

AI in interpreting: Ethical considerations

Ildikó Horváth* 

Department of Translation and Interpreting, ELTE University, Budapest, Hungary

ORIGINAL RESEARCH PAPER

Received: March 20, 2021 • Accepted: October 11, 2021

© 2022 The Author(s)



ABSTRACT

Artificial intelligence (AI) and machine learning technologies have impacted on the language mediation market with the spread of machine translation (MT) and the creation of sub-tasks such as text preparation for translation and post-editing. Up until this time, the impact of machine interpretation on the interpretation profession cannot be felt to the same extent as that of MT on the translation profession. Technological advances, however, have not come to an end and nowadays fully-automated machine interpretation and AI-based computer assisted interpreting (CAI) tools are increasingly common in the interpreting profession. However, the use of AI and big data in interpreting raises several ethical questions in terms of data protection and confidentiality. The earliest references to MT date back to the 1930s. Despite this long history ethical considerations in MT have rarely been discussed in Translation Studies, and to our knowledge, they have not been discussed at all in Interpreting Studies. This article first examines how AI can be used in interpreting as well as the various tools already available, then discusses the ethical considerations raised by the use of AI in general and in interpreting in particular.

KEYWORDS

artificial intelligence, interpreting, ethics, automated speech translation, AI-based CAI, online interpreting delivery platforms

1. INTRODUCTION

Interpreting as a profession is at a crossroads today. New technologies have been gaining ground in the interpretation market, leading to new tools and new interpreting situations. The digital

* Corresponding author. E-mail: horvath.ildiko0821@gmail.com

booth, video remote interpreting, online interpreting delivery platforms, the use of smart pens and tablets in consecutive interpreting, and CAI-tools for interpreters have been present for some time now. However, we can safely say that technological development, reinforced by external circumstances making interpreters work from home or hubs, is revolutionizing interpreting. Its impact is already being felt by professionals involved in the delivery of interpreting services (interpreters, private market players providing interpreting services, conference technicians, event organisers as well as large public institutions offering interpreting services, or professional associations). It is also felt by interpreter trainers who are adapting their training courses to the new reality, where the use of artificial intelligence (AI) is also gaining ground.

There is no single definition of AI, and it seems that it is easier to describe it than define. Green (2018), for example, says that AI “seeks to recreate particular aspects of human intelligence in computerized form”, and asserts that it is a “broad category, including such diverse abilities as vision, speech recognition and production, data analysis, advertising, navigation, machine learning, etc., and just about anything computers can do” (Green, 2018: 10–11). The aim of AI developments is to create formalized strategies, which enable computers to behave in a human-like manner, almost as humans (Wilss, 1993). AI is knowledge programmed into computers and their capacity to learn. There are two types of AI that need to be distinguished: ‘strong AI’ and ‘weak AI’. Artificial General Intelligence (AGI) or ‘strong AI’ or as Etzioni and Etzioni (2017) call it ‘AI minds’, which would presuppose general human cognitive abilities such as thinking, learning, reasoning, planning or creativity. AGI is not available yet (see also Bostrom & Yudkowsky, 2014).

What we have, however, is ‘weak AI’ tools and solutions which can take over and carry out some tasks from humans, such as online shopping or advertising, smart homes, urban infrastructure, self-driving cars, military drones, etc. Kirchner (2020) calls this ‘Functional-AI’ “because these machines can do this one thing and they can do it with enormous precision and speed but they cannot do anything else”. For example, AI beat the world champion in Go but if we present “the Go playing machine with a game of chess or a simple poker game it would be unable to perform even at beginner’s level” (Kirchner, 2020: 2). Furthermore, AI uses deep learning technologies, a form of machine learning (ML) and big data, where enormous amounts of datasets are fed into AI systems. ML today is advancing at an unprecedented speed. Green sees three reasons for this rapid progress: (a) huge increases in the size of datasets, and (b) an increase in computing capacity as well as (c) a “huge improvement of ML algorithms and more human talent to write them” (Green, 2020).

2. AI IN INTERPRETING

AI and ML technologies have also impacted on the interpretation market, and artificial intelligence-based technologies can be used in **automated speech translation** and in computer-assisted interpreting tools. The first experiments to create an automatic interpreter took place at the end of the 1980s and early 1990s. However, language technology available at that time allowed for only a very basic and limited performance of machine interpretation tools. Today, there are several AI-based devices available on the market that attempt to fully automatize the interpreting process both in the consecutive and in the simultaneous mode. A common feature of these devices is that they have been developed for a limited number of specific communication situations. They are used to interpret the most frequent phrases and expressions between



different languages in well-defined contexts such as travel, humanitarian missions, medical care, university lectures, military contexts and also where human interpreters are not available.

Most of these speech-to-speech (S2S) devices use the **cascade model**, consisting of several steps: speech-to-text (STT) conversion using automatic speech recognitions (ASR), machine translation (MT) and text-to-speech (TTS) synthesis (Fig. 1). The first step involves writing down the source language (SL) spoken text, then a translation system converts the written SL text into written target language (TL) text, and the third step involves converting this written TL text to spoken TL speech. Some devices include an extra step after ASR before writing down the SL text to filter the live speech and eliminate disfluencies, restarts, uhms etc.

The cascade model is still the most developed and frequently used model in commercialized S2S devices although end-to-end (E2E) models, which go directly from the SL speech to the machine translation component, are also feasible (Jia & Weiss, 2019).

From a technological point of view all three steps are problematic (Downie, 2020), but in terms of AI ethics the MT element raises the most concerns. MT is the technology computers are using to model natural language processing (NLP) and the human translation process. MT is not a new phenomenon: the first attempts to automate human translation date back to the 1930s (Austermühl, 2001), and it has been regularly discussed in Translation Studies since the 1950s (e.g., Bar-Hillel, 1951; Hutchins, 1986; Melby, 1981; Sager, 1994; Somers, 1992; Vauquois, 1976; Wilss, 1993). MT has gone through three main stages: (1) rule-based from the beginnings, (2) corpus-based (example-based and statistical) in the first few years of the 2000s, and (3) neural network-based machine translation (NMT). Neural network technology appeared in machine translation around 2007 but its use on the translation market became widespread only in 2015, and by 2017 NMT systems outweighed statistical MT systems (Koehn, 2017). NMT meant a breakthrough because the quality of TL texts improved in certain language pairs (for more details see Moorkens, 2018: 377–378).

The neural networks forming the basis for NMT are machine learning systems that use big data and deep learning (Koehn, 2017). Deep learning is a technology where machine learning algorithms use several layers and simulate the functioning of the human brain with the help of robust datasets. The importance of the data fed into these systems cannot be underestimated because the quality of datasets determines translation quality. Today the internet offers an enormous amount of data, which is a positive development in MT. However, if datasets fed into an NMT system are unchecked and of low quality, the machine will use them all the same and provide them in the TL texts at the end of the translation process and also use it for deep learning, which implies that mistakes will be reinforced. Another concern is data efficiency (Niehues, 2020), which means that there are still languages in which there is not enough available data. Furthermore, current systems ‘see’ more data than a human being in their entire lifetime but their performance lags well behind that of humans.

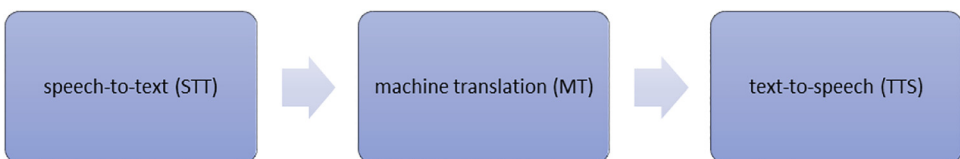


Fig. 1. The cascade model of the automated S2S process



The other domain where AI can be applied in interpreting is **terminology management CAI tools**. One of the conditions for high quality interpreting is thorough, efficient and rapid preparation for the assignment by searching, processing and consolidating information as well as terminology based on the theme of the event. Previously, interpreters used to look for information in libraries or journals or contacted professionals in person. Today, this is mainly done on the internet where it is significantly easier and more efficient to carry out content and terminology preparation using online encyclopaedias, multilingual electronic dictionaries, terminology databases and parallel text banks, which makes assignment preparation easier and more efficient. The terminology work in the case of simultaneous interpreting, however, is also done during interpreting. For this purpose, at least some of the processes built in terminology management devices need to be automated.

The needs for terminology management CAI tools were defined at the beginning of the 2000s, when Rütten designed the model for such software consisting of 5 modules: (1) Online and Offline Research, (2) Document Management, (3) Terminology Extraction and Analysis, (4) Terminology Management, and (5) Trainer. The ‘Trainer’ module is intended to facilitate learning the terms that the interpreter has to know by heart for the assignment (Rütten, 2004: 173–174). The first CAI terminology tools appeared soon afterwards. They were simple and interpreter-friendly in terms of their architecture and functionalities, and made the management of Word and Excel table-like multilingual glossaries possible. These tools did not provide help during the act of interpreting, they simply stored a database from which terminological data could be extracted.

Recently, we have seen several attempts to develop softwares using ASR for interpreters, which provide real time support for the interpreter in the digital booth. Recent developments in neural network technology and deep learning have led to improvement in the field of CAI tools, which enable terminology management and work during simultaneous interpreting in real time. These are what Fantinuoli calls “second generation” CAI tools, for example Intragloss and InterpretBank, which “offer advanced functionalities that go beyond terminology management, such as features to organise textual material, retrieve information from corpora or other resources (both online and offline), learn conceptualised domains, etc.” (Fantinuoli, 2018: 165). The main aim is to provide help with the most typical problem triggers mentioned by Gile as early as the 1990s during simultaneous interpreting, such as numbers, proper names and acronyms (Gile, 1995/2009). These CAI tools would serve as an ‘artificial boothmate’ and offer help with the lexical elements of the SL speech by transcribing the SL speech in real time on the interpreter’s computer screen, extracting technical vocabulary and providing their translation and recording proper names, figures and measures.

3. AI ETHICS

AI is rapidly changing our lives and transforming our society, both positively and negatively. This transformation is already having a significant ethical impact, since as Green (2020) puts it, “AI, as the externalization of human intelligence, offers us in amplified form everything that humanity already is, both good and evil” (Green, 2020).

Ethics as a discipline deals with what is morally right or wrong, good or evil as well as choices made by individuals based on what they perceive as moral duty or obligation. Bird, Fox-Skelly, Jenner, Larbey, Weitkamp, and Winfield (2020) argue that AI ethics “is concerned with the



important question of how human developers, manufacturers and operators should behave in order to minimise the ethical harms that can arise from AI in society, either arising from poor (unethical) design, inappropriate application or misuse” (Bird et al., 2020: 2).

According to Green (2020), the most important areas of AI relevance for ethics are the following: technical safety; transparency and privacy; beneficial use and capacity for good; malicious use and capacity for evil; bias in data, training sets; unemployment/lack of purpose and meaning; growing socio-economic inequality; environmental effects; automating ethics; moral deskilling and debility; AI consciousness, personhood, and ‘robot rights’; AGI (artificial general intelligence) and superintelligence; dependency on AI; AI-powered addiction; isolation and loneliness; effects on the human spirit (Green, 2020).

AI systems are gaining in autonomy and are increasingly taking decisions. One of the issues raised in this respect is that they can make bad decisions if they encounter new situations that have not been fed into the system. Furthermore, because humans do not know exactly how neural networks and such technologies as machine learning (ML) function, AI systems are often black boxes (Bird et al., 2020; Shaw, 2019; Siau & Wang, 2020).

Dignum (2018, 2019) underlines AI’s positive impact on our lives in terms of health, productivity and safety, and she argues that trust in AI systems is essential. For this reason such systems “must be introduced in ways that build trust and understanding, and respect human and civil rights”. Furthermore, in the light of growing AI capacities to make decisions, responsibility needs to be reconsidered and calls for “frameworks to guide design choices, to regulate the reaches of AI systems, to ensure proper data stewardship, and to help individuals determine their own involvement”. In her view, responsible AI “is about human responsibility for the development of intelligent systems along fundamental human principles and values, to ensure human flourishing and wellbeing in a sustainable world” (Dignum, 2018: 1). She further argues that responsibility for ethical AI lies with human researchers and developers who “must make fundamental human values the basis of our design and implementation decisions” as well as the users and owners of AI systems and „governments that legislate about their introduction in different areas, educators, the social organisations providing awareness and critical assessment in their specific fields and all of us” since we need to know “our rights and duties when interacting with these systems” (Dignum, 2019: 5).

Siau and Wang (2020) identify AI ethical issues as those (1) emerging at the design and development phase of AI systems, such as data bias and data privacy and transparency; (2) caused by AI, for example, in terms of unemployment or wealth distribution; (3) in relation to robots’ rights, i.e. “the concept that people should have moral obligations towards intelligent machines” (Siau & Wang, 2020: 76).

A serious ethical concern in AI is data bias, i.e. the fact that in the datasets used to teach algorithms certain social groups are less well represented than others, or negative connotations are attached to them. Since AI systems recycle the data and learn from the datasets they have been trained on they can amplify, for example, gender or racial bias contained in the training data. Boddington (2017) notes that the reason for this “may be because the training datasets for the algorithms are themselves biased in some way, or because the operation of the algorithm itself creates bias” (Boddington, 2017: 16).

Thus the quality and fairness of data play an essential role from the point of view of ethics, since bias most often “occurs when machine learning applications are trained on data that only reflect certain demographic groups, or which reflect societal biases” (Bird et al., 2020: 15). This can have far-reaching implications, for example in law enforcement or national security, it



“could result in some demographics being unfairly imprisoned or detained” or „in some individuals being unfairly refused loans”. Another problem lies in the fact that AI tools are ‘black boxes’, meaning that “it is impossible for the consumer to judge whether the data used to train them are fair or representative” (Bird et al., 2020: 16).

It also needs to be highlighted that datasets must be regularly maintained and updated once fed into AI systems, which is especially true for natural language data since natural language is an open system and can be compared to a living organism, which is continuously changing.

4. INTERPRETER ETHICS

A lot has been written about ethics in interpreting across all types and modes of interpreting. For the purposes of our discussion, it suffices to say that the main ethical principles are reflected in codes of ethics. Several authors have overviewed the recurring principles defined by such codes in the Interpreting Studies literature. Having reviewed various codes of ethical behaviour for interpreters, Kermit, for example, identifies three of them: (1) discretion and confidentiality; (2) neutrality and reluctance towards carrying out other tasks apart from interpreting and (3) all that is said should be translated accurately (Kermit, 2007: 242).

For court and community interpreting, Schweda Nicholson identifies seven recurrent themes in such codes: (1) the interpreter’s overall role; (2) competence and required skills; (3) impartiality; (4) completeness and accuracy; (5) conflicts of interest and grounds for disqualification; (6) confidentiality; (7) continuing professional development (Schweda Nicholson, 1994: 82–95).

In her work on court interpreting Edwards (1995) develops the following themes under the heading of ethics: secrecy, impartiality and having no opinion, keeping out of the case, unwillingness to admit error (Edwards, 1995: 63–71). According to Gentile, Ozolins, and Vasilakakos (1996) in codes of ethics for interpreters ‘the most basic general considerations are confidentiality, impartiality and conflicts of interest’ (Gentile et al., 1996: 58). Horváth (2013) also reviewed five codes of ethics for court interpreters and identified the following recurrent themes: (1) neutrality; (2) confidentiality; (3) professional behaviour; and (4) linguistic accuracy.

Bancroft (2005) carried out the “largest survey of existing codes yet conducted” and identified the following “five (near-)universal or widespread ethical principles”: (1) competence; (2) integrity; (3) confidentiality; (4) neutrality and (5) fidelity. Setton and Prunč (2015) add that “*transparency* [emphasis in original] has been put forward as a key principle in recent literature, although it appears only obliquely in the wording of existing codes” (Setton & Prunč, 2015: 146).

Since there have been initiatives to develop software using AI in conference interpreting, the principles laid down in AIIC’s Code of professional ethics need to be mentioned. These are secrecy; prohibiting deriving any personal gain from confidential information; continuing professional development; protecting the dignity of the profession; qualification; accepting only one assignment for the same period; collegiality; safeguarding the interests of the profession. The central tenet of AIIC’s code is high quality professional conference interpreting, for which the right working conditions need to be ensured (AIIC, 2018).

5. AI ETHICS AND INTERPRETING

Based on the discussion of AI ethics and interpreter ethics above, the following ethical considerations emerge as the most burning issues in AI-based applications used in interpreting: data



bias and data quality, data privacy and data ownership as well as transparency. In fact, unfair and biased datasets, data quality and security as well as secrecy are the most serious ethical risks that need to be taken into consideration when developing and using AI-based interpreting applications.

5.1. Data bias and data quality

AI and MT quality are important for speech translation systems because one of the central components of the cascade models is MT. If data entered into the MT systems running behind speech translation is of unreliable quality, biased or convey a certain bias, it raises concerns of an ethical nature as well.

Training data for language models comes to a large extent from data freely available on the internet. It has the advantage that there is an enormous amount of data that can be used to train such models. A major disadvantage is that if huge unchecked datasets are used to train language models, then the system recycles and replicates data that may be biased, unethical or flawed, thus amplifying them. Google has published about BERT (Bidirectional Encoder Representations from Transformers), their own language model, that some disability-related expressions such as ‘cerebral palsy’ or ‘blindness’ are associated with negative sentiments in the model (Hutchinson et al., 2020).

This is also true in the case of AI-based MT systems, which form the central component of automatized speech translation models. UNESCO (2019) has recently published a preliminary study on AI ethics, which includes the analysis of such systems and states that there are several quality risks related to MT (e.g., lack of linguistic and conceptual correspondence between languages, the untranslatable nature of certain contextual and textual connotations, etc.). Despite the significant improvement of MT systems in recent years, this may result in the fact that MT “is often too unreliable to be used, for instance, in fields where lexical and conceptual precision is crucial, or in cultural expression and literature”. In addition, AI-based MT negatively impacts on linguistic diversity, because MT is likely to be “primarily developed for the main world languages, especially English” because the large datasets required by this technology are rarely available for less widely spoken languages (UNESCO, 2019: 16).

In Translation Studies one of the rare publications on the subject of the ethics of MT and sustainable MT deals with fair MT (Moorkens, Kenny, & de Carmo, 2020), and expresses a need to take AI ethics into consideration when discussing MT. Canfora and Ottmann (2020) distinguish three domains where NMT poses ethical risks: (1) the damage caused to clients and users by erroneous NMT in domains where safety is of critical importance, (2) the issue of liability for NMT results containing errors, (3) the cyber risks of NMT, for example, in the case of the use of free online MT engines (Canfora & Ottmann, 2020).

Data bias and data quality are also relevant to AI-based CAI software, because these tools also use MT engines. Some of them use freely available large MT systems, or non-commercial databases available free of charge on the internet for translators but also for the general public. One such database is the IATE, the shared terminology database developed and maintained by the European Union. Concerning the reliability of data, on the IATE webpage itself there is a disclaimer saying that “[s]ome of the material in IATE is very old and has never been properly checked, so its quality is bound to be lower than we would like [...]. It is therefore important that you assess each solution on its merits. A term with a low reliability and no additional information probably shouldn’t be taken at face value” (<https://iate.europa.eu/faq>).



5.2. Data privacy and data ownership

Kenny (2011) concludes that ethical issues in MT are of concern to developers of statistical machine translation (SMT) systems (at the time of her paper's publication the latest technology was SMT), commissioners and consumers of MT, source text writers, translators whose translations are used to train SMT systems, post-editors, trainee translators and translation teachers. Of these actors several play a role in the creation of datasets and the final quality of MT products. Another ethical concern in this respect is whether commercial MT "developers indicate the human provenance of the data" since "the role of translators in creating vital data has been mostly downplayed or ignored". This issue still prevails and is closely linked to data ownership. Drugan and Babych (2010) mention cases where translation resources are shared such as in the case of Google Translator Toolkit. In this case ethical issues arise when the system retains source and target language texts to be used as training data for Google Translate systems, and confidentiality problems surface, such as texts containing commercial or personal information. Another issue of an ethical nature here is that those who use Google Translate do not know if the data (or translations) they are reusing have been shared by their owner or not. This also means that translators lose the ability to be acknowledged and have control over consent.

This brings us to another topic, namely copyright, which is a very complex one since copyright issues lead inevitably to legal considerations, and legal systems vary around the world. This makes it difficult if not impossible to deal with it in a harmonised way concerning such a global phenomenon as MT. Armstrong noted as early as 1997 that the collection of written data is a major concern in "copyright issues and protecting individual rights" as in the case of medical reports, for example. Furthermore, "gaining permission to use and distribute the material can be a problem" [...] partly because it is hard to find "the responsible authority to grant such rights" [...], thus "copyright issues remain an important legal problem at national and international level" (Armstrong, 1997).

On the same subject, Yanisky-Ravid and Martens (2019) stress that Google Translate, Microsoft Translator, DeepL and Systran translate copyrighted written materials (emails, advertisements, news, articles, songs, literature, books, etc.) thus creating derivative work under the current copyright regime in the US because "legally *authors*, and not *users* or search engines, have the exclusive right to control translations of their works" (Yanisky-Ravid & Martens, 2019:103). However, these companies do not usually contact right holders for permission, which means that they "have neither obtained licenses to use the works nor paid royalties for them. As a result, on an everyday basis many are violating copyright law". So the question arises why wealthy technology companies such as Google, Amazon, and Microsoft should have free access to authors' derivative rights (Yanisky-Ravid & Martens, 2019: 106).

Another ethical concern, which is not directly linked to copyright but is still worth mentioning is the fact that since "such platforms make money off of users' translation queries, their use of the original works is commercial and possibly supplants the market for an authorized translation" (Yanisky-Ravid & Martens, 2019: 103). The authors acknowledge the value of AI translation systems in enabling communication and making cultural exchange possible on a global level for everyday users. However, they advocate programmes adopting fair design, thus respecting copyright. For this there is a need for digital and legal tools that will balance the competing interests, and policymakers should create norms and guidelines. Because of the global nature of the issue, the World Intellectual Property Office (WIPO) should also publish its relevant guidelines.



In Europe, the European Union adopted in 2016 its general data protection regulation (GDPR), i.e., the regulation on the protection of personal data (Regulation 2016/679), a detailed text on the processing and the free movement of personal data of natural persons. However, as Bird et al. (2020) note, it only applies „to personal data, and not the aggregated 'anonymous' data that are usually used to train models” (Bird et al., 2020: 13). This also adds to the argument that data subjects have limited rights to and control over how their data are used.

In this respect, another ethical issue concerns the ASR function, which enables the live transcription of speeches in AI-based terminology software developed for conference interpreters working in the simultaneous mode. More precisely, data privacy and data security, as well as the confidentiality of meetings are at stake. For example, in InterpretBank’s workflow “the acoustic signal that the interpreter receives in the headset is sent to the sound card of the computer equipped with the ASR-CAI tool. The audio signal is then sent to the InterpretBank Application Programming Interface (API) that operates on a server located in Dresden, Germany, and returns the real-time transcript of the speech. InterpretBank uses the Google Cloud Speech-to-Text API2 as the ASR of choice” (Defrancq & Fantinuoli, 2020). In terms of data privacy and protection the trustworthiness of data transfer may raise concerns. Because it is a very recent and not widely used tool it also needs to be clarified how and when speakers are informed and whether they should be asked to consent to their speech being processed on distant servers. If yes, how and in what form is their consent to be expressed, how are they informed of the fact that the content of their speech, the acoustic signal and the text, go outside the venue, situation and the audience they have intended it for. First, interpreters may very easily become inadvertent infringers of the confidentiality principle, one of the strictest and most widely published ethical principles of the profession, because the acoustic signal is transferred from their computer to the server. Second, they may become inadvertent infringers of their speakers’ rights as well.

Data privacy and protection are also factors to be taken into account when promoting and using online remote interpreting platforms, both in the consecutive and simultaneous modes for spoken and signed language interpreting as well. These platforms make the organisation of interpreted events possible with participants and interpreters joining the meetings from locations from all around the world. The use of interpreting delivery platforms surged in the wake of lockdowns during the COVID-19 pandemic in spring 2020, when onsite events either had to be cancelled or were moved online. It is worth mentioning that the technology had already been there for some time but was used relatively rarely before the health crisis. It is impossible to foresee the future, however it seems safe to say that the long term impact of the pandemic will be that online remote interpreting, either from home or from interpreting hubs, is here to stay. Probably it is not going to be used to the extent it was during the crisis because once the health crisis is over, personal meetings are expected to resume. Nevertheless, such platforms have emerged from the crisis as tools that make it possible for international meetings to continue during the crisis. Not only the private market players but also large international institutions such as the EU and the UN had to adapt to this new situation and use online remote interpreting delivery platforms.

This is not to deny the usefulness and viability of such tools and the need for the interpreting profession to keep up with technological development. However, it is important to raise awareness of the ethical issues entailed by the use of such tools and the need to safeguard interpreters and their clients in terms of their personal and professional data by promoting fair use of AI in interpreting.



6. TRANSPARENCY

The third ethical concern related to AI-based tools in interpreting is transparency. Transparency of processes and decision making in AI systems is one of the basic criteria for AI ethics. As [Bostrom \(2011\)](#) states, as early as 2011, “[i]t will be increasingly important to develop AI algorithms that are not just powerful and scalable, but also *transparent to inspection*” ([Bostrom, 2011: 2](#)). However, even today we cannot be sure where large developers of MT systems derive their data from.

[Dignum \(2018, 2019\)](#) mentions the Responsible Research and Innovation (RRI) model, where one of the major building blocks is *Openness and Transparency* (along with Diversity and Inclusion, Anticipation and Reflexivity, Responsiveness and Adaptiveness). Openness and transparency “require open, clear communication about the nature of the [AI] project, including funding/resources, decision-making processes and governance”. She further argues that “[m]aking data and results openly available ensures accountability and enables critical scrutiny, which contribute to build public trust in research and innovation” ([Dignum, 2019: 50](#)). Transparency in AI systems means that we are aware of how the systems make decisions and what kind of data they use. She elaborated the so-called ART (Accountability, Responsibility and Transparency) principles for trustworthy AI, where transparency can be summarised as “the capability to describe, inspect and reproduce the mechanisms through which AI systems make decisions and learn to adapt to their environment, and the provenance and dynamics of the data that is used and created by the system”. Furthermore, it “is also about being explicit and open about choices and decisions concerning data sources and development processes and stakeholders” ([Dignum, 2019: 54](#)).

Transparency is essential for building trust in AI, which means that “it should always be possible to find out *why* an autonomous system made a particular decision, especially if that decision caused harm” ([Bird et al., 2020: 31](#)), an endeavour made difficult for at least two reasons. First, the ways these systems make decisions is rather obscure even for their developers. Second, it is difficult for lack of transparency regarding the provenance of data, for example, in the case of AI applications developed by private companies. As [Bird et al. \(2020\)](#) argue, “most AI applications are developed by private companies, there is not always enough transparency about these data, in contrast to the traditional scientific method that warrants the validity of results by requiring replicability, i.e. the possibility to reproduce them by repeating the same experiments” ([Bird et al., 2020: 11](#)).

7. CONCLUSION

The positive impact of AI-related technological development should be acknowledged insofar as it may lead to facilitating the interpreter’s work by potentially decreasing cognitive load and may facilitate rendering interpreting services in places that are not easily accessible physically, or languages for which interpreters are hard to find locally. However, in the case of such a surge in technological development where AI is used for the first time to a considerable extent and where the spread of cloud-based online interpreting delivery platforms is so sudden, we should bear in mind that there are several new ethical risks that emerge. The most immediate ones are linked to the core of the ethical principles guiding the interpreting profession, regardless of the mode and situation, namely secrecy and the confidential treatment of written and spoken information forming the basis for trust in a client-interpreter relationship. Another one is interpreting



accuracy, faithfulness, in other words, interpreting service quality. Closely linked to this is the data quality and transparency in AI-based systems used either in speech translation systems or in terminology management CAI systems for interpreters. Also, the fact that interpreters are working from home without the help of a technician raises not only technical questions but also questions of responsibility and liability in terms of data privacy and protection. Another question that needs to be answered is who will be held responsible if AI-based automated machine interpretation systems make errors. AI ethics should be a basic consideration when developing AI applications and not only an afterthought. For this reason, it seems that the time has come to discuss AI ethics in interpreting today and include AI ethics in the profession's standard setting instruments, such as codes of ethics or guidelines for professional practice. Furthermore, this aspect of interpreter ethics needs to be included in training programmes.

REFERENCES

- Armstrong, S. (1997). Corpus-based methods for NLP and translation studies. *Interpreting*, 2(1/2), 141–162. <https://doi.org/10.1075/intp.2.1-2.06arm>.
- Austermühl, F. (2001). *Electronic tools for translators*. St. Jerome Publishing.
- Bancroft, M. (2005). *The Interpreters World Tour: An environmental scan of standards of practice for interpreters*. The California Endowment.
- Bar-Hillel, Y. (1951). The present state of research on mechanical translation. *American Documentation*, 2(4), 229–237.
- Bird, E., Fox-Skelly, J., Jenner, N., Larbey, R., Weitkamp, E., & Winfield, A. (2020). *The ethics of artificial intelligence: Issues and initiatives*. European Parliamentary Research Service.
- Boddington, P. (2017). *Towards a code of ethics for artificial intelligence*. Springer.
- Bostrom, N. (2011). Infinite Ethics. *Analysis and Metaphysics*, 10, 9–59.
- Bostrom, N., & Yudkowsky, E. (2014). The ethics of artificial intelligence. In K. Frankish, & W. Ramsey (Eds.), *The Cambridge handbook of artificial intelligence* (pp. 316–334). Cambridge University Press. <https://doi.org/10.1017/CBO9781139046855.020>.
- Canfora, C., & Ottmann, A. (2020). Risks in neural machine translation. *Translation Spaces*, 9(1), 58–77. <https://doi.org/10.1075/ts.00021.can>.
- Defrancq, B., & Fantinuoli, C. (2020). Automatic speech recognition in the booth. Assessment of system performance, interpreters' performances and interactions in the context of numbers. *Target Online*, <https://doi.org/10.1075/target.19166.def>.
- Dignum, V. 2018. Ethics in artificial intelligence: introduction to the special issue. *Ethics and Information Technology*, 20(1–3). <https://doi.org/10.1007/s10676-018-9450-z>.
- Dignum, V. (2019). *Responsible artificial intelligence. How to develop and use AI in a responsible way*. Springer.
- Downie, J. (2020). *Interpreters vs machines. Can interpreters survive in an AI-dominated world?* Routledge.
- Drugan, J., & Babych, B. (2010). Shared Resources, Shared Values? Ethical Implications of Sharing Translation resources. In V. Zechev (Ed.), *Proceedings of the Second Joint EM+/CNGL Workshop "Bringing MT to the User: Research on Integrating MT in the Translation Industry" (JEC 2010), 4 November 2010*, (pp. 3–9). Denver.
- Edwards, A. B. (1995). *The practice of court interpreting*. John Benjamins.



- Etzioni, A., & Etzioni, O. (2017). Incorporating ethics into artificial intelligence. *The Journal of Ethics*, 21, 403–418. <https://doi.org/10.1007/s10892-017-9252-2>.
- Fantinuoli, C. (2018). Computer-assisted interpretation: challenges and future perspectives. In I. Durán-Muñoz, & G. Corpas Pastor (Eds.), *Trends in E-tools and resources for translators and interpreters* (pp. 153–174). Brill.
- Gentile, A., Ozolins, U., & Vasilakakos, M. (1996). *Liaison interpreting. A handbook*. Melbourne University Press.
- Gile, D. (1995/2009). *Basic concepts and models for interpreter and translator training*. John Benjamins. <https://doi.org/10.1075/btl.8>.
- Green, B. P. (2018). Ethical reflections on artificial intelligence. *Scientia et Fides*, 6(2), 9–31. <http://dx.doi.org/10.12775/SetF.2018.015>.
- Green, B. P. (18 August 2020). Artificial intelligence and ethics: Sixteen challenges and opportunities. <https://www.scu.edu/ethics/all-about-ethics/artificial-intelligence-and-ethics-sixteen-challenges-and-opportunities/> Accessed 22 February 2021.
- Horváth, I. (2013). *Bírósági tolmácsolás [Court Interpreting]*. ELTE Eötvös Kiadó.
- Hutchins, W. J. (1986). *Machine translation: Past, present, future*. Ellis Horwood.
- Hutchinson, B., Prabhakaran, V., Denton, E., Webster, K., Zhong, Y., & Denuyl, S. (2020). *Social biases in NLP models as barriers for persons with disabilities*. Google. <https://www.aclweb.org/anthology/2020.acl-main.487.pdf>. Accessed 22 February 2021.
- Jia, Y., & Weiss, R. (19 May 2019). Introducing Translatotron: An end-to-end speech-to-speech translation model. <https://ai.googleblog.com/2019/05/introducing-translatotron-end-to-end.html>. Accessed 22 February 2021.
- Kenny, D. (2011). The ethics of machine translation. In *Proceedings of the XI New Zealand Society of Translators and Interpreters Annual Conference 2011*, Auckland, New Zealand, 4-5 June 2011. http://doras.dcu.ie/17606/1/The_Ethics_of_Machine_Translation_pre-final_version.pdf. Accessed 22 February 2021.
- Kermit, P. (2007). Aristotelian ethics and modern professional interpreting. In C. Wadensjö, B. Dimitrova Englund, & A.-L. Nilsson (Eds.), *The critical link* (Vol. 4, pp. 241–249). John Benjamins.
- Kirchner, F. (2020). AI-perspectives: the Turing option. *AI Perspectives*, 2(2), <https://doi.org/10.1186/s42467-020-00006-3>.
- Koehn, P. (2017). *Statistical machine translation. Draft of chapter 13: Neural machine translation*. <https://arxiv.org/pdf/1709.07809.pdf>.
- Melby, A. K. (1981). Translators and machines – Can they cooperate? *Meta*, 26(1), 23–34. <https://doi.org/10.7202/003619ar>.
- Moorkens, J. (2018). What to expect from Neural Machine Translation: a practical in-class translation evaluation exercise. *The Interpreter and Translator Trainer*, 12(4), 375–387. <https://doi.org/10.1080/1750399X.2018.1501639>.
- Moorkens, J., Kenny, D., & de Carmo, F. (Eds.). (2020). Fair MT. Towards ethical, sustainable machine translation. *Special issue of Translation Spaces*, 9(1), <https://doi.org/10.1075/ts.9.1>.
- Niehues, J. (13 November 2020). *Automated speech translation: Challenges and approaches. A Speech Odyssey – Automated speech translation: Challenges and approaches, scale, uses and edge cases. Artificial intelligence and the interpreter webinar series. AIIC UK & Ireland*. <https://www.youtube.com/watch?v=90E9J1zPxLY>.
- Rütten, A. (2004). Why and in what sense do conference interpreters need special software? *Linguistica Antverpeinsia*, 3, 167–177. <https://lans-tts.uantwerpen.be/index.php/LANS-TTS/article/view/110/57>.



- Sager, J. C. (1994). *Language engineering and translation: Consequences of automation*. John Benjamins. <https://doi.org/10.1075/btl.1>.
- Schweda Nicholson, N. (1994). Professional ethics for court and community interpreters. In D. L. Hammond (Eds.), *Professional issues for translators and interpreters. American Translators Association Scholarly Monograph Series* (pp. 79–97). John Benjamins. <https://doi.org/10.1075/ata.vii.10sch>.
- Setton, R., & Prunč, E. (2015). Ethics. In F. Pöchhacker (Ed.), *Routledge encyclopedia of interpreting studies* (pp. 144–148). Routledge.
- Shaw, J. (2019). Artificial intelligence and ethics. Beyond engineering at the dawn of decision-making machines. *Harvard Magazine*, 2019(1), 44–74. <https://harvardmagazine.com/sites/default/files/pdf/2019/01-pdfs/0119-44.pdf>.
- Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management*, 31(2), 74–87. <http://doi.org/10.4018/JDM.2020040105>.
- Somers, H. L. (1992). Current research in machine translation. In J. Newton (Ed.), *Computers in translation. A practical appraisal* (pp. 198–207). Routledge.
- Vauquois, B. (1976). Automatic translation – A survey of different approaches. *Statistical Methods in Linguistics*, 1976, 127–135. <http://www.mt-archive.info/SMIL-1976-Vauquois.pdf>.
- Wilss, W. (1993). Basic concepts of MT. *Meta*, 38(3), 403–413. <https://doi.org/10.7202/004608ar>.
- Yanisky-Ravid, S., & Martens, C. (2019). From the myth of Babel to Google Translate: Confronting malicious use of artificial intelligence–Copyright and algorithmic biases in online translation systems. *Seattle University Law Review*, 43(1), 99–168. <http://dx.doi.org/10.2139/ssrn.3345716>.

ONLINE RESOURCES

- AIIC (2018). *Code of professional ethics*. https://aiic.org/document/6299/Code%20of%20professional%20ethics_ENG.pdf. Accessed 22 February 2021.
- A Speech Odyssey– Automated speech translation: Challenges and approaches, scale, uses and edge cases. Artificial intelligence and the interpreter webinar series. AIIC UK & Ireland. <https://www.youtube.com/watch?v=90E9J1zPxlY>. Accessed 22 February 2021.
- IATE FAQ. <https://iate.europa.eu/faq>. Accessed 22 February 2021.
- Regulation (EU) 2016/679 Of The European Parliament And Of The Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Accessed 22 February 2021.
- UNESCO. (2019). *Preliminary study on ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000367823>. Accessed 22 February 2021.

Open Access. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited, a link to the CC License is provided, and changes – if any – are indicated. (SID_1)

