LUBLIN STUDIES IN MODERN LANGUAGES AND LITERATURE Vol 48, No 3 (2024)
Maria Curie-Sklodowska University Press

E-ISSN 2450-4580

Ralph Krüger, TH Köln – University of Applied Sciences, Germany

DOI:10.17951/lsmll.2024.48.3.11-23

Outline of an Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication

SUMMARY
This paper first traces the AI-induced automation of the digitalised and datafied language industry, with a focus on neural machine translation and large language models. The paper goes on to discuss a range of digital literacies that have become increasingly relevant in the language industry in light of these technologies, i.e., *machine translation literacy*, *data literacy* and *artificial intelligence literacy*. After highlighting the interface between these three literacies, the paper drafts an outline of an artificial intelligence literacy framework for translation, interpreting and specialised communication. This framework intends to capture an extensive set of competencies required by stakeholders in the AI-saturated language industry.

KEYWORDS
language industry; artificial intelligence; neural machine translation; large language models; machine translation literacy; data literacy; artificial intelligence literacy

1. Introduction: AI-induced automation of the language industry

The rapid evolution of modern artificial intelligence (AI) technologies within the machine learning (ML) paradigm has fuelled the (semi-)automation of intellectual labour in the language industry in recent years (cf. ELIS Research, 2023, pp. 37–39). This AI-fuelled automation has been most pronounced in the translation sector, where powerful neural machine translation (NMT) systems based on the transformer architecture (cf. Vaswani et al., 2017) have led to a widespread shift in production processes from *machine-aided human translation* (MAHT) to *human-aided machine translation* (HAMT) or *machine-translation post-editing* (MTPE). The transformer architecture for NMT systems consists of an encoder and a decoder side. The encoder transforms a given source text into a numerical vector representation which can be processed by the underlying neural network. The decoder then uses this vector representation of the source text to produce the target translation. This encoder-decoder architecture can be split into an autonomous encoder side, which serves as the architecture of so-called *encoder-only language*

Ralph Krüger, Institut für Translation und Mehrsprachige Kommunikation, TH Köln, Ubierring 48, 50678 Köln, ralph.krueger@th-koeln.de, https://orcid.org/0000-0002-1009-3365

models like Google's BERT model (Bidirectional Encoder Representations from Transformers), and an autonomous decoder side, which serves as the architecture of so-called *decoder-only language models* such as OpenAI's GPT-4 model (Generative Pre-Trained Transformer). Due to their size, models such as GPT-4 are also called *large language models* (LLMs). The origins of recent LLMs in the transformer architecture for NMT systems is depicted in figure 1:

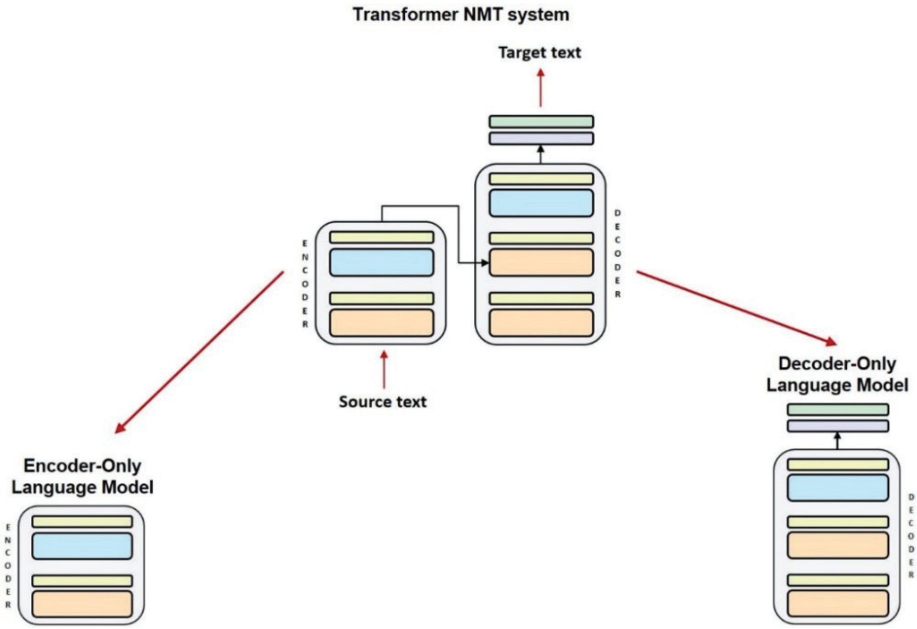

Figure 1: Origins of current LLMs in the transformer architecture for NMT systems

After being trained on massive amounts of data, LLMs exhibit an *in-context learning* behaviour (cf. Dong et al., 2023), which means that they can be conditioned 'on the fly' to perform a wide variety of different tasks via natural language prompting. For example, while dedicated MT systems such as DeepL can only perform machine translation, LLMs such as GPT-4 can be prompted both for machine translation and for a wide range of other tasks, such as autonomous text production, text optimisation or quality evaluation. Recent LLMs such as the current version of GPT-4 or Google's Gemini 1.5 are so-called *multimodal language models*, which can process other modalities besides written language (sound, images, videos). Particularly the ability of recent LLMs to produce autonomous texts and to process spoken language makes them applicable to other sectors of the language industry beyond translation, most notably monolingual specialised communication/technical writing and interpreting. Due to their high versatility, (multimodal) LLMs are also referred to as *general-purpose AI technologies*, which are defined as “machines designed to perform a wide range of intelligent tasks, think abstractly and adapt to new situations” (European Parliamentary Research Service, 2023, p. 1). These general-purpose technologies can potentially be used to further increase the degree of automation in a wide variety of language industry workflows. However, this requires a proper handling of these technologies along various dimensions (e.g., model interaction, workflow implementation, ethical considerations). In turn, this means that relevant language industry stakeholders will require an expanded set of digital competences in order to be able to harness the full potential of these technologies in an efficient and at the same time ethical and sustainable manner.

2. Digital literacies required in the digitalised and datafied language industry

The (semi-)automation of intellectual labour in the language industry through modern AI technologies is the combined product of processes of digitalisation and datafication. Digitalisation refers to the continuous development or evolution of digital technologies (most recently and notably in the form of powerful artificial neural networks) such as NMT systems or LLMs. Datafication, on the other hand, describes the process of accumulating and providing to relevant stakeholders large amounts of digital data (texts, images, videos, etc.) which can be used to train AI technologies in the ML paradigm. In the context of translation, which has been at the forefront of AI-induced automation via NMT, this has led to calls for adequate digital literacies on the part of the various stakeholders in the modern digitalised and datafied translation industry.

Three such digital literacies stand out in particular. The first one is *machine translation literacy*, which is defined by O’Brien and Ehrensberger-Dow (2020, p. 146) as “knowing how MT works, how it can be useful in a particular context, and what the implications are of using MT for specific communicative needs”. With a focus on the professional translation industry, Krüger (2022, p. 249) built on this concept and developed the concept of *professional MT literacy*, which describes “the full range of MT-related competences professional translators (and other language professionals) may require in order to participate successfully in the various phases of the MT-assisted professional translation process”. The second digital literacy recently propagated in the context of translation studies is *data literacy*. The concept is defined by Ridsdale et al. (2015, p. 11) as “the ability to collect, manage, evaluate, and apply data, in a critical manner”. The third and most recent digital literacy with high relevance in a translation/language industry context is *artificial intelligence literacy*, which Long and Magerko (2020, p. 1) define as “a set of competencies that enables individuals to critically evaluate AI technologies; communicate and collaborate effectively with AI; and use AI as a tool online, at home, and in the workplace". Given the pervasiveness of powerful AI technologies in modern societies, voices are emerging that posit AI literacy as one of most important literacies of the 21st century, together with traditional reading, writing, mathematical and overall digital skills (cf. Ng et al., 2021, p. 9). MT literacy, data literacy and AI literacy are not isolated concepts but rather interrelated in various ways, as shown in figure 2:

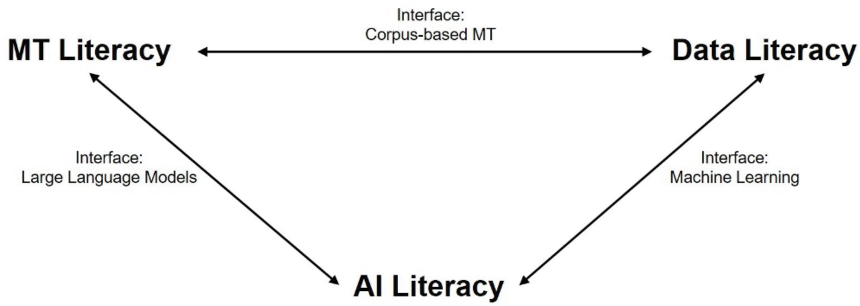

Figure 2: Interfaces between MT literacy, data literacy and AI literacy

The interface between MT literacy and data literacy is the paradigm of corpus-based MT. Contrary to systems from the earlier paradigm of rule-based MT, corpus-based MT systems do not operate on translation rules explicitly coded by humans. Instead, they are trained on large volumes of translation data (source texts and their translations) and derive their own translation rules from these training datasets. NMT is the most current variant of corpus-based MT, which makes data literacy an important component of contemporary MT literacy. This interface between MT literacy and data literacy formed the basis of the DataLit^{MT} research project (cf. DataLit^{MT}, 2023), which developed didactic resources for teaching data literacy in the context of professional MT literacy to students of translation studies/specialised communication programmes at BA and MA levels.

The interface between data literacy and AI literacy is the machine learning paradigm in AI research, which develops AI technologies that are able to acquire knowledge on their own by extracting patterns from training datasets. ML is thus the more general paradigm within overall AI research that informs the more specific paradigm of corpus-based MT. Modern high-performing AI technologies such as LLMs belong almost exclusively to the ML paradigm and are based on an inseparable combination of model algorithms (most notably the transformer) and their training data. Accordingly, Schüller et al. (2023, p. 426) argue that "data literacy and AI literacy cannot be separated from each other as data serves as the fuel for AI".

Finally, the interface between AI literacy and MT literacy is established by recent LLMs which, as discussed in section 1, emerged from the NMT transformer architecture. Given the origins of these LLMs in NMT, several subcomponents of MT literacy can be transferred more or less directly to the wider concept of AI literacy, as will be illustrated in the following section.

3. Outline of an Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication

In this section, I present an outline of an AI Literacy Framework for Translation, Interpreting and Specialised Communication. The framework is based primarily on three existing digital literacy frameworks: 1) The *Professional MT Literacy Framework* (cf. Krüger, 2022, p. 250) developed as part of the DataLit^{MT} project. Expanding upon the definition of professional MT literacy discussed in section 2, the framework distributes overall professional MT literacy over the five dimensions of *technical MT literacy*, *linguistic MT literacy*, *economic MT literacy*, *societal MT literacy*, and *cognitive MT literacy*. Each of these dimensions is divided further into individual subdimensions. 2) The *DataLit^{MT} Framework*, which is an MT-specific data literacy framework also developed in the context of DataLit^{MT} (cf. Krüger, 2022, p. 264). The framework covers the typical data lifecycle of an MT project and includes the five dimensions of *Data Context*, *Data Planning*, *Data Collection/Production*, *Data Evaluation*, and *Data Use* (again, divided further into individual subdimensions). 3) The AI literacy framework developed by Long and Magerko (2020), which is a generic framework structured along the five questions of *What is AI?*, *What can AI do?*, *How does AI work?*, *How should AI be used?*, and *How do people perceive AI?*[1] A reduced version of the proposed framework in its draft version is depicted in figure 3:

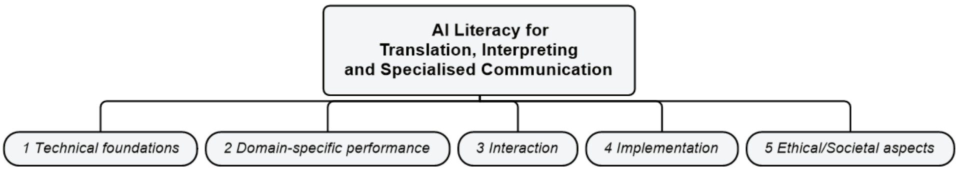

Figure 3: Outline of the Artificial Intelligence Literacy Framework for Translation, Interpreting and Specialised Communication (reduced version)[2]

[1] The Professional MT Literacy Framework and the DataLit^{MT} Framework as well as the interface between the two frameworks are discussed in more detail in Krüger (2022). Long and Magerko's AI literacy framework as well as its interface with the previous two frameworks are discussed in more detail in Krüger (2023).

[2] A digital version of this framework is available under the following link: th-koeln.de/itmk/ai-literacy.

The individual dimensions of the framework will be discussed in the following sections. Since the full framework is too extensive in scope to be elaborated here in full detail, the discussion will summarise briefly the respective dimensions and then focus only on selected sub-dimensions.

3.1. Technical foundations

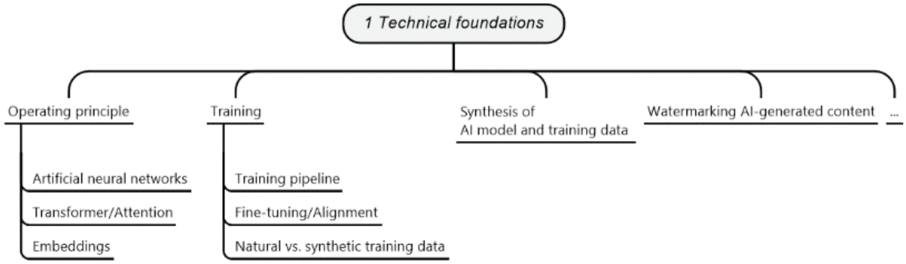

Figure 4: Dimension "Technical foundations"

The first dimension of the proposed framework is concerned with the technical basics of modern AI technologies. This dimension illustrates that the framework only captures a snapshot of the highly dynamic AI landscape and may soon have to be updated. For example, while the transformer is still the state-of-the-art architecture underlying modern AI technologies (and is hence listed under *Operating principle*), competing architectures (e.g., state space models such as *Mamba*, cf. Gu & Dao, 2023) are emerging, which may replace or compete with the transformer as the leading AI architecture in the future. The subdimensions *Training* and *Synthesis of AI model and training data* establish a direct link between this AI literacy dimension and data literacy (see section 2). For example, the data lifecycle of a typical MT project depicted in the DataLit^{MT} Framework basically covers the typical training pipeline of modern AI technologies such as LLMs. The aspect of *Natural vs. synthetic training data* covers a pressing topic in current AI research, namely the tendency to use synthetic (i.e., machine-generated) data to satisfy the extensive training data requirements of these systems, which may negatively affect system performance. For example, Shumailov et al. (2023, p. 1) show that relying extensively on synthetic data in AI model training (at the expense of natural, human-produced data) can lead to what the authors call "model collapse". In a similar vein, Alemohammad et al. (2023, p. 1) find that, "without enough fresh real data […], future generative models are doomed to have their quality (precision) or diversity (recall) progressively decrease"[3]. Watermarking

[3] This technical aspect of modern AI technologies is linked to the aspect of identifying the human added-value vis-à-vis these technologies (see section 3.2). In this context, Shumailov et

AI-generated content is also becoming more and more important in an era where AI technologies can imitate human written and spoken language at a very high level and can produce photorealistic images and videos, which drastically increases the risk of AI-induced manipulation (see section 3.5 concerned with ethical/societal aspects of AI). For example, LLMs could potentially be misused in language industry project management by having them mimic human project managers and using them to manipulate freelance translators, interpreters or technical writers to accept unprofitable jobs, unreasonable deadlines, etc.

3.2. Domain-specific performance

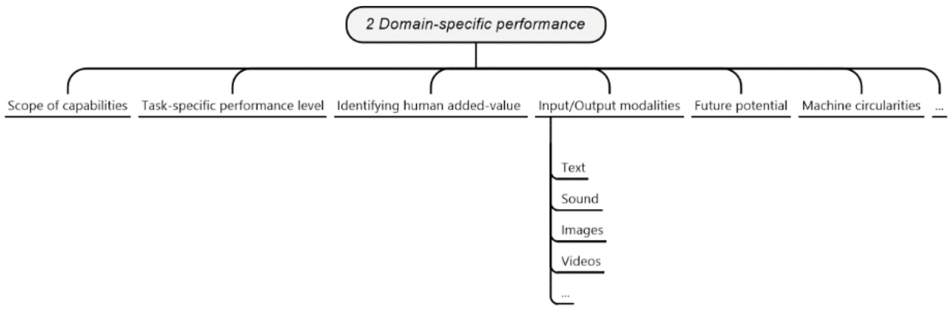

Figure 5: Dimension “Domain-specific performance”

The second dimension of the proposed framework covers the domain-specific performance of current AI technologies such as LLMs. Determining this performance is not a trivial task, since these general-purpose technologies do not betray their affordances in a straightforward way. This means that these systems, contrary to narrow expert systems such as dedicated MT systems (DeepL, etc.), do not readily ‘tell’ their users what to do with them because they can potentially be used for a vast variety of different tasks. Therefore, in order for relevant stakeholders to be able to determine the actual scope of capabilities of current LLMs, to measure their task-specific performance level (which also includes knowledge about the range of input/output modalities these models can handle) and to be able to articulate the added value that humans still provide in AI-fuelled language industry processes, these stakeholders require an adequate AI literacy. Determining this domain-specific performance of current AI technologies is also a prerequisite for integrating these technologies into actual professional workflows (see section 3.4). Given the high pace of current AI development, such an AI literacy also involves the ability to make informed speculations about the future

al. (2023, p. 1) point out that “the value of data collected about genuine human interactions with systems will be increasingly valuable in the presence of content generated by LLMs […]”.

potential of these technologies[4]. The high versatility of general-purpose LLMs may also pose a risk of introducing machine circularities into language industry production processes, e.g., when an LLM such as GPT-4 is asked to pre-edit a text for MT, to then machine translate this text and to also post-edit this text with the aim of optimising its quality. In order to avoid such machine circularities, process chains such as these – even though they could now be handled by a single AI model – should ideally be distributed over different technologies and/or human experts.

3.3. Interaction

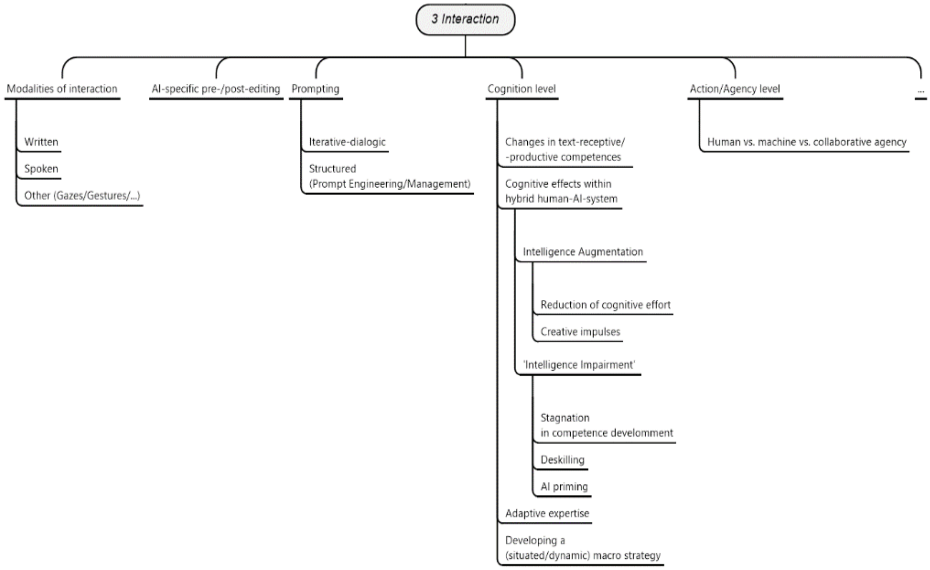

Figure 6: Dimension "Interaction"

This dimension covers aspects of human-AI interaction and is, perhaps unsurprisingly, the most extensive dimension of the proposed framework. The available modalities of interaction are related to the input/output modalities covered in section 3.2. For the near future, standard interaction modalities will most probably be written and spoken language, but other modalities, such as gesture interaction, are already being explored (cf., e.g., the work by Herbig et al., 2019 on multi-modal post-editing). The notion of AI-specific pre/post-editing is informed by MT pre-/post-editing but is wider in scope. For example, notes by design engineers could be structured/optimised by human pre-editors and then be

[4] Which also has an ethical/societal dimension, see the *Impact assessment* subdimension in section 3.5.

fed into an LLM which then produces an operating manual based on these notes (which would then have to be checked by human post-editors). The cognitive dimension of human-AI interaction is very important and therefore features prominently in the proposed framework. Within a hybrid human-AI system, both positive and negative cognitive effects can emerge. Positive cognitive effects can be subsumed under the term *intelligence augmentation*, which "focuses on AI's assistive role, emphasizing the fact that cognitive technology is designed to enhance human intelligence rather than simply replacing it" (Szczerbicki & Nguyen, 2021, p. 381). Examples of such intelligence amplification effects would be a reduction in cognitive effort involved in a particular task or creative impulses provided by the AI system. Negative effects could be subsumed under a neologism such as *intelligence impairment* and would include an AI-induced stagnation in competence development (e.g. a stagnation in translation competence in translation students under the influence of NMT systems), an AI-induced loss of competences (*deskilling*, e.g., professional translators losing the ability to translate from scratch because of the permanent availability of MT) or AI priming, i.e., "the cognitive residue that a task performed with technology has on the human mind" (Markauskaite et al., 2022, p. 6). Modern AI technologies also raise new questions concerning the relationship between human and machine agency and the potential merging of these two forms of agency in human-AI interaction. For example, van Lier (2023, p. 80) conceptualises LLMs and humans as the two components of a collaborative agent system. In such a system, LLMs remain – at least for now – the non-autonomous part, which is under human expert supervision.

3.4. Implementation

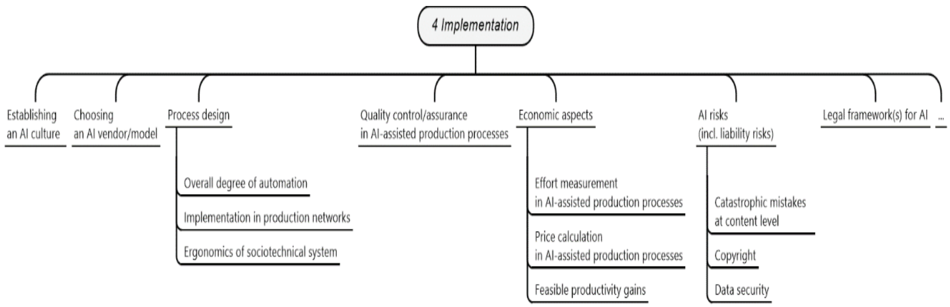

Figure 7: Dimension "Implementation"

This dimension is concerned with the implementation of AI technologies in language industry workflows and is heavily informed by the DataLit^{MT} Framework and the Professional MT Literacy Framework. Establishing an AI culture involves identifying and specifying areas of application where particular tasks could be solved using AI technologies, and establishing guidelines for using these technologies in an ethical and safe manner (see *Establishing a data culture* as part of the *Data Context* in Krüger, 2022, p. 265). This aspect is of particular importance in the language industry and other professional sectors since a recent survey by Salesforce (2023) among employees of international companies found that over half of the survey participants working with generative AI did so without consent from their employer and 7 in 10 participants had never received any training on how to properly use generative AI in the workplace. Process design involves establishing desirable and feasible degrees of automation and implementing AI technologies in production networks[5]. Here, overall sociotechnical considerations and aspects of organisational and cognitive ergonomics have to be taken into consideration. These aspects have been researched extensively in translation studies (see e.g., Ehrensberger-Dow & Massey, 2017) and can also inform process design in production networks fuelled by new AI technologies such as LLMs. The economic and risk dimensions of the proposed framework are derived from the subdimensions of *Effort estimation/measurement in MTPE*, *Price calculation in MTPE*, *Feasible productivity gains in MTPE*, and *Potential business risks of MT* as part of the *Economic MT Literacy* dimension of the Professional MT Literacy Framework. Again, these subdimensions focus on the more narrow use case of MT but can be extrapolated more or less directly to a wider range of use cases involving general-purpose LLMs. A major legal framework governing the future use of AI technologies is the European Union's AI Act (cf. European Parliament, 2023). The AI Act adopts a risk-based approach to AI technologies, which may affect AI implementation in the language industry and other professional sectors.

3.5. Ethical/Societal aspects

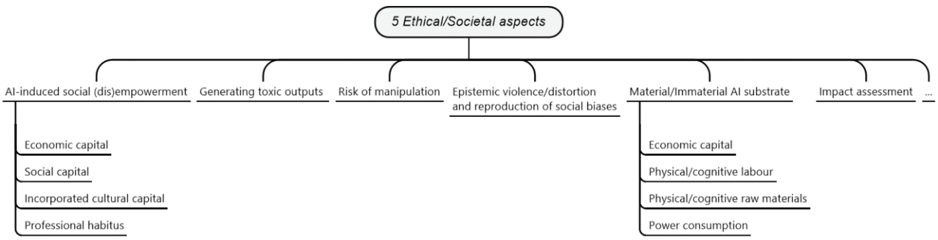

Figure 8: Dimension "Ethical/societal aspects"

[5] While avoiding machine circularities as discussed in section 3.2.

The final dimension of the proposed framework is concerned with ethical aspects of modern AI technologies, which transcend the language industry and other professional sectors and are highly relevant for AI-saturated societies in general (cf., e.g., Crawford, 2021). Again, translation studies has already brought forth a considerable body of work on this topic, which focuses mostly on the ethical/ societal dimension of NMT (cf, e.g., Moniz & Parra Escartín, 2023). One important aspect is potential AI-induced social (dis)empowerment of people affected by these technologies, which the framework models along the Bourdieusian dimensions of *capital* and *habitus* (cf., e.g. the Bourdieusian analysis of MTPE by Sakamoto, 2019). Other relevant aspects are the misuse of LLMs for generating toxic outputs via *jailbreak prompting* (cf. Yong et al., 2024) or the risk of manipulation associated with modern AI technologies (cf. the brief discussion in section 3.1). AI-induced epistemic violence/distortion refers to the potential misrepresentation of reality by data-driven AI systems, for example, by amplifying stereotypes in their underlying training data such as gender or age bias (in an MT context, cf. Bianchi et al., 2023). The notion of *Material/Immaterial AI substrate* involves an awareness of the potential exploitative nature and the environmental impact of AI, which requires large amounts of economic capital and physical/cognitive labour and raw materials and is at the same time a very energy-intensive technology (cf. Crawford, 2021). Finally, powerful AI technologies such as multimodal LLMs also require impact assessments, both at the level of individual industries as well as at overall societal level, in order to analyse the multifaceted consequences of these technologies along relevant dimensions (as sketched in the AI Literacy Framework). Given the high pace of development of current AI research, such assessments must include a forward-looking element, which could be informed, among other things, by ethical frameworks such as Brey's (2012) "anticipatory ethics for emerging technologies".

4. Conclusions

This paper presented an outline of an AI Literacy Framework for Translation, Interpreting and Specialised Communication. The next steps will be to finalise the framework (taking into account its inherent dynamicity and openness due to the high pace of current AI development) and to establish competence levels and competence descriptors for the individual (sub)dimensions of the framework. A blueprint of such competence levels could be Schüller et al.'s (2023, p. 429) three roles of 1) *informed prosumers* (people who produce and consume data and AI in an informed manner), 2) *skilled users* (people who use data and AI in a skilled and responsible manner), and 3) *expert creators* (people who create new insights, solutions, and tools using data and AI). Once the competence levels and descriptors of the framework have been established, they will form the basis for developing didactic resources (in the spirit of the DataLit^{MT} project) for developing an extensive set of AI-related competences required by current and future stakeholders in the AI-saturated language industry.

References

Alemohammad, S., Casco-Rodriguez, J., Luzi, L., Humayun, A. I., Babaei, H., LeJeune, D., Siahkoohi, A., & Baraniuk, R. G. (2023). Self-consuming generative models go MAD. *arXiv*. https://doi.org/10.48550/arXiv.2307.01850

Bianchi, F., Fornaciari, T., Hovy, D., & Nozza, D. (2023). Gender and age bias in commercial machine translation. In H. Moniz, & C. Parra Escartín (Eds.), *Towards responsible machine translation. Ethical and legal considerations in machine translation* (pp. 1–11). Springer. https://doi.org/10.1007/978-3-031-14689-3_9

Brey, P. A. E. (2012). Anticipatory ethics for emerging technologies. *NanoEthics, 6*, 1–13. http://doi.org/10.1007/s11569-012-0141-7

Crawford, K. (2021). *Atlas of AI. Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

DataLit^{MT} (2023). DataLit^{MT} project website. https://itmk.github.io/The-DataLitMT-Project/

Dong, Q., Li, L., Dai, D., Zheng, C., Wu, Z., Chang, B., Sun, X., Xu, J, Li, L., & Sui, Z. (2023). A survey on in-context learning. *arXiv*. https://doi.org/10.48550/arXiv.2301.00234

Ehrensberger-Dow, M., & Massey, G. (2017). Socio-technical issues in professional translation practice. *Translation Spaces, 6*(1), 104–121. https://doi.org/10.1075/ts.6.1.06ehr

ELIS Research (2023). European language industry survey 2023. https://elis-survey.org/

European Parliament (2023). EU AI Act: first regulation on artificial intelligence. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

European Parliamentary Research Service (2023). General-purpose artificial intelligence. *European Parliament*. https://www.europarl.europa.eu/RegData/etudes/ATAG/2023/745708/EPRS_ATA(2023)745708_EN.pdf

Gu, A., & Dao, T. (2023). Mamba: Linear time-sequence modelling with selective state spaces. *arXiv*. https://doi.org/10.48550/arXiv.2312.00752

Herbig, N., Pal, S., van Genabith, J., & Krüger, A. (2019a). Multi-modal approaches for post-editing machine translation. In S. Brewster, G. Fitzpatrick, A. Cox, & V. Kostakos (Eds.), *CHI '19: Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–11). Association for Computing Machinery. https://doi.org/10.1145/3290605.3300461

Krüger, R. (2022). Integrating professional machine translation literacy and data literacy. *Lebende Sprachen*, *67*(2), 247–282. https://doi.org/10.1515/les-2022-1022

Krüger, R. (2023). Artificial intelligence literacy for the language industry – with particular emphasis on recent large language models such as GPT-4. *Lebende Sprachen*, 68(2), 283–330. https://doi.org/10.1515/les-2023-0024

Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In R. Bernhaupt, F. Mueller, D. Verweij, & J. Andres (Eds.), *CHI '20: Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Association for Computing Machinery. https://dl.acm.org/doi/10.1145/3313831.3376727

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S. Tondeur, J., De Laat, M., Buckingham Shum, S., Gasevic, D., & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI? *Computers and Education: Artificial Intelligence, 3*, 1–16. https://doi.org/10.1016/j.caeai.2022.100056

Moniz, H., & Parra Escartín, C. (2023) (Eds.). *Towards responsible machine translation. Ethical and legal considerations in machine translation*. Springer. https://doi.org/10.1007/978-3-031-14689-3

Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence, 2*, 1–11. https://doi.org/10.1016/j.caeai.2021.100041

O'Brien, S., & Ehrensberger-Dow, M. (2020). MT literacy – a cognitive view. *Translation, Cognition & Behaviour, 3*(2), 145–164. https://doi.org/10.1075/tcb.00038.obr

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D., Kelley, D., Matwin, S., & Wuetherick, B. (2015). *Strategies and best practices for data literacy education. Knowledge synthesis report*. Dalhousie University. http://hdl.handle.net/10222/64578

Sakamoto, A. (2019). Why do many translators resist post-editing? A sociological analysis using Bourdieu's concepts. *Journal of Specialised Translation, 31*, 201–216. https://jostrans.soap2.ch/issue31/art_sakamoto.php

Salesforce (2023). More than half of generative AI adopters use unapproved tools at work. https://www.salesforce.com/news/stories/ai-at-work-research/

Schüller, K., Rampelt, F., Koch, H., & Schleiss, J. (2023). Better ready than just aware: Data and AI Literacy as an enabler for informed decision making in the data age. In M. Klein, D. Krupka, C. Winter, & V. Wohlgemuth (Eds.), *INFORMATIK 2023, Lecture Notes in Informatics (LNI)* (pp. 425–430). Gesellschaft für Informatik. https://doi.org/10.18420/inf2023_49

Shumailov, I., Shumaylov, Z., Zhao, Y., Gal, Y., Papernot, N., & Anderson, R. (2023). The curse of recursion: Training on generated data makes models forget. *arXiv*. https://doi.org/10.48550/arXiv.2305.17493

Szczerbicki, E., & Nguyen, N. T. (2021). Intelligence augmentation and amplification: Approaches, tools, and case studies. *Cybernetics and Systems, 53*(5), 381–383. https://doi.org/10.1080/01969722.2021.2018551

Van Lier, M. (2023). Understanding large language models through the lens of artificial agency. In H. Grahn, A. Borg, & M. Boldt (Eds.), *35th annual workshop of the Swedish Artificial Intelligence Society (SAIS 2023)* (pp. 79–84). https://doi.org/10.3384/ecp199008

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., & Polosukhin, Ł. (2017). Attention is all you need. In I. Guyon, U. von Luxburg, S. Bengio, H. M. Wallach, R. Fergus, S. V. N. Vishwanathan, & R. Garnett (Eds.), *Advances in neural information processing systems 30 (NIPS 2017)* (pp. 1–11). https://proceedings.neurips.cc/paper/2017/hash/3f5ee243547dee91fbd053c1c4a845aa-Abstract.html

Yong, Z.-X., Menghini, C., & Bach, S. H. (2024). Low-resource languages jailbreak GPT-4. *arXiv*. https://doi.org/10.48550/arXiv.2310.02446