English translation from the report
Pinngortitalerinermi apeqqutaasinnaasut, assersuutigalugu: 1. Uumasut amerliartussappat imaluunniit ikiliartussappat? 2. Aalisakkat pisarineqartartut alliartussappat imaluunniit milliartussappat? 3. Uumasut takkuttarnerat pisarnermiit kingusinnerulissava imaluunniit siusinnerussappat?	For example: 1 Will the animals be multiplied or decreased? 2 Should the fish grow, or should they grow up? 3 Animals arrive later Be ahead?	Biological questions, e.g.: 1. Are there more or fewer animals? 2. Are the fish caught bigger or smaller? 3. Are the animals arriving/leaving later or earlier than 'normal'?
Ullumikkut pisuussutinik nalunaarsuisarneq. Kalaallit Nunaata pisuussutaanik uumassusilinnik nalunaarsuineq ullumikkut annerusutigut pisanik nalunaarsuinermi katersivimmut Piniarnermut nalunaaruteqarnikkut, ilisimatusarnikkullu misissuisarnernit pisarpoq.	The present - day treasure - making process. Recording the earth's resources is a major event in the world By giving a report to the branch office of Jehovah's Witnesses, and from scientific research.	**Existing resource documentation and monitoring.** Monitoring of Greenland's living resources is currently taking place primarily through reports made to the catch database Piniarneq and research-based studies. Experience indicates that there is great potential for managing knowledge on living resources among Greenlandic hunters and fishers.

Additionally, we tested the translation tool "Nutserut.gl", which is provided by the "Language Secretariat of Greenland". As mentioned previously, this rule-based machine translation tool supports translations between Greenlandic and Danish. We translated the texts from above into Danish and asked a bilingual professional translator to assess the quality. The response was that these results were unusable, and there is still a need for human translators.

General challenges & opportunities concerning the use of AI-based translation tools

Challenges and Risks of Using AI Tools for Translations

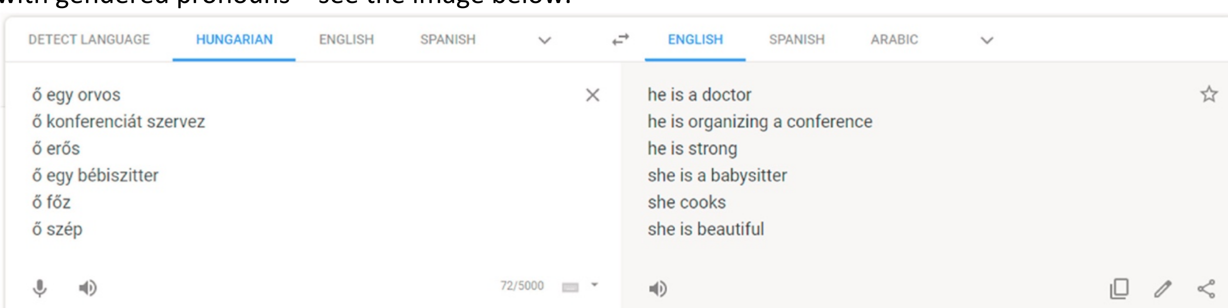
Challenges remaining for the use of AI-based translation tools include:

Unpredictability, which leads to unreliability. AI models can behave in unpredictable ways. As translations/responses are generated afresh in response to each new query, they could generate a correct translation 99 times and then still produce a catastrophic mistake the 100th time. Especially for texts concerning sensitive subjects, AI translation tools should therefore not be used without human oversight – and even if they improve significantly, the chance of a random mistake will always exist. Examples of issues caused by overreliance on AI translation tools include the story of Afghan refugees whose asylum claims in the US were rejected because of mistakes in the automatic translation of their written testimonies.[11] In the translation tests undertaken for this report, the sentence "Detailed charts are not available for all parts of the Arctic" was translated correctly most of the time, but on at least one occasion, the 'not' was left out.

The most extreme form of AI unpredictability concerning machine translation is AI hallucination, which is when tools invent new text while claiming it is a translation. The authors of this study found, for example, that several generative AI tools, which are not designed for translation but are capable of it, will claim to be able to translate from Greenlandic to English, but provide "translations" that bear no relation to the Greenlandic input (the reference of the Alibaba translation of the Greenlandic text above to Jehovah's Witnesses may also be an example of this). Without any knowledge of the source language, such hallucinations can be difficult to recognize. Users must remember that generative AI tools are trained to mimic human language convincingly rather than to always provide factually accurate information, and as a result they often invent facts.

AI models are not programmed in the classical sense, but rather provided with a dataset to learn from and trained until they provide the desired results. It is often impossible to ascertain how exactly they arrive at those results. This is referred to as the AI "black box"; and this opaqueness of AI models and how they come to their conclusions is what makes their behavior unpredictable. It also makes any undesirable behavior difficult to correct and even more difficult to eliminate completely; they cannot be 'debugged' like traditional code.

Bias, stemming from training datasets. AI machine translation tools amplify biases contained in their training datasets. For example, since these datasets contain fewer references to female doctors than to male doctors, the tools will often default to assigning male pronouns or articles to any doctors in the text. This bias is especially strong when translating from languages without gendered pronouns to languages with gendered pronouns – see the image below.



Source: https://www.sciencedirect.com/science/article/pii/S2590291121001352 - when this experiment was recreated in May 2023, Google Translate provided the same translations, but added a disclaimer saying "Some sentences may contain gender-specific alternatives. Click a sentence to see alternatives."

However, even when translating between languages with gendered pronouns, machine translation tools consistently revert to gender stereotypes. For example, in translations of the sentence "The doctor asked the nurse to help her with a test" to languages with gendered articles (e.g., Spanish), DeepL and Google assume the doctor is male ("el médico"), despite the reference to 'her'. In the sentence "The nurse asked the doctor to help him with a test", DeepL and Google both make the nurse female ("la enfermera").

[11] https://restofworld.org/2023/ai-translation-errors-afghan-refugees-asylum/

The datasets used to train translation models contain more sentences referring to men than sentences referring to women. As a result, translations of sentences involving men generally are more accurate than those of sentences about women. Racial biases are more difficult to recognize in translated text but are widespread and well-documented in AI-based image generators, some of which produce white men 97% of the time when asked to generate an image of a "CEO".[12]

Copyright, data privacy and upcoming regulation. There are two major copyright questions related to AI, which different legislative bodies are racing to answer: 1) Should AI tools be allowed to use copyrighted content as training data and to generate responses? And 2) Who, if anyone, can claim copyright over text generated by an AI tool? These questions have mostly been asked in relation to image generation tools, which at times very obviously borrow from specific works, and can even be asked to generate images in the style of a particular artist. However, the answers that legislators come up with could have far-reaching consequences for companies developing all kinds of AI tools and for their users, though it is not clear what exactly those consequences would be. For example, the EU is preparing legislation that would oblige AI companies to disclose the use of copyrighted materials in the training of their systems.[13]

There are also data privacy concerns, as the training data used for AI models contains much personal information, making it easy for users to use AI tools to pool together information on (primarily well-known) persons from a wide range of sources. And when asked to provide a bio of a specific person, AI tools such as ChatGPT will often fill in any gaps with incorrect facts or even complete hallucinations.

Environmental impact. As with most AI-based tools, neural machine translation tools use significantly more energy than their more traditional counterparts, leading to higher emissions, depending on the source of that energy. Stanford's 2023 AI Index Report estimates that the GPT-3 emitted 502 tons of CO2e in 2022, equivalent to 251 round-trip flights from New York to San Francisco.

The challenges outlined above can be addressed. Unpredictability and bias had limited impact on the sample translations and can be largely eliminated either by disclaimers or with minimal human intervention, such as a quick vetting by a native speaker. Upcoming copyright laws and other legislation may limit the power of generative AI tools but are unlikely to result in limited access to translation tools. To reduce their tools' environmental impacts, AI companies have the option to use renewable energy, and any emissions must be weighed against the positive impacts AI can have, including on energy efficiency - e.g., an experiment using AI to optimize cooling in Google's data centers led to 40% energy savings.[14]

Automating Translations in APS/OBPS

Due to the availability of APIs for translation services, knowledge management systems like the APS and OBPS can integrate the option to translate the hosted body of knowledge into their platforms. Depending on the budget and technical setup, there are different ways how such an integration can be accomplished:

- Translation on demand by users: should a user find a document, based on keywords or a basic understanding of the existing document, that they would like to read in a different language, they could trigger a translation request on the platform for either the abstract or the full publication.

[12] https://www.technologyreview.com/2023/03/22/1070167/these-news-tool-let-you-see-for-yourself-how-biased-ai-image-models-are/
[13] https://www.zdnet.com/article/generative-ai-might-soon-face-some-major-copyright-limitations-from-eu/
[14] https://www.deepmind.com/blog/deepmind-ai-reduces-google-data-centre-cooling-bill-by-40

To avoid duplication later on, these translations can be stored in the metadata and/or file storage for later easy retrieval, tagged by language code or similar.

- Translation after submission to APS/OBPS: upon submission and approval of a new publication for the knowledge base, the system could automatically trigger a translation to/from the most common languages and store these in the system. This approach will make it faster to retrieve translations, but will also likely result in higher costs (e.g., storage needs, translation services). Additionally, it is unclear how widely used the translated documents will be. On the other hand, the knowledge base will be consistently translated into all selected languages, allowing for a clearer expectation management approach towards the users of the platform.
- Translation of the most requested/used resources in the APS/OBPS based on user analytics.

There are several resources needed for this effort with costs attached to them. In addition to the programming and integration efforts on the APS/OBPS systems' side, translation providers also charge a fee. Below is an overview of some cloud provider costs (an average abstract has around 2000-3000 characters; this report amounts to around 73700 characters, including spaces):

Microsoft Azure Cloud Translation:

- Free standard translation of up to 2 million characters per month, including free tier custom training
- Beyond 2 million characters, costs are ca. 9 EUR per additional 1 million characters.
- Document translations are never free and cost around 13 EUR per 1 million characters.
- Additional custom translation adds costs for training, translation and model hosting
- Other costs could be incurred with set up of auxiliary support infrastructure, etc.

Google Cloud Translation Service

- Free translation of up to 500,000 characters per month
- Beyond the 500,000 character limit, cost are USD $20 per 1 million characters
- Document translation is USD $0.08 per page (in docx, ppt or pdf format)
- Custom training is calculated by the hour

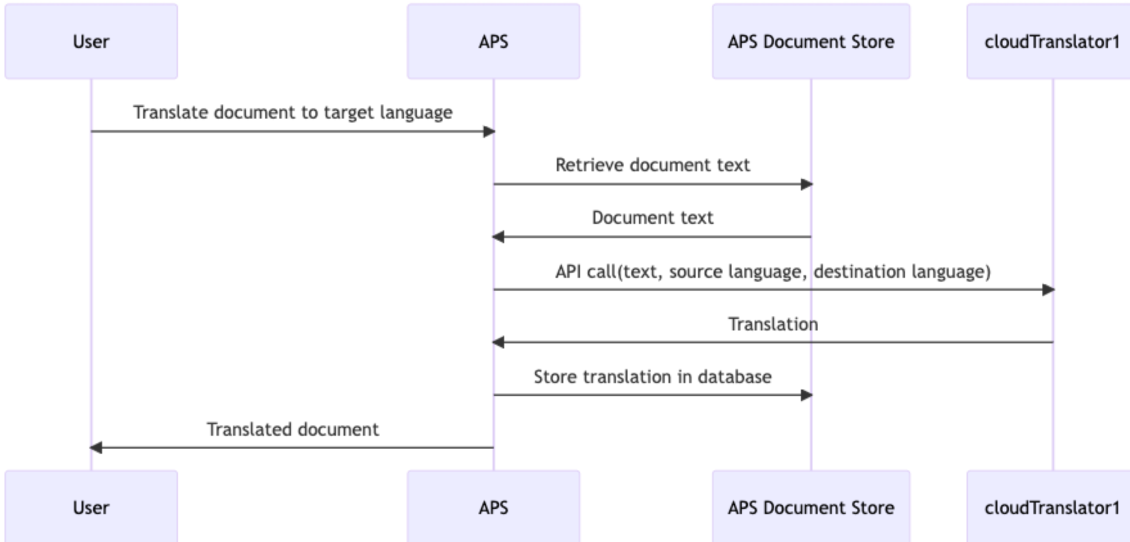
Amazon AWS Translator

- Free translations of 2 million characters per month for the first 12 months
- Beyond that, costs are USD $15 per 1 million characters

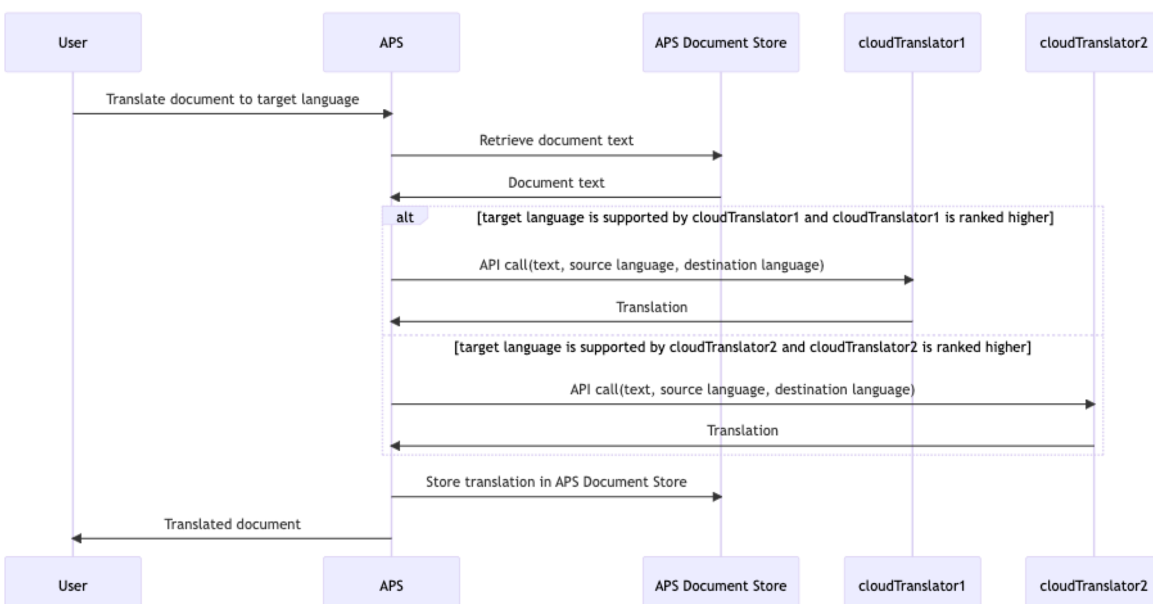
Some cloud service providers offer discounts and sponsored usage credits to non-profit organizations, charities, and research institutions. For example, Microsoft offers 3,500 USD in credits per year to eligible organizations, Amazon up to 5,000 USD and Google also offers discounts and credits after registration (amount unknown).

An integration of translation tools to the APS/OBPS could be accomplished using one or more translation services, following a fairly simple sequence of actions to be built into the platform.

Using only one translation service, this can be achieved through one API call to the translation service. This process, outlined in the Figure below, involves the User initiating the translation to an indicated language within the APS (either by uploading or an active request for translation of a specific document), after which APS retrieves the document text from the Document Store. Using this text, the translation service is called to translate the text into the required language, which it does immediately. The returned text is then stored in APS Document Store for later direct retrieval and is finally also delivered to the User.



Alternatively, considering the differing language offer of different translation tools and their varying quality, it would also be possible to include connections to multiple translation services. While this would add some effort on the platform side, it could yield better results overall, and could end up being cheaper in the long run due to the free tier character allowance of the different services. The process, also shown in the diagram below, is like the one outlined above, with the addition of a decision tree on which translation tool to use first (for example, based on a ranking per language, overall quality, or amount of free translation resources for the time frame).



Implications for the CAPARDUS project

Following the presentation of this work at the final CAPARDUS conference in Greenland on May 21, 2023, additional feedback and ideas were collected. Participants in the event expressed their excitement about the opportunities that automatic translation could bring to the Arctic region. While translation of resources into Danish or English from other major languages could already be of significant use to the APS, the recent innovations to improve translation tools' capacity to translate from and into low-resource languages may generate even more exciting prospects.

The Arctic is home to many indigenous languages; however, conversations between speakers of these different languages tend to be held using a third language (e.g., Danish or English, or both). Direct translations between Arctic languages, using models such as the NLLB one, could therefore allow for a different level of communication and collaboration between communities of all regions in the Arctic, and especially so within the topics of the CAPARDUS project. Potentially, translation tools could:

- Facilitate citizen science projects over the whole region without language barriers;
- Enhance communication between communities in different regions, increasing the potential for exchange of knowledge, but also potentially increase direct trade, and improve intercultural understanding as shared expressions and idioms in similar Arctic languages are likely lost through translation to and from English or Danish;[15]
- Fully integrate indigenous knowledge on Arctic practices into the APS, and allow for its transfer between peoples and languages (Canada to Greenland, etc.);
- Translate indigenous knowledge into major languages for recording and distribution among non-native speakers, to document and preserve it, to allow for more widespread learning, and to facilitate its inclusion in research endeavors, strategic planning, and policy development, also possibly beyond the Arctic.

Summary, Conclusions and Outlook

Conclusions

This report sought to assess whether currently available machine translation tools 1) were able to provide accurate and fluent translations of ocean-related scientific and academic texts to major languages; and 2) offered translations to indigenous and/or rare languages of relevance to the ocean/Arctic domain.

The quality assessment found that, though translation tools are not perfect, they have made great leaps in quality over the past few years and the top performers could be highly useful in the context of the APS/OBPS. Mistakes still occur, however, it is important to remember that human translations also contain errors, as do original texts. Not recommending the use of machine translation tools because of potential harm caused by minor mistakes in translations would be counterproductive and patronizing to users – a disclaimer should suffice to warn people that translations may not perfectly represent the original text. In any contexts in which a flawless translation is crucial, machine translations can be vetted

[15] This only works if the translation tool does not work with an intermediary language. E.g. Dutch and German share the expression 'toekomstmuziek'/'Zukunftsmusik' (nice, but currently unrealistic or overly ambitious future plans), but rather than translate 'Zukunftsmusik' as 'toekomstmuziek' from German to Dutch, it translates it as 'taart in de lucht', which is a literal, unidiomatic translation of the English 'pie in the sky'.

or post-edited by human translators. A simple vetting for major errors would not take much time, and would produce high-quality translations much more quickly and affordably than when using human translators for the entire process.

The assessment showed that machine translation tools struggle with specific sentence constructions and issues such as ambiguity, which could be avoided in original texts. If authors adjusted their writing style to accommodate machine translation, by using shorter sentences with clearer structures and less ambiguity, that would also benefit non-native speakers and lay readers.

The tools tested do not provide any support for the indigenous and/or rare languages of interest to this project. As the Inuktitut translation test using Bing Translator demonstrated, even major tools struggle with low-resource languages, and even if such languages are offered, that does not mean that translations will be of high quality. However, innovative strategies and technologies for training AI based on limited data are emerging that may make high-quality translations from and into low-resource languages possible in the future.

Outlook

Machine translation tools are still developing fast. What improvements are expected in the coming years?

An exciting project in relation to the APS is Meta's No Language Left Behind (NLLB), which has developed a new technique for training language models based on limited datasets. It is using this to offer translations between 200 languages, including Fijian. No Arctic languages are included currently, but the development of the model is ongoing, and it has been made open source so that others can build on it. Current uses of the NLLB include a cooperation with Wikipedia to help editors develop translations of articles into their languages; perhaps a similar cooperation would be feasible with the APS.

Another exciting realm of innovation is that of real-time speech-to-text or speech-to-speech translation, for videos, conference talks, or even fieldwork interviews.

AI Translation and the APS – what use cases could be envisaged? What are the limitations?

First of all, further research is required to determine what information the indigenous peoples of the Arctic would like to access in their own languages. It is also important to assess what current barriers to access exist, besides the language barrier. For example, it may be more important to develop a repository that works on mobile devices, or with low bandwidths, than to translate resources into local languages. Once a clear use case has been established, the APS project could seek to work with technology companies such as Meta to encourage them to offer support for indigenous language translation.

As mentioned previously, the limitations and challenges outlined above can be fairly easily overcome with some human intervention – this will require budget but will be much more affordable than paying for full text translation by human translators. Perhaps users could be granted or asked to pay for a monthly number of credits that can be used to request vetting of a certain number of pages of translated text by a pool of native speakers. These vetted translations can then be made publicly available on the APS for other users, too.

The future of AI translation contains many known unknowns, such as the direction of the further development of NLLB, and the shape of upcoming legislation, but possibly even more unknown unknowns, so any application will have to be flexible and resilient to change.

Potential areas for further research & future projects

Other areas the project could consider looking into further include:

- Cooperation with (EU) research programs, non-profits, and universities to develop translation capacity for indigenous languages. This development could also be encouraged through competitions or hackathons to develop translation models that support low-resource languages.
- Cooperation under the scope of the UN Indigenous Languages Decade (2022-2032). Besides allowing for the translation of information into indigenous languages, advances in machine translation may also be able to help to save endangered languages from extinction, or at least record them for posterity. It could do the same for bodies of text that exist in those languages.
- Other uses of AI for document repositories such as the APS: AI-Powered Knowledge Base tools allow users to collect insights from across different documents based on a single query, generate summaries, and organize documents in a way that suits them. This would make interaction with the platform more organic and reduce the need to download/scroll through large numbers of PDF files.

Annex 1: Full translation evaluations

1: Does the translation tool recognize references and leave them in English?

Translation tool	YES, ALWAYS	NO, NEVER	SOMETIMES
Alibaba Translate		DE	FR
DeepL			DE | NL | FR | DA
Google Translate		NL | FR	DE | DA
Reverso Translation		DE | NL | DA	FR
Systran Translate PRO		NL	DE | FR
Yandex Translate		NL | FR | DA	

2: Does the translation tool accurately convey symbols and subscript/superscript? (e.g., " ° ", "$_4^+$")

Translation tool	YES, ALWAYS	NO, NEVER	SOMETIMES
Alibaba Translate		FR	DE
DeepL	DE | NL | FR | DA		
Google Translate			DE | NL | FR | DA
Reverso Translation		DE | NL | FR	DA
Systran Translate PRO	NL | FR		DE
Yandex Translate	DE | NL | FR | DA		

3: Does the translation tool use the official translation for 'World Meteorological Organization', leave the name in English, or invent a new translation?

Translation tool	USES OFFICIAL TRANSLATION	DOES NOT TRANSLATE	INVENTS NEW TRANSLATION
Alibaba Translate	DE | FR		
DeepL	DE | NL | FR		DA
Google Translate	FR	DE | NL | DA	
Reverso Translation	DE | FR | DA	NL	
Systran Translate PRO	DE | NL | FR		
Yandex Translate	NL | FR	DE | DA	

4: Are 'Ocean Decade' and 'best practices' translated in the same way in the table caption and in the table header?

Translation tool	BOTH THE SAME	BOTH DIFFERENT	ONE THE SAME, THE OTHER DIFFERENT
Alibaba Translate			DE | FR
DeepL	DE | FR		NL | DA
Google Translate	NL | FR | DA	DE	
Reverso Translation		DA	DE | NL | FR
Systran Translate PRO	FR		DE | NL
Yandex Translate	FR | DA		DE | NL

5: Is 'Inuit' translated in the same way throughout excerpt 3?

Translation tool	YES	NO	YES, BUT THE TRANSLATION IS OFFENSIVE	NO, AND ONE OR MORE OF THE TRANSLATIONS IS OFFENSIVE
Alibaba Translate	DE | FR			
DeepL	DE | FR | DA			NL
Google Translate	DE | NL | FR | DA			
Reverso Translation	DE | NL | FR | DA			
Systran Translate PRO	DE | NL | FR			
Yandex Translate	DE | NL | FR | DA			

6: Does the translation tool use formal or informal forms of address consistently?

Translation tool	YES, ALWAYS FORMAL	YES, ALWAYS INFORMAL	NO, IT USES A MIX OF THE TWO
Alibaba Translate	DE | FR		
DeepL	DE | FR	DA	NL
Google Translate	DE | FR	NL | DA	
Reverso Translation	FR	DA	DE | NL
Systran Translate PRO	DE | FR	NL	
Yandex Translate	DE | FR	NL | DA	

7: How would you judge the overall ACCURACY of the translations?

Translation tool	VERY POOR	POOR	AVERAGE	GOOD	VERY GOOD
Alibaba Translate		DE | FR			
DeepL					DE | NL | FR | DA
Google Translate				NL | FR | DA	DE
Reverso Translation		DE	NL | FR | DA		
Systran Translate PRO			FR	DE	NL
Yandex Translate	NL	DE | FR | DA			

8: How would you judge the overall READABILITY/STYLE of the translations?

Translation tool	VERY POOR	POOR	AVERAGE	GOOD	VERY GOOD
Alibaba Translate	DE | FR				
DeepL				NL	DE | FR | DA
Google Translate				DE | FR | DA	NL
Reverso Translation		DE | NL	FR | DA		
Systran Translate PRO			FR		DE | NL
Yandex Translate	DE | NL	FR | DA			

Annex 2: List of Translation Tools

Alibaba Translate

Developer: Alibaba
Website: https://translate.alibaba.com/

Type (mostly): User Facing, API
Allows customization of existing model: ？
Supports self-trained model: ？

Translations on following media:

Text	Images	Documents	Websites
✅	✅	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	✅	✅	❌	✅	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Amazon Translate

Developer: Amazon
Website: https://aws.amazon.com/translate

Type (mostly): API
Allows customization of existing model: ✅
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Apertium

Developer: open source
Website: https://www.apertium.org

Type (mostly): User Facing, open source SW
Allows customization of existing model: ❌
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	❌	❌

AppTek

Developer: Apptek
Website: https://www.apptek.com/

Type (mostly): API, corporate
Allows customization of existing model: ✅
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	?	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Baidu

Developer: Baidu
Website: https://fanyi.baidu.com/

Type (mostly): User Facing, API
Allows customization of existing model: ?
Supports self-trained model: ?

Translations on following media:

Text	Images	Documents	Websites
✅	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
✅	✅	✅	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Bing Translator

Developer: Microsoft
Website: https://www.bing.com/translator

Type (mostly): User Facing
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	✅	❌	✅	✅

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

DeepL

Developer: DeepL SE
Website: https://www.deepl.com/translator

Type (mostly): User Facing, API
Allows customization of existing model: ✅

Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Google Cloud Translate

Developer: Google
Website: https://cloud.google.com/translate/

Type (mostly): API
Allows customization of existing model: ✅
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Google Translate

Developer: Google
Website: https://translate.google.com/

Type (mostly): User Facing
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	✅	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

IBM Watson Language Translator

Developer: IBM
Website: https://www.ibm.com/cloud/watson-language-translator

Type (mostly): API mostly
Allows customization of existing model: ✅
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

iTranslate

Developer: iTranslate
Website: https://itranslate.com/

Type (mostly): User Facing, API
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Lingvanex

Developer: Lingvanex
Website: https://lingvanex.com/translationserver/

Type (mostly): user facing, API
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Mirai Translate

Developer: Mirai
Website: https://miraitranslate.com/en/

Type (mostly): API
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	❌	❌	❌	❌	❌	✅

ModernMT

Developer: translated
Website: https://modernmt.com/

Type (mostly): API
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	✅	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

MS Translator (Azure)

Developer: Microsoft
Website: https://translator.microsoft.com/, https://www.microsoft.com/en-us/translator/business/translator-api/

Type (mostly): API
Allows customization of existing model: ✅
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	✅	❌	✅	✅

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German

✅	✅	✅	✅	✅	✅	✅	✅	✅

Niutrans

Developer: Niutrans
Website: https://niutrans.com/

Type (mostly): API
Allows customization of existing model: ?
Supports self-trained model: ?

Translations on following media:

Text	Images	Documents	Websites
?	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
?	?	?	?	✅	✅

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

NLLB (no language left behind)

Developer: Meta
Website: https://ai.facebook.com/research/no-language-left-behind/

Type (mostly): model
Allows customization of existing model: ❌
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	✅	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

OpenL

Developer: OpenL
Website: https://openl.io/

Type (mostly): User Facing
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	❌	✅	❌	❌	❌	✅

Papago

Developer: Naver
Website: https://papago.naver.com/

Type (mostly): User Facing
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	❌	❌	❌	❌	❌	✅

Phrase NextMT

Developer: Phrase

Website: https://www.memsource.com/

Type (mostly): Computer-Assisted Translation tool (Localization)
Allows customization of existing model: ✅
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	❌	✅	❌	❌	❌	✅

Reverso Translation

Developer: Reverso
Website: https://www.reverso.net

Type (mostly): User Facing, Corporate Version available
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	❌	✅	❌	✅

Systran Translate PRO

Developer: Systran
Website: https://www.systran.net/en/translate/

Type (mostly): User Facing, API

Allows customization of existing model: ✅
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Taia MT

Developer: Taia
Website: https://taia.io/machine-translation/

Type (mostly): API
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	?	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

TextUnited

Developer: TextUnited
Website: https://www.textunited.com/technology/artificial-intelligence-machine-translation/

Type (mostly): Computer-Assisted Translation tool (Localization)
Allows customization of existing model: ✅
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	✅	❌	❌	✅	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Unbabel

Developer: Unbabel
Website: https://unbabel.com/

Type (mostly): Computer-Assisted Translation tool (Localization)
Allows customization of existing model: ✅
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Yandex.Translate

Developer: Yandex
Website: https://translate.yandex.com/

Type (mostly): User Facing, API
Allows customization of existing model: ✅
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	✅	✅	✅

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Promt

Developer: Promt
Website: https://www.promt.com/company/technology/neural-machine-translation/

Type (mostly): API, corporate
Allows customization of existing model: ？
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Tencent Machine Translation

Developer: Tencent
Website: https://cloud.tencent.com/product/tmt

Type (mostly): API
Allows customization of existing model: ？
Supports self-trained model: ？

Translations on following media:

Text	Images	Documents	Websites
✅	?	?	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)

❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	❌	❌	❌	❌	❌	✅

XL8

Developer: XL8
Website: https://www.xl8.ai/

Type (mostly): API
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Youdao

Developer: Youdao
Website: https://ai.youdao.com/#/

Type (mostly): API
Allows customization of existing model: ？
Supports self-trained model: ？

Translations on following media:

Text	Images	Documents	Websites
？	？	？	？

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	✅	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German

✅	✅	✅	✅	✅	✅	✅	✅	✅

Rozetta T-400

Developer: Rozetta
Website: https://www.rozetta.jp/product/?language=en

Type (mostly): API
Allows customization of existing model: ✅
Supports self-trained model: ？

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	?

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

Lilt

Developer: LILT
Website: https://lilt.com/

Type (mostly): Human Translation Interface, API
Allows customization of existing model: ？
Supports self-trained model: ？

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	✅	✅	✅	✅

OpenNMT

Developer: OpenNMT
Website: https://opennmt.net/

Type (mostly): Open Source Translation Software
Allows customization of existing model: ❌
Supports self-trained model: ✅

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
❌	❌	❌	❌	❌	❌	❌	❌	❌

LibreTranslate

Developer: n/a (open source)
Website: https://libretranslate.com/

Type (mostly): Open Source Translation Software
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	✅	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	❌	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
✅	✅	✅	✅	✅	❌	✅	✅	✅

Nutserut

Developer: Oqaasileriffik

Website: https://nutserut.gl/

Type (mostly): User Facing
Allows customization of existing model: ❌
Supports self-trained model: ❌

Translations on following media:

Text	Images	Documents	Websites
✅	❌	❌	❌

Supported languages:

Saami	Greenlandic	Canadian Inuit	Alaskan Inuit	Fijian	Pijin (Solomon)
❌	✅	❌	❌	❌	❌

English	Spanish	French	Danish	Dutch	Norwegian	Swedish	Finnish	German
❌	❌	❌	✅	❌	❌	❌	❌	❌