Chapter 6

Ethics and machine translation

Joss Moorkens
Dublin City University

Neural machine translation (MT) can facilitate communication in a way that surpasses previous MT paradigms, but there are also consequences of its use. As with the development of any technology, MT is not ethically neutral, but rather reflects the values of those behind its development. In this chapter, we consider the ethical issues around MT, beginning with data gathering and reuse and looking at how MT fits with the values and codes of the translator. If machines and systems reflect value systems, can they be explicitly "good" and remove bias from their output? What is the contribution of MT to discussions of sustainability and diversity? Rather than promoting an approach that involves following a set of instructions to implement a technology unthinkingly, this chapter highlights the importance of a conscious decision-making process when designing a data-driven MT workflow.

1 What do we mean by ethics?

The field of Ethics examines morality, good and evil, right and wrong, and addresses questions about how best to live. The earliest surviving texts on this topic originated in Egypt, Babylonia, and India. Greek philosophers such as Socrates introduced the notion of the "good life", one that is worthy and admirable. Aristotle made this a little more concrete by identifying a set of virtues that, when practised, would allow human beings to flourish. These virtues are still abstract, and are not always helpful when deciding whether an action is right or wrong. Subsequently, philosophers and ethicists have suggested ways to decide on a right or moral course of action, based, for example, on the probability of providing the best result for the majority, or by only acting on good or pure motives.

A problem is that what is well-motivated or produces the happiest result for one group may not necessarily produce an equally positive result for another.

Joss Moorkens. 2022. Ethics and machine translation. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 121–140. Berlin: Language Science Press. DOI: 10.5281/zenodo.6759984

There is a tension between the idea that an action can be universally good and moral, such as upholding justice or truthfulness, and the position that values may differ depending on the person or group under examination. There have been many suggestions for ways of untangling whether an action is ethical or unethical based on agency, relationships, or a surrounding narrative. This is where theoretical ethics moves into applied ethics, in trying to guide how we should act in a given situation.

Applied ethics in a working situation will often involve a set of codes or standards to guide professional behaviour. If these codes are too restrictive, they may hamper potential progress or societal benefits. Rigid codes could also cause difficulties as ethical decisions are rarely binary and choices may be governed by the unique scenario and pressures brought to bear on the person making that choice. For this reason, different fields of applied ethics have sprung up to consider common problems and dilemmas within their particular context. This chapter will draw on the fields most relevant to machine translation (MT) including computer and information ethics and data ethics when discussing the ethical use of MT by humans in system development. §3 on the ethical use of MT in professional workflows will draw on business ethics and the growing literature on translation ethics. §4 on computers as ethical agents will draw on machine ethics and computer and information ethics. The final sections will draw on more recent diverse work on ethics and artificial intelligence when looking at sustainability and diversity.

Ethics is a growing area of interest in technology in general, as technology becomes an increasingly integral part of all of our lives and many regions move towards ubiquitous computing. We need to be aware of the impact of the choices we make when we design, implement, or use technology. There is an assumption often expressed that technology is ethically neutral and that bias may be introduced only in our use of that technology. However, the consensus among ethicists and philosophers of science is that technology is not ethically neutral, but rather reflects the values of the designer. These values govern the problem addressed by the technology, the decision to create the technology, the method of implementation, its intended users, the references or training data used, the processing of that data, the location and security of data storage, and the limits to access to the technology based perhaps on cost or geographical location.

The speed and scale of technological development means that regulation is inevitably a step or two behind and we are thus reliant on ethical behaviour on the part of engineers and developers. We rely, to a greater or lesser extent, on large technology companies with political power and wealth to act in our collective best interests, but a series of reports and revelations in recent years have

demonstrated that our confidence in these companies is sometimes misplaced. While technology opens up access to new avenues and benefits, it also exposes the public to risk. By discussing some of the choices and risks inherent in the development and use of MT systems in this chapter, I hope to guide users in making informed and ethical decisions. The focus throughout is mostly (but not entirely) on MT for dissemination, where MT output is not the final step in production.

2 The ethical use of MT by humans in MT system development

2.1 Case studies of data use

This section looks at legal and ethical issues regarding the use of translation data in MT system development. As Pérez-Ortiz et al. (2022 [this volume]) attest, data-driven MT and particularly neural MT (NMT) requires a lot of data for training.[1] While it may be perfectly legal to reuse translation data for training MT systems, is it ethical? It may be helpful to introduce some of the issues to be discussed in this section by considering the following examples.

Translator A has freely signed a contract with their regular employer to carry out a translation on a freelance basis using a proprietary web-based platform, giving explicit permission for their translation data to be reused for MT training. The employer trains MT systems using the data from Translator A and others. In time, NMT quality improves for Translator A's language pair to the extent that the company moves its translation work to post-editing and imposes a unilateral 30% discount on their per-word payment rate. This discount is applied on the basis that productivity has generally improved by roughly 30%, visible to the company from the translation activity data gathered via the translation platform. In order to raise revenue, the company decides to sell MT services externally. This includes some work for an arms manufacturer.

Translator B is opposed to MT as a matter of principle. B accepts work for a company that expects translators to submit their translation memory with translated target texts, which they will repurpose for future human translation. Translator B is not aware that the work has been automatically assigned by an automated project management system, but there is no translation brief and no direct communication with the company. Translator B is also not aware that the company will soon be acquired by a large conglomerate who will use all available

[1] *Data* refers to recorded information in any form, usually stored digitally, and when data are available in huge volumes and processed at scale, we talk of *big data*.

data for MT training and offer it for sale. The data should have all personal information removed before being shared (see §2.5), but this information is retained by accident when the data are uploaded to one purchaser. The company tries to keep this quiet so as to avoid liability.

What ethical issues can you identify in these two scenarios? What would change if the employers or translators made different ethical decisions? In the following subsections, we look at data ownership, permissions, distribution, privacy, and legal frameworks for data sharing. These subsections will, it is hoped, help guide your thinking about the above questions.

2.2 Data ownership

The commonly-used metaphor of data as oil suggests that big data is naturally occurring, whereas in reality it was originally created by humans. The exponential growth in data produced in recent years has meant that there is now more data available for MT training and more demand for translation than ever before, far more than is possible for human translators to produce. MT training data are usually stored in the form of parallel or aligned bilingual segments of text that have been translated by humans, often in translation memories (although MT output is also sometimes used for MT training). The source of this translation data is likely to be shared in public repositories, such as the European Commission's Directorate-General for Translation data, which can be "re-used and disseminated, free of charge… both for commercial and non-commercial purposes" (Steinberger et al. 2012: 457), privately held repositories of translation data, or parallel data crawled from the web.

The Berne Convention, first enacted in 1886, forms the legal basis of copyright for translations, considering them to be derivative works that "shall be protected as original works without prejudice to the copyright in the original work" (Article 2, World Intellectual Property Organization 1979). The convention grants the author of an original work the exclusive right to authorize a translation, although it fails to define "original work" or "originality", allowing for different interpretations in different jurisdictions. Troussel & Debussche (2014) believe that an argument could be made for originality in a creative translation, although this has yet to be tested in courts. The authors further believe that ownership rights to a translation memory database may be asserted where there has been "a substantial investment in either the obtaining, verification or presentation of the contents", according to the European Database Directive (European Parliament 1996: Article 7). In practice, translation memories are usually sent to the client, whether or not there is a contractual agreement in place for waiving any claim of ownership on the part of the translator.

At scale, big translation data has become a valuable resource for MT and machine learning system training (Moorkens & Lewis 2019a). This does not mean that translators receive any secondary payment however, and the granular reuse in MT training means that the source of training data is usually not identifiable. This is also true of data gathered by webcrawling for parallel texts. Translators A and B in our case study probably have little option other than to hand over their translation data and to accept the consequences, especially considering that most translators work on a freelance basis, and thus have limited scope for argument with their employers. It is reasonable to argue that a more equitable system of data ownership would contribute to the sustainability of the translation industry (see also §5).

2.3 Permission to use data

In some jurisdictions, it is considered that the employer who pays for a translation is the rightful owner, whereas in others, ownership may be transferred, granting permission for reuse. We might assume that there is "a degree of collegiality at play among those translators who favour resource sharing" with fellow translators, even with those who they do not know (Moorkens & Lewis 2019b: 8). However, the acceptance of this reuse for human translation may be eroded when translation data are instead used for training MT systems, especially among translators who believe that progress of MT technology is not in their best interests. Some translation contracts may explicitly state that translation data will be reused within human or machine translation workflows, but it is rare for translators to control how their work is reused. There is also no evidence of permission being sought or granted for reuse of webcrawled data.

This means that Translators A and B will contribute towards future projects, the purpose and end use of which will be opaque to them. Translator A may be ethically opposed to working for the arms manufacturing company, but will nonetheless be an unknowing participant in their work. This opacity is a problem faced more generally by those whose data are collected and reused, along with those who contribute work towards large technology projects without the opportunity to ask the questions "What is the final application and use of the products of my work?" and "Am I content or ashamed to have contributed to this use?" (see Moorkens 2020, Weizenbaum 1986).

When data are created during translation, depending on the format for recording and exchange, a number of attributes are recorded. These usually include a name or ID for the translator (see §2.5), the date and time of creation, language codes, software used, and a project ID. This information is useful for deciding

when and where to reuse the data based on the translator, the project, or the creation date. Translator activity data, including more detailed timings, editing actions, and records of individual keystrokes may also be recorded, particularly when a proprietary web-based platform is used, as in Translator A's project. Such data can be useful for monitoring translators' work, but is commonly removed for MT training so that only parallel sentence pairs are used. Once any possible identifying metadata (data about data) are removed, preferences for future use or reuse cannot be recorded and individual contributions cannot be measured, even if there is a retrospective change to agreements that means that contributions should earn a royalty. On the other hand, this will improve anonymization of translation data, which is important if the data are to be shared or exchanged.

2.4 Data distribution

In the early days of translation memory sharing, Topping (2000) wrote that, of individual translators, localization agencies and localization customers, only translators felt that it was ethical to share memories. This view seems quaint when we have companies in 2021 whose business model is shaped around amassing and reselling data for MT and other machine learning purposes.[2] This business model works, as the amount of training data and the level of care in curation of this data will make an impact on both the quality and value of an NMT system.

As mentioned in §2.2, certain datasets can be made freely available and distributed through projects such as Opus, the open parallel corpus (http://opus.nlpl.eu, Tiedemann 2012), due to their licensing agreements or because they are covered by the European Union (EU) directive (2003/98/EC) on the re-use of public sector information. This does not necessarily mean that translators have expressly given permission for all possible forms of reuse, but they are aware that the data will be shared with the general public.

Data may otherwise be distributed on the basis of agreements between companies, or due to one company being acquired by another, which is common within the language service industry (Moorkens 2020). It may be bought and sold or donated for research or philanthropic purposes, all without the necessity for approval from the data creator. This is perfectly legal as long as the data cannot be classed as personal data, in which case restrictions apply.

[2]Please see the introduction to this volume for more on machine learning. For the purposes of this chapter, we understand machine learning as a use of computers to achieve an end by inference from big data rather than from input of an explicit command.

2.5 Privacy, personal data

If data can render an individual identifiable, it can be considered personal data. This includes translation memories with a named or coded (pseudonymized) creator. Within the EU, the General Data Protection Regulation (GDPR)[3] has restricted the sharing and reuse of personal data since 2018, providing guidelines for national legislation that would impose heavy fines in the case of a data breach. This has had the effect of increasing cybersecurity and limiting the use of servers in non-EU locations. Any secondary use of personal data must be covered by the permissions given for the original use, with some exceptions for research purposes. There are a number of other national regulations outside of the EU that govern the use of personal data.

Companies should report any data breach, but there have been many media reports of breaches being covered up. One reason to do so is to avoid GDPR fines (of up to €20 million or 4% of annual turnover, whichever is the greater, according to Article 83 of the Regulation), but companies may also wish to avoid negative publicity, loss of consumer confidence, and loss of stock market value, in the case of publicly-traded companies. Ideally, transparency would lead to greater public trust in an organization, but the problem of data security in large organizations is not always well understood, with their sheer size presenting a data protection difficulty. What is more, not all data breaches are equal, as they can be due to "ethical hackers" who are employed by the organization to identify cybersecurity vulnerabilities or who identify vulnerabilities to protect the public, or malicious hackers who intend to access data for their own gain.

Once personal attributes are removed and data are anonymized, personal data becomes (just) data, and shareable.[4] Of course, even without the metadata, some data may be recognizable if its content or style is identifiable or if biometric data could be used to link to an individual. If data are shared or pooled, data from an individual or group may be used to make inferences, for example about attributes that are carefully protected by the GDPR such as race or sexuality. As these inferences are made on the basis of combining data rather than being explicitly contained in any single data set, they are not usually covered by the GDPR (Wachter & Mittelstadt 2019). This presents a risk to "group privacy", where a group may be discriminated against due to the content of data that does not identify any individual (Floridi & Taddeo 2016).

[3] EU Regulation 2016/679, available from https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[4] There are a number of ongoing efforts to automate anonymization for translation data at the time of writing, but this is difficult to do reliably.

Translators A and B, for example, could have their translation data aggregated with other personal data, allowing a third party to make inferences about them individually or as members of a group. The use of web-based platforms for translation is increasingly common, giving translators less control of their translation data and allowing surveillance of work activities. If personal circumstances lead to a temporary downturn in productivity or translation quality as gauged via translator activity data gleaned from the work platform, that could negatively affect their prospects for future employment. If identifiable translation activity data for an individual that encodes this downturn is shared outside of a single organization, that could have far-reaching consequences for that individual. This does not necessarily mean that it is unethical to monitor quality or productivity. An agency or company needs to be able to stand over their translations. However, by automating employment decisions or communication, as is the case with project management in the example of Translator B, a company will leave the translator with no opportunity to explain translation choices or to build a long-term relationship based on trust. There is no guarantee of ethical behaviour on the part of the translator or user of a platform at the best of times, but when relationships are purely transactional, research has shown that trust and the assumption of good faith on both sides are particularly undermined, with knock-on effects on satisfaction and performance (Whipple & Nyaga 2010).

2.6 Ethics in MT evaluation

Rossi & Carré (2022 [this volume]) look at methods of human and automatic MT evaluation. There are a number of ethical issues related to MT evaluation that are worth raising here. Most MT systems' output is evaluated automatically during training and again afterwards for a quick, easy, and cost-effective measure of quality. In competitive shared tasks, where development teams pit their systems against one another, either automatic or crowd evaluation tends to be used. Based on these evaluations, output may be considered to have reached parity with human translation quality if a segment-level crowd rating or automatic evaluation achieves the same score as a "reference" human-translated target text. If the evaluation score for MT surpasses that of the human reference, it is considered "super-human" output.

This language is problematic, especially when disseminated more widely in research publications and marketing materials, which in turn may be reported in news media, giving the impression that MT produces perfect quality output without risk and that human translators are no longer necessary. However, automatic evaluation metrics tend to show little correlation with human judgment

and there are several problems with crowd evaluation, in which anonymous and presumably untrained internet users rank or rate segments of sequential translated material. Freitag et al. (2021) found that expert (professional translator) evaluators produced markedly different results to crowd workers when carrying out a detailed error analysis with access to full source and target documents, and demonstrated a clear preference for human rather than MT output. Additionally, there are issues with crowd work related to poor rates of pay, labour conditions, opaque user rating systems, and use of humans (crowd workers) as research participants without oversight or ethical review. Nonetheless, published results based on automatic evaluation and crowd work are almost always reported cursorily, devaluing human translation and creating an unrealistic and uncritical perception of MT among the general public, including translation clients. This perception increases the likelihood of MT being introduced into professional workflows.

3 The ethical use of MT in professional workflows

3.1 Translation stakeholders

The case studies presented in §2.1 prompt us to think not just about ethics and MT system development, but also about the ethical use of MT in professional workflows. For example, the company who engages Translator A makes a unilateral decision to move translation production to post-editing, when ideally the introduction of MT into a workflow would be based on consultation and agreed by all stakeholders. The stakeholders in the use cases in §2.1 are the translation agency or language service provider, comprising a number of internal roles, and the freelance translators. In addition, the translation client should be aware that MT will be used as part of the translation process and cognizant of the attendant benefits and risks (see §3.2). Almost all research on post-editing productivity has shown a boost in output when compared with translation from scratch or using translation memory. However, the orthodoxy is that the use of translation automation should relate to the shelf-life and level of risk attached to the translated text, and the client relies on translation agency expertise in choosing an appropriate and cost-effective workflow. The end user relies on the client to provide them with a text that does not expose them to unexpected risk. In addition, Pym (2012) suggests that, when even a low-risk target text is made less comprehensible by poor quality MT, the translation may conform to the needs of the client (who reduces costs by applying light or full post-editing) but will require extra effort on the part of the end reader.

Translation software developers are also important stakeholders in the use of MT in workflows, and the values and related design decisions mentioned for technology developers in §1 also apply to them. A translation tool in use can reshape activities and their meaning. Interaction with MT may be via interactive MT, where MT is used for autosuggest and edited dynamically, or post-edited, appearing as an extra translation suggestion or automatically propagated within the target text window. A tool can focus on usability, for example incorporating familiar keyboard shortcuts, providing an uncluttered interface, and maximizing customizability. Translation activity data (see §2.3), if collected, can be made transparent to users so that they can see what is being collected and use it themselves if they wish. Alternatively, translators could work within a disempowering platform, accepting jobs as soon as they appear, with no visibility of data gathering and a very limited user interface, and have their performance rated with no option for feedback or discussion.

Translators have the option to accept or not accept work offered to them based on the text domain and working conditions. For those who are not aware of the variable ways that MT may be used within professional workflows, the decision to accept or not accept work may be difficult, particularly if the agency has not been transparent. It is also important that translators are transparent in their use of tools, particularly MT, so that agreed confidentiality arrangements are not breached, and risk is not introduced for other stakeholders (or themselves, as may be seen in §3.2) without their knowledge. For translators, the use of a *code of ethics*, a set of rules to guide ethical behaviour, falls under the rubric of deontological ethics. While such codes are associated with a narrow interpretation of the role of ethics in translation, they are nonetheless useful for decision-making.[5] As with many other professional organizations, translators' associations often provide such a code to encourage professional conduct, impartiality, honesty, and respect for confidential material. These codes also promote trust on the part of current and potential clients. At the time of writing, a review of many of these codes found no explicit mention of MT, even though (as we have just seen) the decision as to whether or not to make use of MT in a translation project may be an ethical one. Chesterman (2001), who has written widely about ethics and trust in translation, suggests a general ethics of service, focusing on loyalty to the terms and quality requirements of the client, to the source text and its author(s), and to the target text reader.

The use of MT in translation production does not necessarily entail a loss of quality, and the cost and effort of human translation is not appropriate for all

[5]Lambert (2018) and others propose that the assumption of neutrality in translation, central to many Codes of Ethics, is just as flawed as it is for technology.

types of texts, particularly those with a short shelf-life that present little risk. However, for critical texts in which a mistranslation introduces risk, the use of MT must be considered carefully and subject to review. There is some evidence that certain project managers do not want to know or prefer to turn a blind eye when their translators use MT (Sakamoto 2019), but there may be good reasons for translation clients to be aware of MT use and to stipulate in contracts whether or not it may be used.

3.2 Risk and liability

Translation contracts may attribute ownership of translation data or give permission for retention of a translation memory to the translator or client, as described in §2.2., Canfora & Ottmann (2020) introduce two other contractual areas relevant to the use of MT: liability and confidentiality.

Translators may be found to be in breach of contract due to negligence or failing in their duty of care to a client. Liability can only refer to human behaviour, which means that a person must bear responsibility for an injury or loss related to an error introduced by MT in a translation process. Liability aside, the fact that MT might introduce a risk to end users is an ethical problem. Raw MT should never be used for safety-critical content. Current translation and post-editing standards do not mention liability or risks to end users.

The ISO translation standard does, however, stress the importance of "safe and confidential handling […] of all relevant data and documents" (ISO 2015, 3.2.a). Users of free online MT systems grant service providers the right to use the data entered for online translation, and there have been instances where confidential and sensitive material has been made available through unthinking use of free online MT. This cybersecurity risk introduced by such MT systems is why Canfora & Ottmann (2020) feel that subscription MT services that do not retain data are a better option, and ideally recommend that companies use closed platforms where server architectures are not open to the public – or choose not to outsource translation work at all in order to protect confidentiality. Freelance translators, for their part, might object to the loss of control over translation data and translator activity data that a closed platform entails, as discussed in Sections 2 and 3.1.

The ISO standard for MT post-editing (ISO 2017) makes no mention of confidentiality or risks to data security, which is rather surprising considering that the process necessarily entails the use of MT, introducing the associated risks. Trust is a key part of risk reduction, as standards, guidelines, and contracts are only of

value if the translator feels that they are in a trust partnership, without which they may rationalize unethical behaviour (see, for example, Abdallah 2010).

Aside from concerns about risk, translators and users may not wish to use MT due to the processes described in §2 or due to the impact of artificial intelligence (AI) on the world of work and sustainability. The following section examines the latter point with respect to NMT.

4 Sustainability

4.1 Payment, conditions, job satisfaction among translators/post-editors

Translation is a highly skilled task, but portions of the workflow have been automated (to an extent) in the examples of Translators A and B, with automatic job assignment, the imposition of post-editing, and the repurposing of translation data for tasks that the translators may not expect. There is growing consensus that AI will have a major impact on work in many areas previously considered to be immune from automation. While this might not directly cause higher unemployment rates, the changes could affect economic returns, work organization, and skills management in ways that are difficult to predict. These are considerations for the future in many industries, but in translation the impacts are well underway for a couple of reasons. Firstly, MT post-editing has been the fastest-growing area of the translation market since 2010 or so, predating the shift from statistical to neural MT. Stockpiling of translation data has been commonplace since the advent of translation memory tools in the early 1990s, although the collection of translation activity data for monitoring and automation is relatively recent. Secondly, the largely freelance workforce means that translators have flexibility and autonomy, but work on a project-by-project timeframe. This has created a disparity of power, whereby translators have little say in processes and conditions that can be changed unilaterally by agencies and employers from one project to the next. The effect of the disparity of power is apparent from the discussions regarding data in §2. As the pace of mergers and acquisitions has increased, creating large publicly-traded translation conglomerates, the disconnect has grown between those making decisions on business operations and freelance workers doing translation, post-editing, revision, annotation, review, subtitling, or another of the vast and growing array of roles that engage directly with texts. Suggestions from the industry to automate project management and to use blockchain to attribute authorship or contribution are not likely to improve this situation. More generally, the translation industry has not historically

shown strong leadership on ethics and sustainability, as discussed by Moorkens & Rocchi (2021).

The early view of translation technology, as expressed by Kay in 1980, was that the translator should remain in control with technology assisting with work that is mechanical and routine, and possibly boring (Kay 1980). What we have seen instead for many translators is that their work has gradually been circumscribed. Some translators have seen their work reduced to quality checks, annotation, or the correction of repetitive errors in mixed-quality MT output. In the latter case, the MT output may even have been decomposed to individual sentences that lack any accompanying context, with some automatically passed and others marked for review.

Some translators enjoy post-editing and, even with discounts applied (as in the case in the first case study in §2), find the work worthwhile and lucrative. However, there is a balance to be struck between short-term efficiency and long-term gains for all stakeholders in a technologized translation process.

Workers find satisfaction in doing work that has meaning, in mastering their task, and in working with supportive colleagues. They are motivated by a sense of achievement and recognition for that achievement. If this is not a consideration in translation production, better workers will leave and there will be a shortage of skilled translators and/or post-editors. Such a shortage would affect reliable access to multilingual information and the gathering of high-quality bilingual data on which MT training relies. Docherty et al. (2008: 4) consider that a sustainable work system must satisfy the needs of many rather than few stakeholders, and that instead of focusing exclusively on "short-term, static efficiencies such as productivity and profitability; we must also focus on long-term, dynamic efficiencies such as learning and innovation". UN Sustainable Development Goal (SDG) 8 is to provide decent work and economic growth, but environmental sustainability, as addressed by SDGs 13 and 15, is also relevant to MT.

4.2 Environmental concerns

It might reasonably be argued that there is a contradiction between setting a goal for economic growth and for environmental sustainability. Cronin (2017) makes this point particularly about the growth dependency of the localization industry. The ICT industry, on which MT relies, requires the mining of rare metals and has a reputation for poor recycling and polluting. Neural MT is particularly resource intensive, requiring powerful GPUs (Graphical Processing Units) for training and large amounts of power. Strubell et al. (2019) estimate that training for one large transformer neural network model will produce almost five times

the CO2 output of a car (including fuel) during its full lifetime.[6] However, most training instances are far less resource-intensive than the one reported in this paper. Furthermore, while hardware becomes more powerful and costly to engineer and produce, optimization of power consumption and the potential to run massive amounts of parallel processes mean that the power required for training is dropping. Nonetheless, it remains the case that training an NMT system is costly and requires a good deal of power. How that impinges on the environment will depend on the source of that power. There is currently no agreed benchmark for power consumption when publishing details of MT systems, although some have been proposed in the context of suggestions for sustainable AI development. The point made strongly by Van Wynsberghe (2021) is that without a focus on sustainability in the development and deployment of AI (and, by extension, NMT), AI development itself will not be sustainable.

5 Diversity

5.1 Among developers and users

The cost and power requirements are a huge barrier to entry into NMT development. The data requirements meant that early systems had to use publicly available data (see §2), usually creating systems for major European languages. It comes as no surprise then, that initial published work on NMT was conducted mostly by well-resourced academic research groups in North America and Europe. This has changed somewhat for two main reasons. Firstly, large technology companies have thrown their weight behind research efforts in NMT, building very well-resourced teams that lead the way in optimizing MT systems between major languages. This means that many academic research groups struggle to compete in major European languages and have moved to the more "niche" area of low-resource and minority languages. Secondly, the ability to create *synthetic* parallel data by machine-translating monolingual data from the intended target language into the intended source language[7] has led to a jump in quality for under-resourced language pairs. Thus the Fifth Conference on Machine Translation (WMT20) includes translation in Inuktitut and Tamil to and from English. However, another way to improve quality for low-resource languages is to build large multilingual systems, which are usually the preserve of the large commercial teams.

[6] We note also that only the largest companies can afford the costs of training such large-scale models.

[7] MT researchers call this process "back-translation". It is not to be confused with "back-translation" used as a glossing technique in standard translation studies sources such as Baker (2018).

There has been no survey of the diversity of MT research teams. A search of papers will find a reasonable amount of research published on MT using Simplified Chinese, Bengali, and Hindi – languages with huge numbers of speakers, but ones that are non-European and, in the cases of Bengali and Hindi, historically under-resourced. It is probably less likely that a great deal of diversity will be found among the leaders of these research teams and those who set the research agenda. In discussions of bias outside of MT, such a lack of diversity has been highlighted as a problem that has contributed to a number of well-publicized errors and blind spots in systems that use machine learning, such as facial recognition tools intended to identify likely criminals that pick out disproportionately high numbers of ethnic minorities.

In their article on the societal impact of MT, Vieira et al. (2020: 13) find that inappropriate use of MT can cause harm to vulnerable people in medical and legal use cases reviewed. They find that it can exacerbate inequality, given the "disproportionate availability of data and resources for a relatively small number of the world's languages", combined with a lack of MT literacy among those deploying MT. On the other hand, there is also a benefit in democratizing communication, as more and more under-resourced languages are being catered for in free online MT systems. At the time of writing, Google Translate covers 109 languages, including such under-resourced examples as Chichewa, Scots Gaelic, Uyghur, and Tatar.

5.2 As reflected in MT outputs

In 2016, Jones calculated that, of over 6,000 non-endangered languages, only 1% were catered for by any sort of MT. This situation has improved a little due to the research efforts mentioned in §5.1, and as evinced by the growing number of languages covered by Baidu and Google. However, this takes place in a world where information tends to flow from well- to poorly-resourced languages. Because MT has been shown to exert source language interference to a greater degree than human translation (see Toral 2019), the worry is that poorly-resourced languages will be impoverished in the long run.

This could be the case for all machine-translated languages, especially if a shortage of new human-translated data means that MT systems struggle to keep up with contemporary language. Vanmassenhove et al. (2019) illustrate how lexical diversity is lost when NMT engines are trained up to the point of so-called convergence,[8] suggesting that if training was stopped at an earlier juncture, the NMT output produced would be less standardized and more lexically diverse.

[8]The point at which iterative NMT system training is stopped, as automatic evaluation scores show no improvement in output quality. See §7.2.

Vanmassenhove (2019) shows that this standardization of output presents an algorithmic bias that exacerbates existing gender bias in training data, whereby a noun, or other form, is most commonly associated with one or other gender (in binary gender systems), and the less common gender is standardized out of the output data. This can result in output that emphasizes societal bias, with genders assigned inconsistently, even within a single segment. The following section reflects on contemporary efforts to neutralize biased output and the broader role of computers as implicit ethical agents.

6 Computers as ethical agents

Mainstream discussions of ethics in AI are concerned with safety and expanding machine autonomy and the application of the technology to a greater number of tasks. Machine ethics, meanwhile, is distinguished by the fact that it sees "the machine" as a subject with *agency* (that is, a willingness and ability to act) rather than as an object. A small but growing body of research in MT is concerned with bias (see §5.2) and risk, expanding discussions to include the values inherent in MT and AI more generally (see §5.1). This leads to the notion of computers as implicit ethical agents, whereby ethical decisions are implicit in their design. While efforts are being made in this regard, we regularly see biased or discriminatory output shared online from scenarios that were not easily predictable for developers. The stories of unintended consequences of technology, particularly AI, when intelligence appears to be demonstrated by machines, have sparked a series of books and articles that demonstrate unethical uses of that technology.

The problematic aspect of the technology may involve data gathering, the algorithm used to inductively extract patterns from the data, or the nature of the data itself. The consequences of technology may also be intended, whereby the *affordances* of the technology, i.e. the uses and interactions suggested by the tool, nudge the user towards acting in a way that has negative repercussions for them or for others.

The research focus in the small amount of research on bias in NMT has been exclusively on gender bias rather than on other attributes such as race or sexuality, and a number of solutions have been proposed to "de-bias" the MT output. Suggestions for the correct use of gender and the removal of gender bias in NMT output have included the use of gender tags, similar to those that may be used for politeness and register; de-biasing word embeddings; and treating gender bias similarly to domain adaptation for NMT, using transfer learning on a small gender-balanced dataset (Tomalin et al. 2021). In 2020, Google rolled out a system

whereby, given certain languages, for every male-gendered output, an almost identical female-gendered output (and vice versa) is created. The user can then evaluate the different options and choose which one to use. This, of course, assumes a binary view of gender, but has improved the quality of Google's gender-specific translations overall (Johnson 2020).

The future may see computers act as explicit ethical agents, with the ability to process information about each situation and to autonomously determine the best or most ethical course of action. Technology has not yet reached that level of sophistication, which is one reason why machine creations cannot claim copyright and machines cannot be held liable for loss or injury. Even if or when this happens, we cannot assume that a computer will act ethically. We have established that technology is not benign. While its use can bring great personal and societal benefits, there are risks and consequences that may not have been considered before the technology was deployed.

There is no doubt that the availability of free online MT has aided communication for many, but the seamless interfaces and improving output may lead end users to assume that its use is harmless (and it usually will be). The professional translation workflow stakeholders in the use cases in §2.1 can be assumed to have expertise in language technology, but this is not true of the general public, who might expect the same level of coherence and comprehensibility in a machine translation that they would find in a (good) human translation. They may not even be aware that they are reading a machine translated text, and even if the output were to be labelled as MT-generated, they are unlikely to be aware of the risks of mistranslation.

7 Summary

As MT quality improves, the technology facilitates more communication either directly or as part of a translation workflow. However, there are ethical concerns to be considered by MT developers, translation buyers, translation agencies, translators, and consumers of translation. As with all technologies, neither MT development nor MT output should be considered neutral, but rather as promulgating the perspective of the developers or the translators who created the training data, in the tools for interaction with MT and in the output text. Uncritical reporting of positive MT evaluation results minimizes public awareness of risk and bias in MT output while potentially devaluing the work of human translators. Readers may find it useful to consider the issues raised in this chapter when working with and using MT, and reflecting on how the related processes fit with their own values, purposes, and principles.

Ethical considerations as laid out in this chapter begin with the source of translation and translator data, ownership, permissions, copyright, and mode of distribution. The ethical use of MT within professional translation workflows may depend on the attitudes of all stakeholders, rules of confidentiality, and the design decisions behind MT platforms. This relates to sustainability, modes of interaction with MT, and the degree of autonomy and ownership of the process allowed for translators. The methods by which we can ensure diversity and de-bias MT systems and data are perhaps least developed, and will no doubt require further discussion and adjustment over time.

References

Abdallah, Kristiina. 2010. Translator's agency in production networks. In Tuija Kinnunen & Kaisa Koskinen (eds.), *Translator's agency*, 11–46. Tampere: Tampere University Press.

Baker, Mona. 2018. *In other words*. London: Routledge.

Canfora, Carmen & Angelika Ottmann. 2020. Risks in neural machine translation. *Translation Spaces* 9(1). 58–77.

Chesterman, Andrew. 2001. Proposal for a hieronymic oath. *The Translator* 7(2). 139–154. DOI: 10.6509.2001.10799097.

Cronin, Michael. 2017. *Eco-translation: Translation and ecology in the age of the anthropocene*. London: Routledge.

Docherty, Peter, Mari Kira & A. B. (Rami) Shari. 2008. What the world needs now is sustainable work systems. In Peter Docherty, Mari Kira & A. B. (Rami) Shari (eds.), *Creating sustainable work systems: Developing social sustainability*, 1–22. London: Routledge.

European Parliament. 1996. *Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases*. https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:31996L0009.

Floridi, Luciano & Mariarosaria Taddeo. 2016. *What is data ethics?* DOI: 10.1098/rsta.2016.0360.

Freitag, Markus, George Foster, David Grangier, Viresh Ratnakar, Qijun Tan & Wolfgang Macherey. 2021. Experts, errors, and context: a large-scale study of human evaluation for machine translation. *Transactions of the Association for Computational Linguistics* 9. 1460–1474. DOI: 10.1162/tacl_a_00437.

ISO. 2015. *ISO 17100:2015. Translation services – requirements for translation services*. https://www.iso.org/standard/59149.html.

ISO. 2017. *ISO 18857:2017. Translation services – post-editing of machine translation output: Requirements*. https://www.iso.org/standard/62970.html.

Johnson, Marvin. 2020. *A scalable approach to reducing gender bias in Google Translate.* https : / / ai . googleblog . com / 2020 / 04 / a - scalable - approach - to - reducing-gender.html (20 May, 2020).

Kay, Martin. 1980. *The proper place of men and machines in language translation* (Report CSL- 80-11). Palo Alto, CA: Xerox Corporation.

Lambert, Joseph. 2018. How ethical are codes of ethics? Using illusions of neutrality to sell translations. *Journal of Specialised Translation* 30. 269–290. https://www.jostrans.org/issue30/art_lambert.php.

Moorkens, Joss. 2020. "A tiny cog in a large machine": Digital Taylorism in the translation industry. *Translation Spaces* 9(1). 12–34.

Moorkens, Joss & David Lewis. 2019a. Copyright and the reuse of translation as data. In Minako O'Hagan (ed.), *The Routledge handbook of translation and technology*, 469–481. London: Routledge.

Moorkens, Joss & David Lewis. 2019b. Research questions and a proposal for the future governance of translation data. *Journal of Specialised Translation* 32. 2–25.

Moorkens, Joss & Martha Rocchi. 2021. Ethics in the translation industry. In Kaisa Koskinen & Nike K. Pokorn (eds.), *The Routledge handbook of translation and ethics*, 320–337. London: Routledge.

Pérez-Ortiz, Juan Antonio, Mikel L. Forcada & Felipe Sánchez-Martínez. 2022. How neural machine translation works. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 141–164. Berlin: Language Science Press. DOI: 10.5281/zenodo.6760020.

Pym, Anthony. 2012. *On translator ethics: principles for mediation between cultures.* Amsterdam: John Benjamins.

Rossi, Caroline & Alice Carré. 2022. How to choose a suitable neural machine translation solution: Evaluation of MT quality. In Dorothy Kenny (ed.), *Machine translation for everyone: Empowering users in the age of artificial intelligence*, 51–79. Berlin: Language Science Press. DOI: 10.5281/zenodo.6759978.

Sakamoto, Akiko. 2019. Unintended consequences of translation technologies: from project managers' perspectives. *Perspectives* 27(1). 58–73.

Steinberger, Ralf, Andreas Eisele, Szymon Klocek, Spyridon Pilos & Patrick Schlüter. 2012. DGT-TM: a freely available translation memory in 22 languages. In *Proceedings of the 8th international conference on language resources and evaluation (LREC 2012)*. ELRA. https://www.aclweb.org/anthology/L12-1481/.

Strubell, Emma, Ananda Ganesh & Andrew McCallum. 2019. *Energy and policy considerations for deep learning in NLP*. https://aclanthology.org/P19-1355.pdf.

Tiedemann, Jörg. 2012. *Parallel data, tools and interfaces in OPUS*. In Proceedings of the 8th International Conference on Language Resources & Evaluation (LREC 2012). 2214-2218. Luxembourg: ELRA. http://www.lrec-conf.org/proceedings/lrec2012/pdf/463_Paper.pdf.

Tomalin, Marcus, Bill Byrne, Shauna Concannon, Danielle Saunders & Stefanie Ullmann. 2021. *The practical ethics of bias reduction in machine translation. Why domain adaptation is better than data debiasing*. 419–433. DOI: 10.1007/s10676-021-09583-1.

Topping, Suzanne. 2000. Sharing translation database information: Considerations for developing an ethical and viable exchange of data. *Multilingual* 11(5). 59–61.

Toral, Antonio. 2019. Post-editese: An exacerbated translationese. In *Proceedings of machine translation summit XVII*, 273–281. EAMT. https://www.aclweb.org/anthology/W19-6627/.

Troussel, Jean-Christophe & Julien Debussche. 2014. *Translation and intellectual property rights*. Luxembourg: Publications Office of the European Union.

Van Wynsberghe, Aimee. 2021. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics* 1. 213–218. DOI: 10.1007/s43681-021-00043-6.

Vanmassenhove, Eva. 2019. *On the integration of linguistic features into statistical and neural Machine translation*. PhD Thesis. Dublin City University.

Vanmassenhove, Eva, Dimitar Shterionov & Andy Way. 2019. Lost in translation: Loss and decay of linguistic richness in machine translation. In *Proceedings of machine translation summit XVII: Research track. Dublin: EAMT*, 222–232. https://www.aclweb.org/anthology/W19-6622.pdf.

Vieira, Lucas Nunes, Minako O'Hagan & Carol O'Sullivan. 2020. Understanding the societal impacts of machine translation: A critical review of the literature on medical and legal use cases. *Information, Communication & Society* 24(11). 1515–1532. DOI: 10.118X.2020.1776370.

Wachter, Sandra & Brent Mittelstadt. 2019. A right to reasonable inferences: Rethinking data protection law in the age of big data and AI. *Columbia Business Law Review* 2019(2). 1–130.

Weizenbaum, Joseph. 1986. Not without us. *Computers and Society* 16. 2–7.

Whipple, Daniel F. Lynch, Judith M. & Gilbert N. Nyaga. 2010. A buyer's perspective on collaborative versus transactional relationships. *Industrial Marketing Management* 39(3). 507–518. DOI: 10.1016/j.indmarman.2008.11.008.

World Intellectual Property Organization. 1979. *Berne convention for the protection of literary and artistic works*. (as amended on September 18, 1979). Geneva: WIPO. http://www.wipo.int/wipolex/en/details.jsp?id=12214.