Machine Translation Use in University Contexts Policy

July 2023

Introduction

The most important in a nutshell

This policy provides Bern University of Applied Sciences staff and students with guidelines for the use of machine translation for understanding and writing texts in learning, teaching, research and administrative work.

In the past few years, neural machine translation (NMT) based on artificial intelligence has made astounding progress. This is reflected in the widespread use of NMT in society, including in university contexts. NMT can now provide a quality that is sufficiently high to serve a variety of purposes and to promote multilingual communication. Nevertheless, this technology has its limitations and comes with risks that users have to be aware of, including concerns over data protection standards, confidentiality and privacy.

In addition, while media reports often present NMT as a "ready-to-go" solution, current state-of-the-art NMT systems still present several issues related to accuracy, adequacy or appropriateness. However, the grammatical fluency of MT output and the unpredictability of MT mistakes makes it difficult for users to identify and correct errors, a phenomenon known as false fluency. Infrequently used words, long or complex sentences, specialised terminology, and intentionally ambiguous wording to convey caution or probability (known as hedging) are particularly challenging for MT systems. Bias in MT output, especially cultural, racial and gender bias, remains a problem with MT systems, as does the quality of MT output for lesser-used languages. Finally, most NMT systems that are available to the general public mainly work at the sentence-level, which leads to major issues with respect to the cohesion and logic of a text, such as unclear pronoun references or inconsistent terminology.

While NMT is expected to improve over time, many aspects concerning meaning (accurately capturing the message), context (adapting references to suit your intended audience) and text (appropriate style or structure) are still not handled well with current MT systems. This policy therefore seeks to provide Bern University of Applied Sciences staff and students with guidelines for the use of machine translation in learning, teaching, research and administrative work.

Definitions

The most important in a nutshell

Machine translation (MT) is the translation of a text from one language to another by a computer software without human involvement. The MT output may be used in its raw form (raw MT output) or it may be revised (post-edited MT output).

Gist translation
The use of machine translation to gain a general yet perhaps imprecise understanding of the original meaning of the source text.

High-risk communication
Texts that have a high visibility (e.g. published externally), are meant to elicit a certain behaviour or response in the reader or whose misunderstanding could result in reputational, financial or legal risk.

Low-risk communication
Texts that have a low-visibility (e.g. limited to internal use) or are consulted only occasionally and for information purposes only.

Machine translation
The translation of a text from one language into another by a computer software without human involvement.

Neural machine translation
The latest-generation of data-driven or corpus-based machine translation based on artificial neural networks. Neural machine translation (NMT) systems are capable of producing high quality output and are trained in a manner so they can "learn" to translate new texts and new words that they have not encountered during training.

Post-editing
The process of editing MT output to correct errors of language, meaning or style. Post-editing can be divided into different categories depending on the person conducting the post-editing, as well as the text, its purpose and its target reader:
– Post-editing performed by casual users
– Post-editing for low-risk communication (see the definition below) to correct any errors that can be detected
– Post-editing performed by professionally trained translators
– Light post-editing for high-risk communication (see the definition below) to ensure that meaning has been accurately transferred while ignoring stylistic issues
– Full post-editing for high-risk communication (see the definition below) to attain a high level of quality comparable to that of human translation

Pre-editing
The process of preparing a document for machine translation, for example by correcting mistakes in the source text, simplifying sentence structures or removing ambiguities.

Raw MT output
Machine translation output that has not been edited or revised by the user.

Source language
The language from which a text is to be translated.

Source text
The text that is to be translated.

Target language
The language into which a text is to be translated.

Target text
The translated version of the source text.

Translation
The process of transferring the message and meaning of a text from the source language to the target language.

Data protection

The most important in a nutshell

When using a free MT system (e.g. Google Translate, DeepL), do not enter any personal data or confidential information (including original research). When using a fee-based MT system (e.g. DeepL Pro), you must ascertain the privacy policy of the company. Some companies might provide protection to all data in order to allow the usage of personal data. Irrespective of this, you must comply with the General Data Protection Regulations, particularly with regard to personal data, as well as any obligations of confidentiality or secrecy. Never enter sensitive personal data.

a. Free MT systems (e.g. Google Translate, DeepL)

There is no such thing as a free MT system. Users pay with their texts. This means that any data supplied to the MT system for translation may be re-used as training data for the algorithms or sold to other companies. Therefore, it is not lawful to enter any personal data or confidential information (including original research) when using a free MT system.

b. Fee-based MT systems (e.g. DeepL Pro)

Some MT systems offer a fee-based subscription in which they guarantee that a user's text will not be re-used. Companies might even provide protection to all data in order to allow the usage of personal data or internal information. However, users must comply with the General Data Protection Regulations when using these systems, particularly with regard to personal data, as well as any obligations of confidentiality or secrecy. You must therefore ascertain the privacy policy of the company in question when using MT systems. In any case, you may not enter the following into any MT system:
– confidential information as defined by the BFH information classification structure
– information that is bound to professional secrecy (in particular patient and/or health data)
– information that is bound to contractual obligations of secrecy particularly sensitive personal data (data on health, privacy or ethnicity; data that refers to religious, political or trade union affiliations or activities; data on administrative or criminal prosecutions or sanctions; and data on social assistance measures or involuntary commitment).

Transparency

The most important in a nutshell

As a general rule when using MT output, you should always insert a disclaimer indicating if the text is raw MT output or post-edited MT output.

As a general rule, MT output should always be indicated to the readers of a text. For example, when emails are written using an MT engine, a disclaimer should be inserted indicating if the email consists of raw MT output or post-edited MT output. This will help the readers of the message know how to situate the quality of the translation and potential mistakes. Below are some examples of disclaimers that can be adapted by you depending on the level of post-editing.

For raw MT output or light post-editing by a casual user:
'*This text was automatically translated into [insert target language]. It may contain errors. You will find the original [insert source language] text below.*'

For minimal post-editing by a professional translator (known as 'light post-editing'):
'*This text was translated using machine translation with minimal post-editing to ensure readability and accuracy. The [insert target language] version is offered for information purposes only*.'

No disclaimer is required for full post-editing undertaken by a professional translator.

How to use MT

The most important in a nutshell

You must always take into consideration privacy and/or confidentiality issues when using MT. In addition, the following principles apply.

a.   MT use in learning and teaching
Running a text that you have not written yourself through an MT system and using the MT output without citing the original source is plagiarism. Students must follow any guidelines set by their instructors with respect to MT use for assignments and assessments.

b.   MT use in research and administrative work
Before running a text through an MT system, consider whether a potential error generated by the software could result in offence, reputational loss, legal or financial damage, or grave misunderstandings with serious consequences. A professionally trained translator must always be involved for high-risk communication. Read the full policy for more information about which types of texts are suitable for MT.

Copyright and other ethical issues must be taken into consideration when using MT. In this respect, the following principles apply.

a.   MT use in learning and teaching

Running a text that you have not written yourself through an MT system and using the MT output without citing the original source is plagiarism.

– Students and instructors must indicate if they are using raw MT output or post-edited MT output.
– Students must follow any guidelines set by their instructors with respect to MT use for assignments and assessments, including exams, term papers, theses and the like.

b.   MT use in research and administrative work

The use of MT depends on how the translated text will be used, the purpose of the text and the target reader. When texts are used for temporary and/or casual internal communication, some inaccuracies in the translation may not have severe consequences. However, when texts are used for permanent and/or formal external communication, deficiencies can cause reputational damage.

Temporary (unpublished) and/or low-risk internal communication can consist of raw and/or post-edited MT output provided that you
– insert a disclaimer indicating that the text was translated using MT,
– copy the original text below so that the reader can compare the translated text to the MT output in case of ambiguity, uncertainty, etc., and
– follow the steps in the MT checklist with respect to pre-editing and post-editing.

Examples of temporary and/or low-risk internal communication include emails, internal presentations, non-published texts and internal reports.

Permanent (published) and/or high-risk external communication should be post-edited or translated by a professional translator. A disclaimer should be inserted indicating if it has been lightly post-edited. No disclaimer is needed if it has undergone full post-editing.

Examples of permanent and/or high-risk external communication include official transcripts, assessment texts, academic texts, as well as highly specialised texts such as research publications and creative texts that require cultural adaptation such as website content, annual reports and marketing materials.

Help and advice

The most important in a nutshell

If you have any questions, please contact the Bilingualism + Multilingualism Office at bilinguisme@bfh.ch.

MT output can vary considerably depending on the text type, language pair and MT engine. Consult with the translation service at your university or with a professional translator if you are uncertain as to whether a text is suitable for MT or if the text should be professionally post-edited or translated.

If you have any questions about this policy, please contact the Bilingualism + Multilingualism Office at bilinguisme@bfh.ch.