

Machine Translation



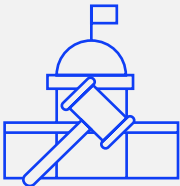



Recommendations for Public Administration

Machine translation offers considerable potential for public administration. The Commercial Register of the Canton of Schwyz and the Integration Unit of the Canton of Zurich implemented two case studies within the scope of the “Innovation Sandbox for Artificial Intelligence (AI)”. As the analysis of these case studies has shown, human translators remain indispensable for translation of official documents. However, training translation models specifically for public administration and integrating public administration terminology into existing solutions can offer clear added value. Specialised Swiss providers of machine translation can increase the quality of translations and improve data security. Use of specific models for public administration that can be adapted to the needs of individual public offices or units is a promising approach in the long term.

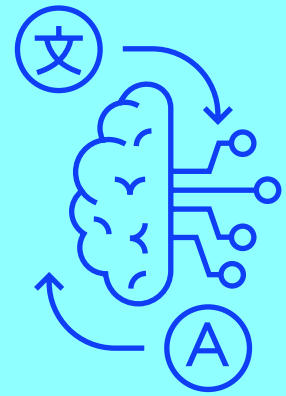
Innovation Sandbox for Artificial Intelligence (AI)

This document was created within the scope of the Innovation Sandbox for Artificial Intelligence (AI). The sandbox is a test environment for the implementation of AI projects from various sectors. This broad-based initiative involving public administration, industry and research, is designed to promote responsible innovation by allowing the project team and participating organisations to collaborate closely on regulatory questions and enabling the use of novel data sources.

[More information](#)

	I. Relevance of machine translation	04
	II. Differentiation of fields of use	06
	III. Legal framework of public administration	08
	IV. Commercial Register and Integration Unit as case studies	10
	V. Conclusions and recommendations	19
	VI. Annex with benchmarking results	21

I. Relevance of machine translation



Machine translation* offers considerable potential for public administration. For a country like Switzerland where four official languages are spoken and several cantons are multi-lingual, this technology is particularly relevant. Furthermore, the need for translations in all Swiss cantons spans numerous areas, including migration offices, commercial register offices, labour market offices, courts and law enforcement.

Progress and future of machine translation

Significant progress has been achieved in the domain of machine translation in the past decade. Modern systems use methods such as **neural networks** which have markedly improved the quality of translations. These are known as AI-powered **translation models**. These technologies enable translation tools to better understand context and, thus, deliver more accurate and more natural-sounding translations. Looking to the future, the expectation is that innovations in AI technologies will enable machine translations to become even more precise and versatile, especially in regard to processing “rarer” languages and dialects as well as specialised jargon.

The objective of machine translation in public administration is to simplify language accessibility for public administration employees who work with different languages and, furthermore, to improve the quality of communication with the general public. At the same time, machine translation can reduce recurring costs of professional translations in selected cases. However, AI-based translations

will never fully replace the role of humans. If what is required are legally reliable or officially certified translations, public administration should limit the use of machine translation to support only.

However, there are numerous less critical areas of use for machine translation, including fact sheets, informal official correspondence and standardised administrative documents. Machine translation offers an efficient and low-cost alternative in these areas. The challenge is to find a balance between automation and human expertise, and to thus ensure the accuracy of translations in each individual case.

“Machine translation offers considerable potential for Swiss public administration!”

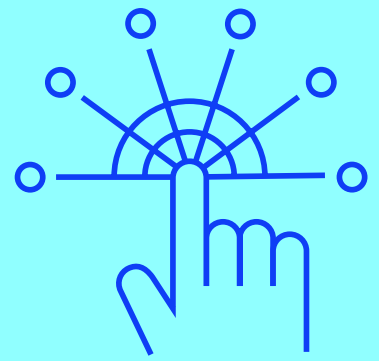
Raphael von Thiessen, Head of Innovation Sandbox for AI

Two public administration units gained practical experience with machine translation as a part of the “Innovation Sandbox for AI”. The presented case studies highlight the significance of data security – especially the need for a data storage solution within Switzerland, as opposed to freely available, international cloud applications. The report at hand also analyses the training of translation models specifically for public administration and the integration of public administration terminology into

* Words in bold in the text are explained in the glossary.

existing solutions. Approaches of this kind allow for machine translation to be used more effectively and more tailored to the requirements of public administration, which may improve the quality of the results. Chapter four of this report describes the specific case studies. But, first, it is worth differentiating the fields of use and considering the legal bases for use of machine translation in public administration.

II. Differentiation of fields of use



Use of machine translation in public administration requires thorough and differentiated consideration of fields of use. This is important so as to increase efficiency, improve translation accuracy and guarantee the security of sensitive data.

Where can machine translation be used?

Not every use case for machine translation is the same. On the one hand, the need for professionally reviewed and certified documents will remain for documents which must be legally reliable, e.g. court documents, as accuracy and reliability are paramount in these cases. Even then, machine translation may offer added value, provided the results are critically reviewed and revised by professionally trained experts. On the other hand, ninety-five per cent accuracy can be considered sufficient in many areas. These approximate accuracies are suitable for general information or for less critical areas of communication, i.e. areas that do not require a perfect translation.

The question as to whether machine translation should be limited to written documents, or whether oral translations should be included as well, also influences the scope of use. Whereas the focus in the examined case studies was on written translations, the option of oral (simultaneous) translations is another area worth exploring – offering considerable potential for automated translation of public administration transcriptions.

Choice of language is another significant factor. The decision as to whether translations should only extend to Switzerland's national lan-

guages, or include other languages as well, is a highly relevant one. **Translation models** vary considerably in terms of quality, reliability and consistency (see chapter four of this report). Many public administration offices are faced with languages that are not widely spoken in Switzerland, and for which the required training data for machine translation is insufficient. Machine translations are also relevant in crisis situations. The beginning of the war in Ukraine is a case in point, when the availability of translators was limited, and when fast, machine-generated translations can make a significant difference in the expeditious provision of information to the respective target groups – in this case, to refugees from Ukraine.

Considerations from the perspective of public administration

The decision as to whether public administration should provide content in various languages calls for consideration of strategic and ethical factors. Should public administration, at all, translate content and provide it in other languages? In principle, the assumption is that a public authority will only communicate in its own respective official language. Prior to the existence of automated translations, this principle made sense and was in keeping with the times. The only exceptions were official proceedings, e.g. within the scope of court proceedings, for which translations were required. Most public administration units assume that anyone living in Switzerland will learn one of the country's national languages. However, the reality is often different, especially for migrants who have only been in

Switzerland for a short time. In addition, the availability of translated content could raise expectations that will be difficult to meet, especially in terms of choosing which languages to translate documents into. According to which criteria would public administration choose these languages? Based on the number of language speakers? Or on account of political and social situations (e.g. refugee crises)? It can be argued, furthermore, that if no translations are provided, then foreign language speakers can resort to free online translation services. But do truly all individuals concerned have access and the skills required for responsible use of AI-based translation tools?

A further consideration relates to how public administration should manage the use of freely available translation tools, such as DeepL and **Google Translate**, or new large language models (LLMs) like **ChatGPT**. Despite bans or directives to the contrary, in practice, public administration offices often resort to freely accessible services. How can public administration ensure that sensitive data does not end up in the hands of providers who use it for their own purposes?

And how does public administration handle the risks arising from faulty translations? As well as normative aspects, this also extends to questions of liability, as public bodies are under the obligation to provide correct information that the general public can rely on (see chapter three).

Opportunities and challenges of available technology

The aim of the Sandbox projects was not to answer these normative questions. Some public administration units have already issued directives for use of machine translation tools. In the future, decision-makers in public administration will need to address these normative and ethical questions. Legislators and regulators may also need to take action on a political level. The focus of the project at hand was a different one, namely to assess the current status of technological solutions based on two specific case studies. Interesting approaches include: training specific **translation models** for public administration, feedback from professional experts, and storage of public administration

terminology to achieve consistent and improved translations in the context of public administration. A further important aspect concerns data security and protection of sensitive information. This relates, in particular, to personal data and internal public administration documents which are largely subject to strict directives.

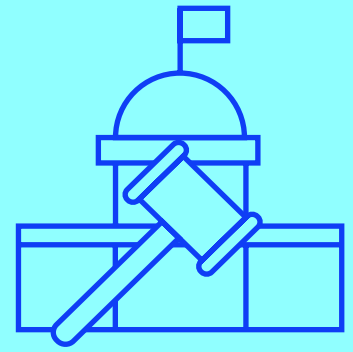
“Machine translation must be used responsibly and with a differentiated approach.”

Raphael von Thiessen, Head of Innovation Sandbox for AI

It is important to note that the results presented in this report are exclusively related to the specific case studies. This report does not provide comprehensive and generally applicable assessments of technological solutions. Rather, the results are to serve other administrative units in terms of taking their own initiative in the domain of machine translations and to adapt solutions to suit their own needs. Before describing the case studies, the following chapter provides an introduction into the relevant legal framework.

III.

Legal framework of public administration



Public administration must adhere to specific particularities when using machine translation services. As a rule, information to be translated will leave the internal (public administration) server for the purpose of translation. Depending on the content of the documents to be translated, a machine translation provided by a third party may, therefore, not be possible, or only possible if strict conditions are adhered to. Many public administration units have directives that govern which texts or information may be machine-translated, and under which conditions. Use of internal tools is conceivable, with information kept within the public administration entity, or for which there is assurance that data will neither be stored nor passed on to third parties.

Adherence to confidentiality and security specifications

Processing of a wide range of information within public administration is subject to special confidentiality and security requirements – be it because it concerns specially classified information (“confidential” or “secret”) or because it is personal data or other sensitive data.

Machine translation of documents that include confidential or secret information is only conceivable if the following conditions are met:

- The solution is installed on the public administration’s internal servers or the external solution concerned is certified (e.g. according to ISO 27001).
- Full transparency regarding data flows and cybersecurity is guaranteed.
- The public authority concerned audits and authorises the solution.

For personal data, public administration must also check the legal situation for each individual case, with restraint required in this area. Use of machine translation tools is unproblematic for texts that serve the purpose of personal communication exchanges within the public administration concerned.

“Machine translation can pose reputational and liability risks for public administration. A clear and transparent indication of the possibility of errors is thus an important requirement.”

Dr. Stephanie Volz, ITSL University of Zurich

As a rule, the only option available for documents that are subject to special confidentiality and security requirements is use of internal or certified external machine translation tools. In such cases, the emphasis is on integrity of the information included when processing. This applies, in particular, to information made available to the general public.

Consideration of reputation and liability risks

In principle, when information is made available to the general public, or is translated using internal

tools, this does not mean that they are accessible to a machine translation without any restrictions. Quite the contrary: despite the demonstrated good quality of machine translations, linguistic inaccuracies or translation errors are to be expected. These errors can cause reputational damage. In the case of official public administration information, erroneous statements present a liability risk if the document in question does not clearly indicate (e.g. through watermarking) that it is a machine translation and may contain errors. Consequently, for official public administration information that cannot be published with such a clear reference, the public administration office concerned should favour a professional translation.

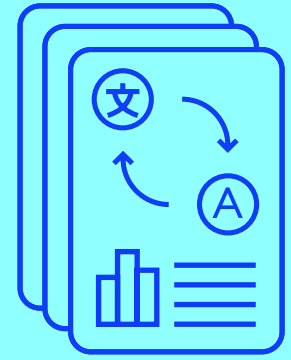
Ensuring transparency

It is important for public administration offices to inform about the use of a machine translation service. It must be clearly discernible to users that a text was machine-translated. It is also important to include a statement advising that the machine translation may contain errors or inaccuracies.

The next chapter describes two case studies from different areas of public administration. In addition to the legal framework, the focus was on investigating different translation options. The purpose is not to compare specific products or services, but to show public administration entities how they can go about finding AI-based translation solutions to suit their needs.

IV.

Commercial Register and Integration Unit as case studies



Two start-ups specialised in machine translation independently submitted AI projects to the “Innovation Sandbox for AI” between March and June 2022. Due to the thematic overlaps, the participating partners implemented both case studies simultaneously within the scope of the Sandbox – in order to examine various aspects of the same topic.

- **Neur.on** is a legal tech start-up that addresses the specific translation challenges and confidentiality needs in the domains of law, tax and banking with the help of AI
- **Textshuttle** is a spin-off of the University of Zurich and develops customised and AI-based translation systems with a focus on data security and controllability

Collaboration with both service providers proved to be highly productive. Both solutions have clear strengths, with the two case studies having very different objectives. The opportunities and challenges of the respective solutions cannot be generalised beyond the specific case studies presented here. Machine translation innovations and product developments by Neur.on and Textshuttle are advancing so fast that the results of the report at hand will soon be outdated, with new technical functionalities being introduced.

The aim is to show ways to advance machine translation constructively and responsibly within the context of public administration. The following section of this report is dedicated to describing the two case studies.

Specific translation model for the Commercial Register of the Canton of Schwyz

Cantonal commercial register offices are a central port of call for business activities in Switzerland. They record and manage important information about companies located in a canton and make that information publicly available. Excerpts from commercial registers play a decisive role in this regard. They serve as official documents that summarise data about a company’s name, legal structure, headquarters, purpose, authorised representatives as well as details on capital structures.

Public accessibility to these excerpts is of major significance. It ensures transparency and trust in business operations by allowing potential business partners to obtain information and check credibility.

Commercial register excerpts in foreign languages

The need for commercial register excerpts in various languages is a consequence of international business activities. Switzerland, with its multilingual residents and its role as a global economic player, has to do with a multitude of international investors, partners and customers. The availability of excerpts in several languages facilitates international trade and collaboration across borders by reducing language barriers and simplifying access to relevant company information for interested parties who do not speak German.

The translation of commercial register excerpts generally occurs in two ways: on the one

hand by way of generic translation tools, available online and free of charge, for fast and simple translations. These tools are designed for general purposes and offer basic translation quality which can be sufficient for simple texts. On the other hand, especially when it concerns officially certified excerpts, the services of accredited translation experts are required. These translation experts are not only in a position to ensure linguistic accuracy, but also understand specific legal terminology and concepts often included in such documents. Since, in many cases, there is no direct equivalent of legal terminology in other languages, the expertise of a specialised translator is a must-have to ensure that the meaning is conveyed correctly and to avoid any misunderstandings.

Regularity – an advantage for machine translation

Commercial register excerpts are particularly suitable for use of machine translation technologies. This is because of the standardised structure of excerpts, and the recurring use of specific terminology and phrases. Machine translation systems, particularly AI-based ones, can benefit from this regularity. They are able to discern and learn patterns, enabling efficient and consistent translations.

Despite these advantages, the challenge remains that commercial register excerpts often include complex legal terminology which requires a diligent and accurate translation. Herein lies the potential for combining machine translation with human expertise, thus ensuring efficiency as well as accuracy.

Specific model for commercial register excerpts

Neur.on has developed a specialised translation model tailored to the needs of the Commercial Register of the Canton of Schwyz. Some 20,000 commercial register entries served as a basis for training the specific model. The project initially started with the development of a model for translations from German into English. French and Italian, i.e. two of Switzerland's national languages, are planned as a next step. Optimisation of the model constituted a

decisive aspect – based on the expert input of the head of the Commercial Register and with inclusion of specialists from Neur.on. The aim is to provide a download function, with each document clearly marked as a “machine-generated translation”, so that users can directly download high-quality translations. However, that being said, a machine-translated commercial register excerpt cannot replace a translation performed by a specialist translator.

Implementation of this project involved several challenges:

- **Data export:** The first challenge was the export of data from an information management system which was not originally designed for such purposes. However, the IT service provider who developed the system managed to export the data successfully.
- **Data preparation:** Neur.on subsequently converted the data so that it could be processed by translators as well as AI specialists.
- **Challenges with terminology:** An analysis of the various types of companies (e.g. general partnership, company limited by shares, etc.) revealed complex challenges related to terminology, especially in regard to the search for English equivalents for legal terminology. The project team also took the latest developments in gender-inclusive language into account.
- **Legal research and discussion:** The next step encompassed legal research and preparation of a list of proposals to discuss with experts from the Commercial Register. The links between legal aspects and current language developments proved particularly interesting in this respect.
- **Implementation and consistency:** Inconsistencies were revealed during the processing of commercial register excerpts with respect to word combinations and source data expressions. To avoid errors and misunderstandings, the project team corrected these inconsistencies in the target text so as to ensure consistent use of terminology.

Findings from the commercial register excerpt case study

The project team achieved a good and consistent data basis by way of the described measures. This data basis is a prerequisite for training a robust, specialised translation solution. As an evaluation of content translated into English showed, the specific model is capable of handling considerable variability and delivers reliable results.

However, these results are not easily transferrable to other languages or to document types in other areas of public administration (e.g. courts, migration offices, etc.). That notwithstanding, there are some general findings, as listed below.

- **Good and consistent quality of translations:**

A narrow field of use allows a specific model to deliver highly reliable results. The project team achieved a high quality with machine translation of commercial register excerpts into English thanks, in particular, to the interaction between training the model and input text from experts. Con-

sistency of the translations was also very good in the area of use presented here.

- **Particularities of specific models:**

Neur.on will, however, need to adapt the specific model in the event of legislative changes. The specificity of such a model also harbours risks: reliability of the translation model may decline when the model processes unknown data. A case in point is the input field “Purpose” on the commercial register excerpt, where a company’s field of activity is described. The greatest risk of error is where content varies the most.

- **Transparency through labelling as**

“machine-generated translation”: Labelling as “machine-generated translation” is very important, given that machine translations are not perfect. Despite controlled processes, providers can never fully guarantee accuracy. This transparency is crucial in terms of making users aware of the limitations of machine-generated translations.



Figure I: This mock-up shows the content of the commercial register excerpts machine translated by Neur.on. Once the project has been completed, users will be able to download the translated excerpts directly from the Commercial Register’s website.

Data security with Swiss providers: Not all Swiss providers guarantee Swiss dataflows, especially when using cloud **GPUs**. Whereas international cloud service providers in Switzerland or the EU may also be suitable for non-confidential commercial register excerpts, when data is confidential, the users must be aware of the jurisdiction which Swiss servers may also be subject to. Neur. on is certified in accordance with the ISO-27001 standard. This means dataflows, data processes and data management are clearly defined, documented and auditable.

Benchmarking of translations for the Integration Unit of the Canton of Zurich

The Integration Unit of the Canton of Zurich is responsible for coordinating measures to encourage integration in the Canton of Zurich. These measures complement the integration support provided within existing standard structures at federal, cantonal and municipal level. Collaboration within these standard structures and advising thereon in the domain of encouraging integration constitutes a key task of the Integration Unit. Civil society organisations are, furthermore, important points of contact and cooperation partners. The Integration Unit also has an important role in providing information, and supports the municipalities in the canton with implementing their integration tasks. The unit also contributes to making sure integration concerns are considered in legislation and public administration.

The website “Welcome to the Canton of Zurich” provides newcomers with initial information for a successful start in their new location of residence. The website encompasses topics ranging from health to mobility and taxes in comprehensible language, so that individuals who are new to Switzerland can get a first idea of their new surroundings. Many newcomers will be unfamiliar with German. The purpose of the information is to reach precisely these non-German speakers. Thus, in the past, accredited translators were tasked with translating the website and related leaflets into the most broadly spoken foreign languages in the canton. These languages include French, Italian, English, Spanish and Portuguese. Furthermore, the Integration Unit also

regularly responds to enquiries received in foreign languages by e-mail.

Provision of information in additional languages

How can the Integration Unit make important initial information for newcomers available to as many people as possible? Considering the cost factor, translations by accredited translators are only possible within a limited scope. What are the alternatives? Can the individuals concerned translate the information themselves with the help of generic online tools? And, if so, do they have the skills required to translate key information on tax obligations and healthcare regulations? Machine translations could offer the option of making content available to as many people as possible while, at the same time, ensuring the quality of translated content. Furthermore, public administration employees would no longer have to resort to freely accessible online tools which do not comply with cantonal administration directives.

Integration of public administration terminology

Textshuttle has enhanced its business solution for the Integration Unit by adding humanly translated specialist terminology. The aim of this solution was to reproduce public administration terminology consistently and accurately in English, French, Italian and Polish translations. The primary goal was to find out if the specific translation solution is capable of achieving a better quality than generic online translation tools, especially with respect to Swiss peculiarities in public administration language.

The project team considered various aspects in regard to storing public administration terminology:

- Avoidance of complex and long expressions consisting of several words
- Correct use of singular and plural
- Correct spelling, including upper- and lower-case letters
- Particular form of writing words (e.g. gender-inclusive language)

The model uses stored public administration terminology flexibly so as to ensure good quality translations. The tool indicates where public administration terminology has been used in translations. This highlighting sensitises users to (official) public administration terminology.

Use for informal public administration correspondence

The Integration Unit used the Textshuttle tool for informal communications with newcomers to the Canton of Zurich, with translations based on the stored public administration terminology. The competent employees added an important note in German and in the respective foreign language to every piece of correspondence, as follows:

“The content of this email has been machine-translated from German. The translations may contain inaccuracies or errors. The users are aware of and bear the risk of any errors and/or inaccuracies in the translation. The Integration Unit does not assume any liability for this. The German version is binding.”

Integration Unit of the Canton of Zurich

The Integration Unit used Textshuttle’s machine translation tool for English, French and Spanish, as the majority of received enquiries were in these languages. In day-to-day work, the Textshuttle tool replaced foreign-language emails previously drafted by the responsible individuals themselves. In the absence of access to a specific public administration tool like Textshuttle, public administration employees will often resort to generic and freely available online services that store the content of translations externally and use it for further development.

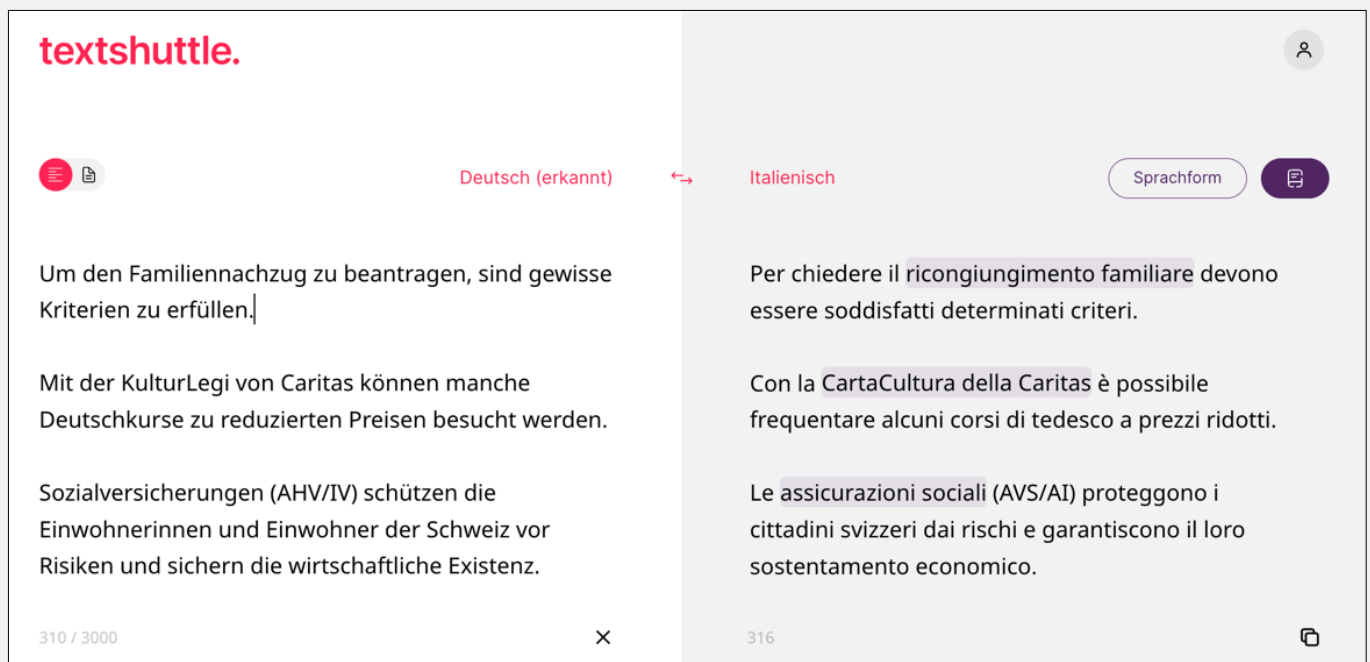


Figure II: Above an example of the Textshuttle business solution with stored public administration terminology of the Integration Unit, Canton of Zurich.

Switching to and from different applications proved to be a challenge resulting in additional work, which raised the question as to whether use is worthwhile from an efficiency perspective. Ideally, machine translation tools will be integrated into the respective public administration application. While Textshuttle offers this option, it was consciously excluded by the project team for the case study at hand.

The Textshuttle tool also delivered very good results for translation of complete documents, retaining the respective document formats correctly. Since the project team did not collect any direct feedback on quality of translations in relation to informal public administration correspondence, it was difficult to draw any conclusions in that regard. Therefore, a separate benchmarking was conducted with translation experts evaluating the tool in terms of performance and accuracy of translations.

Benchmarking based on initial information for newcomers

The mentioned text on the ["Welcome to the Canton of Zurich"](#) website provided the basis for benchmarking the translation solutions. The said text is characterised by its ease of understanding. Its content is suitable for evaluating machine translations of public administration language because of the broad spectrum of public administration terminology used from diverse subject areas (e.g. "premium reduction" or "residence permit").

Several translation variants were evaluated by professional translators in a blind test, including **DeepL Pro**, Textshuttle with stored public administration terminology as well as a human translation. The objective was to find out whether Textshuttle could achieve advantages through storage of public administration terminology. Depending on language, the Textshuttle model used different amounts of the stored public administration terminology, i.e. between 119 (Polish) and 139 (English) public administration terms.

Preliminary remarks regarding the evaluation

It is important to emphasise that this evaluation is not a scientific study, but a practical review. With just three evaluations conducted per language, the number of evaluations was very low (a total of twelve evaluations). It is very likely that the qualitative evaluation of the three translation variants would be different if carried out by other evaluating translators. There were major differences in the translators' evaluations who, e.g., rated communicative translations produced by humans negatively compared to literal translations. A possible explanation for this is likely to be the evaluation system. In practice, communicative translations often make good sense. The aim of the evaluation was, therefore, to not suggest a fictitious accuracy in the assessment of specific solutions. Rather, the benchmarking was to show how public administration units can compare and rate translation solutions for their own purposes. The results are not to be generalised beyond this specific study.

The evaluation of the translations drew on the following criteria*:

- **Accuracy:** does the translation convey the original message precisely?
- **Text flow:** does the translation flow and is it easy to read?
- **Terminology:** in particular, is public administration terminology, e.g. "residence permit", used correctly and consistently?
- **Grammar:** do the translations use grammar correctly?
- **Completeness:** has the full content, without adding or omitting information, been translated?
- **Cultural appropriateness:** does the translation take cultural nuances and sensitivities of the target group into account, e.g. newcomers with a migratory background?

The results of this benchmarking provide valuable insights into the performance and limitations of the various translation services. They also offer impor-

tant points of reference for further development of and adjustments to translation tools in the context of public administration, especially in regard to the needs of newcomers to Switzerland.

*Detailed results along the listed criteria are available in the Annex.

DeepL Pro

- **English:** DeepL Pro's translation into English was rated very positively, especially in terms of grammar.
- **French:** DeepL Pro's French translation was also rated as very good, particularly for grammar and completeness.
- **Italian:** DeepL Pro's Italian translation was, by contrast, rated rather negatively, particularly in regard to public administration terminology with major issues identified. Text flow and cultural appropriateness were also rated negatively.
- **Polish:** DeepL Pro also achieved very good results for Polish, which is a less widespread language in Switzerland. This underscores the international orientation and competency of the provider.
- **Varia:** DeepL Pro translations were very positively rated across all languages for grammar correctness.

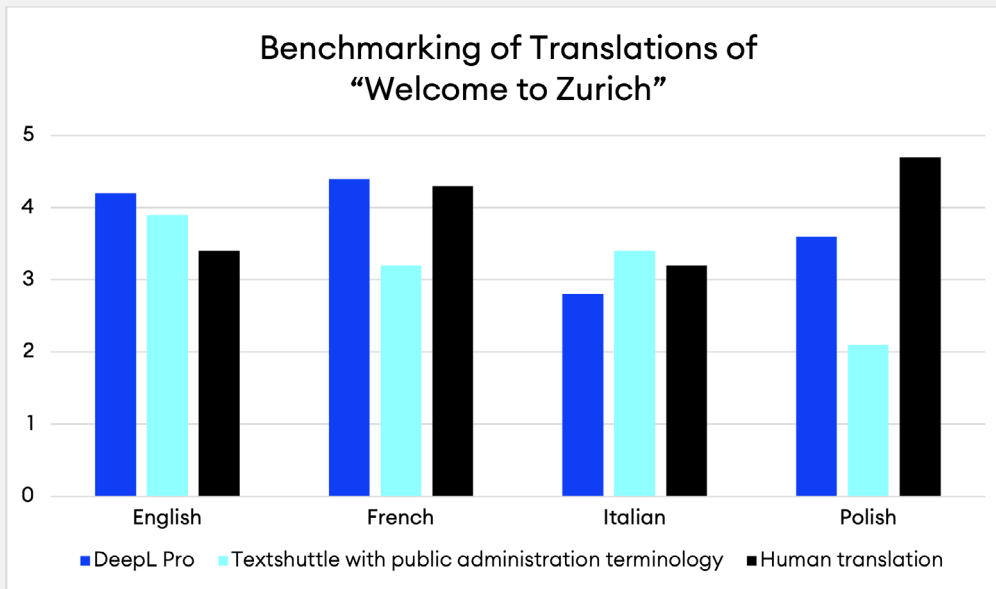
Textshuttle with public administration terminology

- **English:** Textshuttle's English translation was rated positively, especially in terms of accuracy, text flow and completeness. This underscores the capabilities of Textshuttle as a local provider focussed on widely spoken languages in Switzerland, such as English.
- **French:** the results for French were neither particularly good nor particularly bad. Text flow was rated positively.

- **Italian:** a clear added value was achieved in regard to the Italian translation by virtue of the public administration terminology. Text flow was also rated positively.
- **Polish:** the Polish translation was rated rather poorly. Even the stored public administration terminology failed to produce any positive effects in the overall context.
- **Varia:** text flow in the Textshuttle translations was rated very positively across all languages.

Human translation

- **English:** the English translation was rated as neither particularly good nor particularly bad, but with text flow receiving a positive rating.
- **French:** the French translation was rated as good, though partly perceived as awkwardly phrased (individuality of human translation).
- **Italian:** the Italian translation received fairly positive ratings, with particular mention of accuracy.
- **Polish:** for complex languages like Polish, human translation proved to be markedly better, showing the greatest discrepancy between machine and human translations along all criteria.
- **Varia:** communicative translations into English, French and Italian were rated negatively, which is likely due to how the evaluation was interpreted. The evaluations showed major discrepancies overall, which reflects the individuality of human translation.



Caption – see evaluation criteria:

- Fully applies (4–5)
- Tends to apply (3–4)
- Neither nor (2–3)
- Tends not to apply (1–2)
- Does not apply (0–1)

Benchmarking conclusions

These benchmarking results deliver important insights into the strengths and weaknesses of the various translation services. They show that machine translation can almost compete with human translators for some languages, especially Switzerland’s national languages and for English. However, machine translation often faces difficulties with more complex languages, or with specific specialised terminology. These findings are valuable for further optimising and adjusting translation tools in the domain of public administration.

- **Differences in quality between languages:** The benchmarking showed that translation quality varies considerably between the respective languages. This also applies to Switzerland’s national languages and to English, which are particularly relevant in public administration.
- **Few tendencies across all languages:** The study also revealed that differences in translation quality between the various languages were more significant than deviations in regard to specific

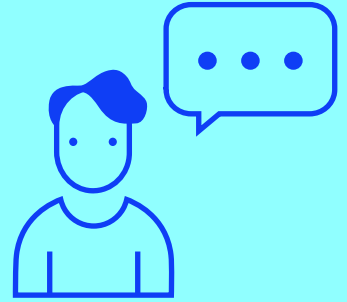
evaluation criteria. In most cases, none of the evaluated translation solutions was consistently more precise or more fluent in all languages than the other evaluated solutions. This emphasises the fact that every language has its individual challenges for machine translations.

- **Caution with some languages:** Special caution is called for with respect to certain languages which are not widely spoken in Switzerland and that have a high level of complexity (in this case study: Polish), as the quality of machine translation offered by some tools may still be insufficient.
- **Human translation in the blind test:** The human translations did not consistently score best in the blind test. Possible reasons for this could be that these translations are freer and more focussed on conveying the meaning, and are thus not a word-by-word translation of the source text. In the case study at hand, differences in the evaluations are more likely a result of the evaluation system than of translation quality.

- **Expert evaluation tendencies:** Although the evaluations provided by the experts varied greatly in some cases, clear tendencies were nonetheless discernible. It must be noted here that the results would certainly have been different with other evaluating experts, as three evaluations are not representative.
- **Added value by virtue of public administration terminology:** Adapting translation services through storing public administration terminology can offer clear added value, as the example of the Italian translation carried out by Textshuttle shows.

v.

Conclusions and recommendations



This chapter provides a summary of the findings and recommendations based on the evaluations of two case studies: use of machine translation within the Commercial Register of the Canton of Schwyz and benchmarking of translations for the Integration Unit of the Canton of Zurich. These case studies offered insights into the practical use and challenges of machine translation in public administration. In particular, they highlighted the level of performance of various translation tools –ranging from generic online tools right up to specialised translation services– in the context of different languages and specialised terminologies. These findings provide the basis for the following conclusions and strategic recommendations to help drive forward the use and development of translation technologies in public administration moving forward.

- **Human translation as the gold standard:** Despite advances made in the domain of machine translation, human translation is still the gold standard for legally reliable and officially certified documents. This study has also shown that there are significant challenges of machine translation in regard to less critical translations as well.
- **Variability and quality of human translations:** The study showed human translations to have a high degree of variability in the context of public administration, and as not per se superior to machine translations.
- **Limitations to generic translation tools:** Generic translation tools reach their limits when having to record and correctly translate public administration jargon specific to Switzerland. The quality

of translations varies considerably depending on language.

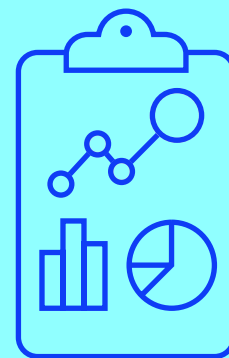
- **Added value through adaptations to public administration terminology:** Adapting translation services by storing public administration terminology offers clear added value. Doing so can considerably improve the accuracy and consistency of translations.
- **Advantage of Swiss providers:** Compared to international providers, the assumption is that Swiss providers of machine translations will invest more in the quality of translations in Switzerland's national languages, and will also take better account of idiosyncrasies of Swiss languages. Compared to many international and freely available providers, and regardless of the quality of translations, Swiss machine translation providers can achieve significant advantages with respect to data security by offering local data storage.
- **Handling generic models and online machine translation services:** In the everyday work setting, public administration employees often resort to generic models and freely available online machine translation services. A specific translation tool managed by the respective public administration entity is thus recommended, in order to ensure quality and data security. Ideally, these tools would also allow for translation of sensitive data without any additional measures.
- **Development of specific translation models for public administration:** Training and adjusting translation models, based on public adminis-

tration texts and specific needs of the various offices, offer significant advantages. The creation of a comprehensive translation model for public administration, which can be adapted to individual preferences and terminologies, presents an effective strategy to improve the quality of translations.

- **Significance of expertise and preferences:** The expertise of specialists is indispensable for fine-tuning. Different preferences of public administration offices or units must be taken into account in this regard.
- **Trade-offs in regard to specifically trained models:** Whereas specifically trained models work well within narrowly defined areas of use, they show weaknesses when it comes to more general translation tasks. However, there are clear signs that further development of large language models (LLMs) will lead to improved general applicability.
- **Need for checking by trained translators:** Despite steady improvement of machine translation systems, critical review and adjusting by trained translators is still indispensable, especially with regard to official and important documents.
- **Integration into public administration applications:** For the sake of productivity, translation tools should be integrated into public administration applications, so as to enable a more efficient and seamless use. In practice, this involves considerable effort given that the integration into existing systems is often technically demanding.

VI.

Annex with benchmarking results



The benchmarking results deliver valuable insights into the performance and limitations of various translation services. They also offer important points of reference for further development of and adaptations to translation tools in the context of public administration, especially with a view to the needs of newcomers in Switzerland.

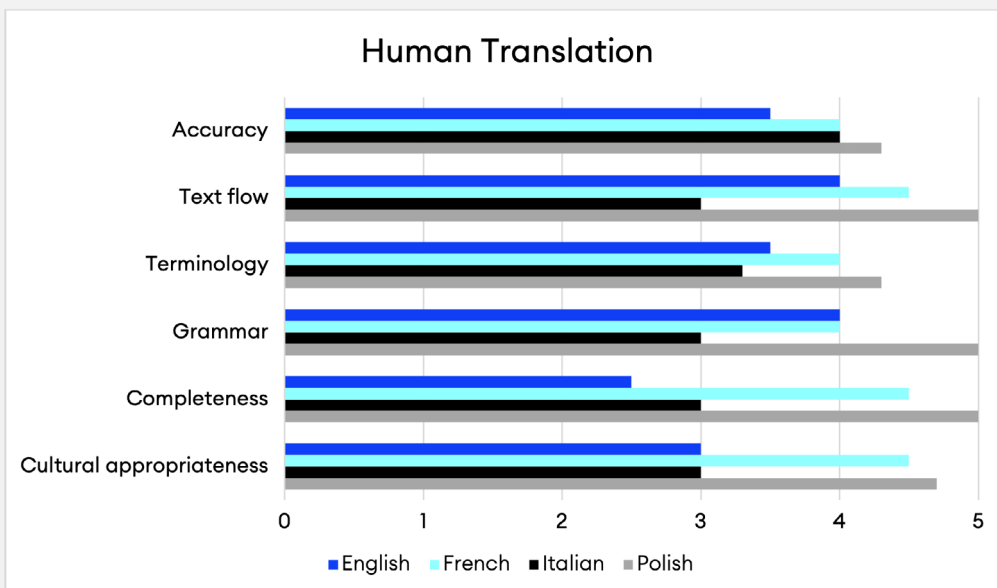
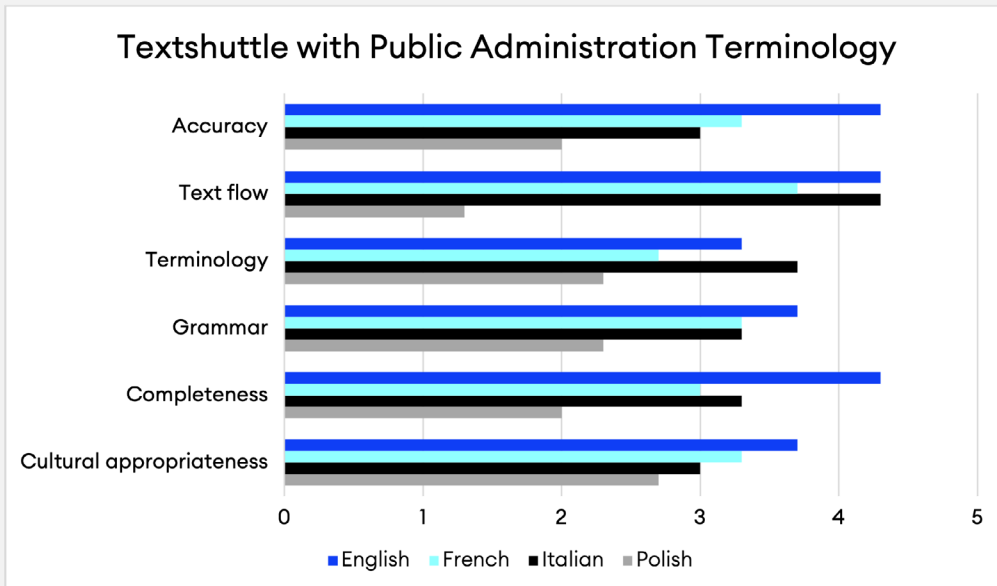
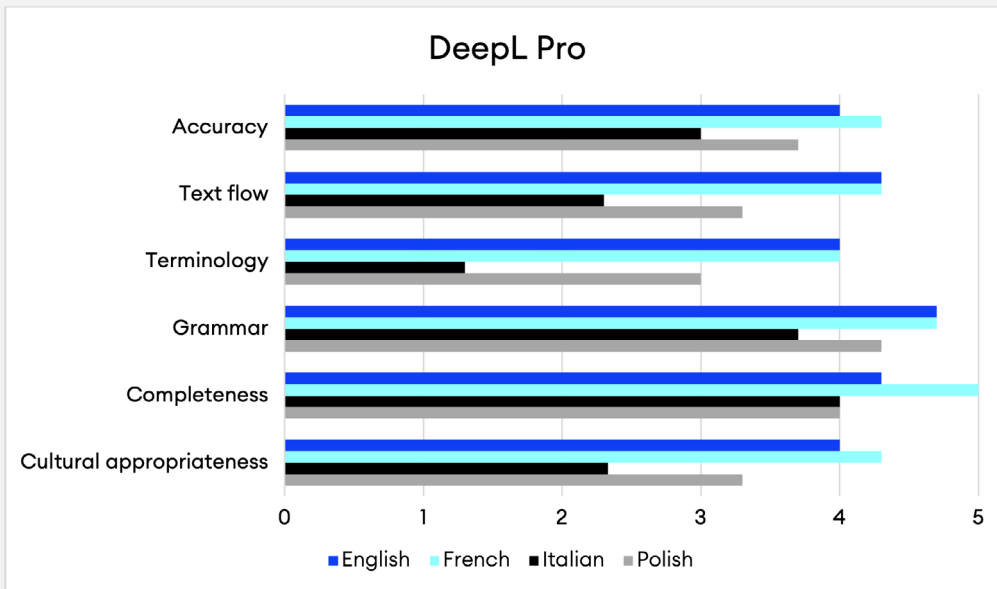
Below the criteria drawn on for evaluating the translations:

- **Accuracy:** does the translation accurately convey the original message?
- **Text flow:** does the translation flow well and is it easy to read?
- **Terminology:** is, in particular, public administration terminology, e.g. “residence permit”, used correctly and consistently?
- **Grammar:** are the translations correct grammar-wise?
- **Completeness:** has all content, without adding or omitting information, been translated?
- **Cultural appropriateness:** does the translation take cultural nuances and sensitivities of the target group, e.g. newcomers with a migratory background, into account?

Note that the following rating is based on just three evaluations performed by trained translators in the respective languages (i.e. a total of twelve evaluations). What is presented here is not a scientific study, but a case study from practice. The results cannot be generalised beyond the presented case (see chapter four). The aim is to show how public administration can evaluate machine translation solutions in view of specific cases of use. The evaluation is based on six criteria. In addition, the evaluating translators documented specific examples (inaccuracies, errors, etc.).

Caption – see evaluation criteria:

- Fully applies (4-5)
- Tends to apply (3-4)
- Neither nor (2-3)
- Tends not to apply (1-2)
- Does not apply (0-1)



Glossary

ChatGPT an advanced large language model of OpenAI that can generate and understand human-like texts, including for machine translation..

DeepL Pro a machine translation service based on neural networks and known for high-quality translations with a good text flow in various languages.

Google Translate a widely used machine translation service by Google that supports numerous languages and is known for its fast, user-friendly translations.

Graphics processing unit (GPU) a specialised processor originally designed for computer graphics, but also for training and use of machine learning models, including neural networks.

Large language models (LLMs) advanced AI models which process large quantities of text data in order to understand and generate language, and which form the basis for many modern translation systems and language assistants.

Machine translation a process by which computer programmes translate texts automatically from one language into another, which is particularly useful in public administration for fast translation of large quantities of text. Advanced models, such as Transformer, improve accuracy and natural-sounding language of translations.

Neural networks computer models inspired by the human brain which enable machine learning and are fundamental for advanced machine translation systems.

Translation model AI-based translation models use machine learning and neural networks to automatically translate a text from one language into another. These models learn from comprehensive datasets and take context, idioms and cultural nuances into account

Individuals and organisations involved in this report

Expert interviews

Mara Bertamini, Textshuttle

Nadia Sambuco, Integration Unit of the Canton of Zurich

Othmar Aeschi, Commercial Register of the Canton of Schwyz

Paula Reichenberg, Neur.on

Authors



Dr. iur. Stephanie Volz,

Regulatory Expert Innovation Sandbox for AI, ITSL University of Zurich



Raphael von Thiessen,

Head of Innovation Sandbox for AI, Division of Business and Economic Development, Canton of Zurich

Case studies provided by the Innovation Sandbox for AI

The companies Neur.on and Textshuttle served to implement case studies within the “Innovation Sandbox for AI”. Both organisations simultaneously submitted project proposals in relation to machine translation to the AI Innovation Sandbox. With the Commercial Register of the Canton of Schwyz and the Integration Unit of the Canton of Zurich serving as implementation partners, two case studies were carried out between January and December 2023. The content of this report is based on these specific case examples.

Imprint

Publisher

Division of Business and Economic Development,
Canton of Zurich
Metropolitan Area Zurich Association
Innovation Zurich

Project conception and coordination

Raphael von Thiessen
Location Promotion Canton of Zurich
8090 Zurich
raphael.vonthiessen@vd.zh.ch

Concept in collaboration with:

Stephanie Volz
Isabell Metzler
Patrick Arnecke

Authors

Dr. iur. Stephanie Volz
Raphael von Thiessen

Design

Sibylle Brodbeck, sibyllebrodbeck.ch

Publication

This report is published exclusively in digital format
and in the languages German and English

Translation

Mila Myrsep, Word for Word Ltd. Liab. Co.

Copyright

All contents of this publication, especially texts and
graphics, are copyright protected. The copyright is
held by the Location Promotion Canton of Zurich.
The publication can be passed on to third parties
with the copyright information, and it may be
quoted from with complete source references.

© 2024 | Canton of Zurich

Project Steering

Division of Business and Economic Development,
Canton of Zurich
Statistical Office, Canton of Zurich
Division of Digital Government, Chancellery,
Canton of Zurich
Office for Economy, Canton of Schwyz
Metropolitan Area Zurich Association
ETH AI Center
Center for Information Technology, Society, and Law
(ITSL) University of Zurich