



Castledown

 OPEN ACCESS

# Technology in Language Teaching & Learning

ISSN 2652-1687

<https://www.castledown.com/journals/tltl/>

*Technology in Language Teaching & Learning*, 6(1), 1–16 (2024)

<https://doi.org/10.29140/tltl.v6n1.1192>

## “Just a Pocket Knife, Not a Machete”: Large Language Models in TEFL Teacher Education and Digital Text Sovereignty



JULES BUENDGENS-KOSTEN 

*Goethe University Frankfurt, Germany*  
*buendgens-kosten@em.uni-frankfurt.de*

This paper approaches AI in TEFL teacher education from a perspective of digital text sovereignty (*digitale Textsouveränität*). Digital sovereignty (*digitale Souveränität*) is a concept that goes beyond media literacy and data literacy as a set of skills, to include *personal* competences in a more Humboldtian vision of education. Digital *text* sovereignty focuses on all those aspects of digital sovereignty that apply to the reception and production of texts, from the ability to produce and consume digital texts to the development of a critically-reflexive attitude to these texts (Frederking, 2022; Frederking & Krommer, 2022), making it especially useful in contexts of TEFL teacher education.

This paper is based on an empirical study conducted within the EFL teacher education department of a German metropolitan university during the 2023 summer term. It analyzes 21 student essays written after an interactive 90-minute training session on ChatGPT, conducted in three intact TEFL teacher education seminars.

Qualitative content analysis is used to identify domains of digital text sovereignty discussed in these essays, and to identify key themes related to four aspects of digital text sovereignty: Mediality, Source Code, Intentionality, and Veracity. The findings might be of interest for researchers and teacher educators interested in (a) modelling digital text sovereignty as it applies to AI, and (b) developing teacher education material and programs that target digital text sovereignty in an AI context, or that aim to support (future) teachers in developing an AI-informed pedagogy.

### Large Language Models (LLMs) in TEFL and TEFL Teacher Education

For several years, TEFL and TEFL teacher education have seen recurring debates on the impact of AI on these fields. Grünewald, for example, asked in 2019 if foreign language education was still needed in an age of AI translation tools. In CALL specifically, different AI-based tools such as machine

---

**Copyright:** © 2024 Jules Buendgens-Kosten. This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. **Data Availability Statement:** All relevant data are within this paper.

translation tools (Lee, 2020; Lee, 2021) or AI chatbots (Jeon, 2022; Yang et al., 2022) have been of ongoing interest. With the introduction of ChatGPT, these discussions have developed more urgency and have brought an increasing academic interest as well as a growing interest on the part of teachers and teacher education students to learn more and to develop their teaching and assessment practices in response to it.

Even though a technical description of large language models (LLMs) and their creation process via Machine Learning is not the focus of this paper, some features of LLMs in general and ChatGPT specifically will be outlined very briefly. ChatGPT – developed by OpenAI (<https://openai.com/>), originally published in 2022 and currently available in version 4 – is a pre-trained large language model, essentially “a system for haphazardly stitching together sequences of linguistic forms it has observed in its vast training data, according to probabilistic information about how they combine, but without any reference to meaning” (Bender et al., 2021, p. 617). Importantly, it allows access through a natural language interface. Users can formulate a prompt in ways similar to a teacher writing a prompt for learners, i.e. by specifying the content, text type, length, language, and style of a text, without being bound to a specific input format. Alternatively, users can also pose questions, which the LLM will then answer. A text can be the result of multiple ‘turns’ between the user and the tool, and a broad range of languages can be used in both prompt and output (including generation of multilingual texts and translation prompts). The response is generated – anew for each prompt – by drawing on a large language model created through machine learning. ChatGPT has no understanding of the world, nor an understanding of the rules of language, beyond information on which words co-occur with which other words. It has been described, consequently, as a “stochastic parrot” (Bender et al., 2021).

From an educational perspective, easily accessible LLMs like ChatGPT have been both applauded and viewed critically. A range of scholarly papers in addition to blogs, websites, and other channels have published (often strongly practice-focused) texts describing the potential of ChatGPT and similar tools for the language classroom (Bonner et al., 2023; Kohnke et al., 2023; Koraishi, 2023; Mahlow, 2023; Vogt & Flindt, 2023), and for data analysis in language teaching research (Pack & Maloney, 2023). *Empirical* research on ChatGPT, on the other hand, is still scarce, though the available literature is expanding nearly by the day.<sup>1</sup>

One trend, observable in currently available work, might be referred to as a “pedagogy of the gaps”: an approach to teaching in an age of digitalization in general, and of easily accessible LLMs and other machine-learning and/or AI tools specifically, that focuses on what the machine cannot yet do in determining the methods of teaching and assessment as well as the goals pursued through said teaching and assessment. It is necessarily reactive, changing as the technology changes. This reactivity is most pronounced on the level of method, where even small changes in the technology might lead to the need to fundamentally revamp how teaching and assessment happen. At the level of educational goals (as discussed below), a certain reactive factor, if balanced out by other factors, might be desirable to ensure learners are indeed prepared for life in contemporary society. Therefore, I will only use the term “pedagogy of the gaps” in those cases in which goals are determined nearly entirely by technological advances, and refrain from using the term when technological advances are seen merely as an encouragement to reconceptualize skills and competencies more broadly.

A look at AI and foreign language education that is not purely reactive is provided by Berthele and Udry (2023), which focuses on AI more generally, but the arguments which are made can be applied to

---

<sup>1</sup> Work on LLMs in (language) education is still in its infancy. Consequently, I have made the choice to also include preprints in my literature review.

LLM contexts and ChatGPT specifically without difficulty. The authors look at educational standards for foreign language teaching in Germany and Switzerland and discuss which of the goals laid down in educational policy documents can be reached through tools and without learning. They identify reading and writing as areas that can easily be achieved without learning by using e.g., translation tools, and they identify spontaneous interactive speech, intercultural competences and metalinguistic skills as less ‘replaceable’. Instead of falling into a “pedagogy of the gaps”, they suggest an update of the CEFR to include tool support more systematically in the “can do” statements.

Mahlow (2023), looking at German as a first, second and foreign language, also comes to a balanced view on the degree to which AI tools can ‘replace’ the knowledge and skills of the user:

“By embracing the opportunities presented by language technology and incorporating it thoughtfully into language learning environments, we can enhance writing instruction and empower learners to become effective communicators in German and other languages.” (Mahlow, 2023, p. 190)

A theme that runs through a number of these papers is the need for critical skills in dealing with LLMs and other AI tools. Vogt & Flindt (2023), for example, focusing mostly on older AI tools but touching on ChatGPT as well, make the argument that, as students are already using AI tools, the EFL classroom has the potential to integrate reflection on these “in a way that is conducive to learning for a variety of foreign language competences.” (Vogt und Flindt, 2023, p. 182). I suggest that this expressed need for critical skills fits well into the “digital text sovereignty” model discussed below.

### LLMs and Digital Text Sovereignty

Frederking coined the term *digital text sovereignty* (*digitale Textsouveränität*)<sup>2</sup>, drawing on the already established concept of *digital sovereignty* (also: *data sovereignty*) (Friedrichsen & Bisa, 2016). *Digital sovereignty/Data sovereignty* refer to “meaningful control, ownership, and other claims in data” (Hummel et al., 2021), though beyond this core, the exact interpretation of the term and related terms differ substantially (e.g., whether it refers to something on an individual or on a group level, and whether it is a right or a skill/competency; Hummel et al., 2021). *Digital text sovereignty*, focusing on the individual and skill/competency aspects of the term<sup>3</sup>, adds to this the dimension of textuality. Frederking argues that the “digital world” can be understood as a “complex structured digital text World, a kind of digital meta-text consisting of an infinite number of individual digital texts” (Frederking, 2022, p. 3, translation by the author), “text” taken here in its broad sense, encompassing a plurality of codes and modes.

Frederking argues that digital text sovereignty has four aspects, two on the semiotic level (Mediality [*Medialität*]; Source Code [*Quellcode*]), and two on the semantic level (Intentionality [*Intentionalität*]; Veracity [*Wahrheitsgehalt*]). These four aspects can each be viewed from a functional-technological perspective and a personal-reflexive perspective, producing a total of eight dimensions. These eight dimensions, in turn, each contain a productive and a receptive element (Table 1).

<sup>2</sup> As I am citing work originally published in German and the translations I have chosen might not be the preferred translations of the author, I am including the German original when I first discuss a term. In all following discussions, though, I will limit myself to the English term.

<sup>3</sup> Frederking frames digital text sovereignty as a habitus, rather than as a (set of) skill(s) or competence(s), even though this does not appear as a strict delineation within this text, in which he also references skills as components of digital text sovereignty.

**Table 1** *Frederking’s Eight Dimensions of Digital Text Sovereignty (Frederking, 2022)*

semiotic	Mediality	functional-technological	receptive
		personal-reflexive	productive
	Source Code	functional-technological	receptive
		personal-reflexive	productive
semantic	Intentionality	functional-technological	receptive
		personal-reflexive	productive
	Veracity	functional-technological	receptive
		personal-reflexive	productive
		functional-technological	receptive
		personal-reflexive	productive

I argue that this model has relevance when viewing the topic of LLMs in EFL. Concerning Mediality, understood here as “ability to comprehend(ing) reception and production of literal, visual, auditive and/or audio-visual text or media elements, their multimodal or symmedial composite structures and their hypermedial, connective and communicative connections” (Frederking, 2022, p. 6f, translation by the author), the relationship is very clear. ChatGPT as a large language model generates textual output based on textual input. The ability to use it for text generation, and the ability to (critically) work with such generated texts, are important in all languages a person uses, including their foreign languages – maybe even more so in foreign languages, due to the potential of the tool to scaffold lower text production skills and to produce model texts for analysis and reproduction.

Beyond Mediality, though, the Source Code dimension is equally important. Frederking’s (2022) text focuses strongly on functional-technological and personal-reflexive skills/habitus related to markup languages such as HTML, relegating programming languages to non-language subjects while acknowledging the importance of algorithms (Frederking, 2022). In the case of LLMs, Source Code could be interpreted more broadly as an understanding of what this tool is and is not, can do and cannot do. Only with a solid foundation in this can semantic skills related to Intentionality and Veracity be properly developed. Such a solid understanding of LLMs can, for example, prevent the misconception that LLMs have an understanding of the world, and express this understanding of the world in language, rather than the correct understanding of LLMs as stringing words together in ways that reflect probabilities based on their training data.

Intentionality comes into play when benign and less benign reasons for utilizing LLMs and LLM-generated texts are discussed. Frederking understands Intentionality as “recognizing or transparent making of the explicit and implicit Intentionality of a text (...) in the act of reception and production” (Frederking, 2022, p. 7, translation by the author). Frederking, drawing on Eco (1992), distinguishes three types of Intentionality: *intentio operis* (text intention), *intentio auctoris* (author’s intention), and *intentio lectoris* (reader’s intention). He argues, again following Eco, that text intention is the one that is best deducible on the basis of textual evidence alone (Frederking, 2022, p. 14). In the case of LLM-generated texts, where it is impossible to attribute Intentionality to the LLM itself and it may be

difficult to attribute Intentionality to the (machine-generated) text, complex but fascinating questions are opened up. Finally, Veracity, understood here as “successful assessment of the truth content of digitally consumed or presented text statements” (Frederking, 2022, p. 7, translation by the author) ties into discourses of the reliability of LLM-generated texts and biases within these texts (cf. e.g., Bender et al., 2021).

Finally, regarding the distinction between functional-technological and personal-reflexive digital text sovereignty, “functional-technological” refers to a habitus of applied, competent, self-determined reception and production of digital texts. Personal-reflexive, on the other hand, is a habitus “that understands digital texts as phenomena which can be used for creative self-assurance, questions them within the horizon of one’s understanding of self and world, and (...) constructively-critically reflects their consequences” (Frederking, 2019; 2020, p. 7, translation by the author).

I argue that digital text sovereignty as a concept is especially suited to the context of TEFL teacher education, as its focus on textuality makes it especially relevant in language learning contexts – more so than other concepts, such as AI literacy (overviews AI literacy in teacher education & school: Casal-Otero et al., 2023; Olari & Romeike, 2021).

### LLMs and Digital Text Sovereignty in Teacher Education

The discussion above looked at LLMs such as ChatGPT and digital text sovereignty mostly from the perspective of learners. In the remainder of the text, I will focus on TEFL teacher education in the context of LLMs, using digital text sovereignty as my theoretical lens.

LLMs deserve to be discussed in the context of (TEFL) teacher education for two reasons. Firstly, to respond to the easy availability of LLMs – and not only with “pedagogy of the gaps” – (future) teachers need what might be framed as an “AI-informed pedagogy”. As part of this, in order to be able to foster digital text sovereignty in an age of AI, they also need to *possess* digital text sovereignty, including both the functional-technological and the personal-reflexive aspects.<sup>4</sup>

Mishra und Koehler’s TPACK model (Mishra, 2019; Mishra & Koehler, 2006) describes how Technological Knowledge is not enough. Only if Pedagogical Knowledge and Content Knowledge – and, importantly, the overlaps between these three main types of knowledge – are present can teachers successfully teach with media and/or in a context of media usage. Analogous to this model, I do not claim that digital text sovereignty alone leads to a successful AI-informed pedagogy and its implementation, but that without digital text sovereignty (a term that combines aspects of technological aspects and content knowledge, but goes beyond these with the inclusion of personal-reflexive aspects) such an AI-informed pedagogy is not meaningfully possible.

Secondly, teacher education in itself also uses texts and writing as methods. The writing of term papers is not merely an assessment practice, but also a route for skill and knowledge development. Teacher education students take notes, summarize research, brainstorm, engage in translation, etc. as

<sup>4</sup> It is worth mentioning that many challenges in education cannot be resolved only through competent actions by individual teachers. As Vogt and Flint (2023) stress, institutional support plays an important role in technology integration, and this will most likely impact teachers’ responses to AI as well. Though in this paper I use a purely “individual” view of digital text sovereignty, the arguments brought forward by Vogt and Flint remind us of the origin of the term in the notion of digital/data sovereignty, and how that term is frequently used to denote features of communities (such as the state), or rights, rather than the skills of individuals. Such a broader view of digital text sovereignty, even if not applied in this paper, has some promise.



part of their teacher education. Tools that can support them in that – or that can, if used injudiciously, take away from them the opportunity for skill development – therefore deserve discussion in teacher education.

Yet, supporting TEFL teacher education students in developing their digital text sovereignty is not without its challenges. Vogt and Flindt (2023), for example, stress that one key element in preparing teachers for using and/or reflecting on AI tools in the EFL classroom is the affective dimension. Furthermore, in drawing on experiences with teacher education in the context of digitalization more generally, they stress the importance of institutional support.

Another challenge in preparing teachers for teaching in an age of generative AI and of LLMs is that teacher educators, too, require the necessary skills. Moorhouse & Kohnke (pre-publication) looked specifically at English language teacher educators in Hong Kong. The 13 EL teacher educators interviewed (in April and May 2023) indicate a high level of needing to learn more about this topic, and a perceived urgency to adapt to a world with generative AI tools.

## **Research Question**

This paper attempts to contribute to what is at the moment a small number of empirical studies focusing on TEFL teacher education for an age of AI, including easily accessible LLMs. It attempts to answer the following question: How do TEFL teacher education students address the different dimensions of digital text sovereignty related to AI and LLMs, and to ChatGPT specifically, after participating in a 90-minute interactive intervention?

It is hoped that answering this question may play a role for (a) modelling digital text sovereignty as it applies to AI, and (b) developing teacher education material and programs that target digital text sovereignty in an AI context, or that aim to support (future) teachers in developing an AI-informed pedagogy.

## **Method**

### **Participants**

Students from three intact university TEFL teacher education seminars at a German metropolitan university participated (Table 2). All lecturers (excepting adjunct-equivalent teachers) within the department were invited to participate with their courses. Four lecturers volunteered. One course was excluded, as the specific format of the course and the types of writing in which students engaged in the course would not have harmonized with the intervention planned.

The remaining three courses were all part of a four-year degree program, with two courses targeting students in the first two years of study (roughly BA-level equivalent), and one course targeting more advanced students (roughly MA-level equivalent). All courses were open to a range of TEFL teacher education specializations, which could include TEFL for primary school, TEFL for different types of secondary school, TEFL for vocational school, or TEFL with a focus on special education. The exact composition of courses varied.

Based on the courses students were in, they must have previously completed one or two TEFL didactics modules, and some students would also have already completed their teaching practicum. No demographic data on participants was collected.

**Table 2** *Overview Over Sample*

Seminar	Seminar G1	Seminar G2	Seminar G3	Total
Academic level	BA-level equivalent	BA-level equivalent	MA-level equivalent	
Academic focus	Digital Media in Teaching	Teaching Speaking Skills	CLIL & Multilingualism	
Participants who provided informed consent for use of their essays	7	9	8	24
Total number of essays retained in the study	7	6	8	21

## Materials and Procedure

The core of this study is a 90-minute intervention that could be integrated flexibly into TEFL teacher education seminars. The intervention was a “learning experience(s) focused on understanding AI”, (Casal-Otero et al., 2023, p. 6) in the form of an interactive, discussion-based format. It consisted of (a) an introduction to LLMs, (b) a group discussion of the purpose of writing in TEFL and in TEFL teacher education, (c) individualized and flexible work in a “work stations”-based mode, and (d) an end-of-session reflection in which students shared what they had learned. A follow-up assignment was set as homework.

For the individualized and flexible work in phase (c), students received a worksheet that included three tasks. Within the classroom, three stations had been prepared that students could access at their own pace. Students worked through the three tasks in any order they wished, at the speed they wished, and while working with any partner(s) they wished. The author and the lecturer of the course were present, answering questions and engaging in discussions with students to encourage them to justify their opinions, challenge their assumptions, or consider a question from multiple perspectives. The three tasks used can be found in the Appendix. They are also discussed in more detail in Buendgens-Kosten (forthcoming), which documents the intervention’s pilot during the 2022/2023 winter term, with a focus on multiliteracies.

The follow-up assignment was identical for the three courses:

*Write ~500 words on what ChatGPT and similar tools mean – or will mean – for EFL teaching and learning, including assessment. In your text, draw on the activities done today. Also draw on additional resources (formal or informal) that focus either on:*

- *How to use ChatGPT as a tool to support your writing/thinking/planning...*
- *The ethics of tools like ChatGPT (using existing texts to build a language model, low-paid workers, cheating, risks for society, bias in AI, liberating individuals from mind-numbing busywork...).*

*Please indicate all sources following APA 7<sup>th</sup> edition.  
Upload your reflection to your course LMS area.*

The intervention took place in a regular seminar session and supported learning goals of that seminar; consequently, participation in classroom activities associated with the intervention was mandatory. Only data from students who provided informed consent was used. Students who did not wish to provide

informed consent had no negative ramifications, and could enjoy the same educational experience as students who provided informed consent.

For the purposes of this paper, only the 24 essays created as follow-up assignments will be considered. A detailed discussion of findings related to the worksheets within the pilot study can be found in Buendgens-Kosten (forthcoming).

### **Data Analysis**

The author and/or the lecturer in the respective class anonymized the essays. As the essays were written at home, and strong AI usage standards were not yet in place and non-use of AI could not be effectively monitored, the need to establish whether texts were indeed written by students arises. Consequently, a minimal standard was applied to essays to improve the probability that the texts reflected actual student ideas. All essays had to reference the course activities and reference additional published material, or, in the absence of references to published material, required especially clear references to course materials. This minimal standard was chosen as references to the course material were clear evidence that learners were actively engaged in producing the text, if not in writing the whole texts themselves, then in designing prompts that included descriptions of course activities or specific ideas they wanted to express. References to published material were one of the weaknesses of early ChatGPT. At the point in time that data was collected, the presence of correct references therefore implied traditional writing or at least editing of a generated text. This procedure led to the exclusion of G2\_2, G2\_6 and G2\_7. Text G2\_6 was clearly AI generated: it made no reference to in-class activities and referenced several nonexistent (‘hallucinated’) texts, a tell-tale sign of early ChatGPT. G2\_2 and G2\_7 are more complex cases that might reflect disregard of the task instructions more than actual AI use, especially as their style was unlike the ‘overly-polished’ style frequently found in ChatGPT-generated essays. Still, these texts were excluded out of an overabundance of caution.

The remaining essays were analyzed using top-down qualitative content analysis. The author identified sections in each that reflect ideas or arguments related to digital text sovereignty, based on Frederking’s (2022) eight dimensions, focusing on Mediality, source text, Intentionality and Veracity. For borderline cases, it was decided to include them rather than to exclude them, to provide the broadest picture possible. When statements could fit into more than one category, they were coded for both categories. In a second step, the author identified key topics within each category.

## **Results and Discussion**

### **Mediality**

Every student (21 out of 21) addressed questions related to Mediality. Primarily, these were descriptions of what the tool can do, plus discussions of how this is done and what weaknesses might exist. Descriptions of use cases were very varied, and in general realistic, encompassing ways that both teachers and learners might use the tool, focusing on text generation and input – though one student mentioned multimodality and one other student mentioned audio input (G1\_1, G3\_4). Options for multilingual and mediation/translation use were mentioned by several students. Descriptions of “how to” tended not to focus on technological aspects, but rather on integration in e.g. the writing process.

A recurrent theme was that ChatGPT is good at some tasks and not so good at others, with several students pointing out that outputs based on assumedly simple questions could be of a surprisingly low



quality. Similarly, many students commented that reflections and other text types that require drawing on personal experience and emotions were often of lower quality:

“Chat GPT is a suitable tool for simple questions, but it is useless for critical reflection on personal experiences. (...) During the daily classes, students may use ChatGPT here or there, but if it is a game-changer for students also depends on the way of teaching. Since we are taught the dominance of TBLL/TBLT (...), ChatGPT is just a pocket knife, not a machete.” (G1\_7)

There seems to be a focus on productive use, though the distinction might not be very helpful in this context, as productive use by the learner might imply receptive use by the teacher, and vice versa. The codes collected for “Mediality” seem to fall primarily under “functional technological”, but this may be attributable to the analytical procedure chosen here, as “question(s)[ing] them within the horizon of one’s understanding of self and world, and (...) constructively-critically reflect(s)[ing] their consequences” (Frederking, 2022, p. 7, translation by the author) touched on Intentionality (consequences of academic dishonesty) and Veracity (bias). Some statements clearly touch on the personal-reflexive, though, such as this one:

“An example could be that students can ask ChatGPT for topics for an essay or a presentation. However, it could also lead to the students being less creative (...)” (G1\_1)

Another example is this comment, which touches on potential impacts of AI on society more broadly:

“A page that could create paintings or designs in seconds is thus able to do without illustrators/cartoonists. Or to people using a writing AI and selling books on renowned platforms such as Amazon (Klotz, 2023).” (G3\_1)

### Source Code

Most students (15 out of 21) touched upon Source Code, but often in a very superficial way.

At the most basic level, a technical-functional understanding of what ChatGPT is involves identifying it as an AI tool, more specifically an LLM. Both concepts were introduced in the intervention in a brief, non-technical way. Two thirds of the texts (14 out of 21) explicitly mention AI/Artificial Intelligence (or the German equivalent, KI). Usually little additional commentary is provided, making it difficult to judge the degree to which this concept is understood. G3\_5 is an exception here:

“In the modern era of technology, artificial intelligence (AI), which refers to the replication of human intelligence processes by machines, particularly computer systems (Burns et al., 2023) has emerged as a pivotal subject in discussions concerning technological advancement.” (G3\_5)

Contrary to G3\_5, G3\_1’s response seems to conflate ChatGPT with AI tools more generally:

“In the past year, humans have been introduced to artificial intelligence, aka AI.”

Four students also explicitly refer to (large) language models (LLMs), though most of them (three) either prefer to use the more general term “language model”, or misremembered the term as used in class. Again, barely any additional commentary was provided. The fact that so few used the term

“(large) language model”, suggests that this was a new concept for students attending the intervention – unlike AI, which students were already familiar with, at least at a superficial level.

Interestingly, two students (G2\_9, G3\_1) refer to bots, a term not introduced in the intervention, which is however not technically incorrect:

“ChatGPT is an artificial intelligence-based chat-bot that is available online and can generate written cohesive responses to the requests and inputs of the user.” (G2\_9)

G1\_3 goes in a similar direction while stressing the natural language processing skills of ChatGPT, referring to it as a “natural language processing tool” and to “natural language queries”, technically correct terminology that was, again, not used/not stressed in the intervention. It shows that in addition to the intervention, students drew on prior knowledge, on interactions with fellow students, or on readings.

I will not reiterate what students wrote about the capabilities and limitations of ChatGPT, as this overlaps with the relevant section on Mediality. Some of the statements of what ChatGPT can or cannot (assumedly) do, though, reflect misconceptions of how it works, which I will touch upon now.

G1\_3 demonstrates a minor misconception by writing of watermark technology as if it were already widely implemented. Some students made comments the veracity of which is hard to judge, as it depends on precise ways in which the LLM is trained and user input data is stored and used. This applies primarily to G1\_4, G1\_5, G1\_6, G2\_3, and G3\_2. The example of G2\_3 shows, though, how in practice their understanding would be sufficient for competent use of the tool:

“To be more specific, the language model is in some ways limited when it comes down to generate answers. The following should be considered, before using ChatGPT:

- 1) The language model is not up to date and can therefore provide altered information, studies etc. that are not entirely correct nowadays.
- 2) The language model has no access to all the sources on the internet and especially academic sources. Therefore, it is difficult to know whether the provided text/answer is very reliable, if it will not be checked.
- 3) ChatGPT does not cite (at least not always correctly). As the language model pulls its information out of other sources, users are likely to plagiarize.” (G2\_3)

Questions related to factual correctness of generated text will be taken up again below.

As I will also discuss under “Veracity” below, there is some degree of uncertainty as to whether students understood that ChatGPT is primarily a stochastic parrot (Bender et al., 2021) as well, i.e. that it generates texts based on probabilities derived from other texts, rather than based on world knowledge.

What may be most interesting is which misconceptions were *not* found in the data. No student attributed intelligence, creativity, sentience, etc. to the tool itself. It was clearly labelled as a tool, including by metaphorical references to physical tools. Students did not personify the tool, even in the contexts in which they referred to it as “conversation partner” (G1\_2). Students, in general, seemed to be aware that the tool generated text rather than just copying text – though expressions like in G2\_3’s text above might hint at some vagueness in their understanding in that regard. These correct assumptions are important in many ways, e.g. for understanding how the tool can be used for academic dishonesty, and

the options teachers do and do not have in identifying this through technical means – a topic that will be discussed in more detail under the header of “Intentionality”.

In general, if “productive use” is understood here as the ability to develop (large) language models or IT applications that build upon them, no evidence for productive skills/habitus can be found in the dataset. At the same time, students seemed to possess enough understanding – even if, possibly, not always technically quite correct – to approach generated texts with an understanding of their limitations and of the risks involved in the (injudicious) use of those texts.

### Intentionality

All but one text touched on the topic of Intentionality. By far the biggest topic in this section were comments related to academic dishonesty, cheating and plagiarism, focusing on learners as perpetrators. With the exception of two students (G2\_1 and G1\_1), all (i.e. 19 out of 21) student texts addressed one or more of these. As stated above, Frederking, drawing on Eco (1992), distinguishes three types of intention: *intentio operis* (text intention), *intentio auctoris* (author’s intention) and *intentio lectoris* (reader’s intention), arguing, again following Eco, that text intention is most clearly deducible on the basis of textual evidence alone (Frederking, 2022, p. 14). G3\_1, though, commented more on the author’s intention, framing academic dishonesty as a mismatch between teachers’ and learners’ intentions:

“At this point, how ethical is the use of artificial intelligence programs? It is indeed a tool that, if used correctly, helps to optimize working time. However, we have already seen that many students, for example, use these technologies for more free time. A college and a university should be an establishment where students express their thoughts and share them with others. What is happening so that there is no desire to express themselves?” (G3\_1)

This, in turn, is framed by several students not just as an issue for assessment, but as an issue for learning:

“Using Chat GPT to prepare written assignments compromises the success of learning this competence.” (G2\_9)

The consensus seems to be that (school and university) students can and do use ChatGPT in ways that do not reflect their writing skills and their own ideas. Many students commented on the *difficulty* in identifying if ChatGPT or similar tools had been used (receptive, technical-functional Intentionality). On the other hand, G3\_3 for example argues that identifying use of ChatGPT should be easy via comparison with a text sample:

“In the case of assessment, I noticed during the session with Dr. [author’s name] that ChatGPT writes too sophisticated and without personal style. Hence, the teacher could easily spot whether a text was written by ChatGPT or the students if he/she knew the style of the students. My idea would be to take a photo of a piece of writing that has been written in a lesson (perhaps at the beginning of the school year and then later in class) to have this as reference point when it comes to assessing a homework that should have been written by the students.” (G3\_3)

The kind of analytical skills described by G3\_3 perfectly reflects the digital text sovereignty of a teacher, who is seen as able to deduce intent of a text (or its author) through detailed textual analysis.

Importantly, none of the students seems to attribute intention to ChatGPT. Instead, intention was attributed to the users of ChatGPT, reflecting an understanding of what the tool is and is not (cf. section on Source Code).

Another aspect that could have been addressed here is bias, as bias can be understood as an (in this context) unintentional content feature of texts. I decided instead to discuss these text segments in the following section, which is concerned with Veracity.

### **Veracity**

Only 13 out of 21 essays were coded as touching upon Veracity.

There were two main themes in the data. Firstly, many students (7 out of 21) referred to ChatGPT and similar tools as sources of information. See for example G3\_8, referring to a paper they had selected themselves:

“If students require direct answers, they could use AI as a reliable option (AlAfnan et al., 2023, p. 65).” (G3\_8)

Or, even stronger, G3\_2, again drawing on a paper they had selected:

“Additionally, Biswas argues that ChatGPT could for example help find a solution for climate change. (...) ChatGPT can compress complex topics. This makes it easier to gain an overview and develop possible strategies. In order to prepare students for life as a responsible member of society, issues like climate change and social problems like poverty are discussed in school. Therefore, it makes sense to explain ChatGPT to students, as they will be confronted with the major problems of our society as responsible members of society. ChatGPT has the potential to help find a solution to these big, complex problems.” (G3\_2)

It would be wrong to conclude, though, that students viewed this uncritically. They provided many caveats, using expressions such as “not up to date” (G2\_3), “not (...) checked” (G2\_3), “outdated” (G1\_5), “wrong” (G1\_5), “faulty” (G1\_5), “shallow, unsatisfactory” (G1\_4) and “correct-sounding nonsense” (G1\_4). On the other hand, at least one student saw the risks related to Veracity as opportunity to develop students’ critical thinking skills (G2\_3).

G3\_3 argues that the Veracity of texts generated by ChatGPT reflects patterns of texts written by humans. Drawing on an article they had chosen and read, they argued:

“The correct usage includes educating students that ChatGPT answers their prompts with the data available to the system (Wiggenbröcker, 2023). This means that the amount of data determines the answers (Wiggenbröcker, 2023). For instance, if more articles on the internet propose that the earth is flat than round, then ChatGPT will give the former as answer as it cannot think for itself (Wiggenbröcker, 2023). Especially in the age of misinformation and fake news this is a huge concern. Students need to be aware that information given by ChatGPT must be checked, as it is not always correct.” (G3\_3)

While there was a widespread awareness of risks related to incorrect information obtained through ChatGPT, the exact mechanisms of how this occurs (i.e. that ChatGPT is a large language model

that uses patterns in language to generate texts, not world knowledge) is not necessarily reflected in these responses.

Finally, four students address bias, prejudice and stigma. As G3\_6 describes it:

“Additionally, ChatGBT [sic] is fed with information developed from human thinking. Therefore, wrong human thinking for example racism is reproduced, which can be dangerous.” (G3\_6)

For Veracity specifically, drawing the line between functional-technical and personal-reflexive is difficult. The need to fact-check can be seen as a step within a writing or reading process (productive perspective), but as a reader (receptive perspective), the knowledge that ChatGPT texts cannot be trusted, even if the prompt author has only the best intentions, will teeter into the personal-reflexive.

### **Limitations**

Importantly, as essays were written after an intervention, this study does not document the knowledge, skills and habitus of ‘naïve’ students. Care needs to be taken before drawing conclusions about teacher education students or in-service teachers in general.

This study was conducted in three intact seminars and followed university seminar logic in that the essays analyzed were learning tasks for the students as much as they were sources for research data. The specific design of in-class activities and writing prompts will have impacted the topics addressed by students, and other activities and other writing prompts might lead to slightly different findings.

The approach chosen was not optimal for detailed diagnostic information on comprehension and skill in the area of Source Code. In future studies, pen-and-paper tests might contribute more precise information, albeit at the risk of disregarding the TEFL-specific focus.

### **Conclusion**

In this study, I analyzed students’ essays, written after a 90-minute interactive intervention, with a focus on digital text sovereignty. These essays touched on the four aspects of digital text sovereignty, Mediality, Source Code, Intentionality and Veracity, though with differing frequency and differing intensity. Applying the distinction between productive and receptive use and between technical-functional and personal-reflexive was attempted, but not always straightforward.

Several conclusions can be drawn, and several questions arise from the analysis. Firstly, while digital text sovereignty still appears a very valuable theoretical lens for AI in foreign language teaching and learning, in practice, several of the aspects overlapped. For example, Source Text and Veracity are intertwined when talking about generative text AI, as without an understanding of how the text is generated, judging its veracity is limited to fact-checking.

Secondly, there is a range of frequently reported misconceptions that students either never held or that were remedied by an interactive 90-minute intervention. More work on students’ knowledge and beliefs can establish which of these two interpretations is more likely. Both interpretations, though, are promising regarding supporting students in developing high levels of digital text sovereignty in the context of AI and LLMs.



Secondly, establishing how well students understood some aspect of Source Code was hard to establish through the means chosen here. At the same time, it is unclear if vagueness, for instance about the mechanism that causes ChatGPT to ‘hallucinate’ false information, impacts the ability to deal with the risks inherent in that. Or, phrased as a question: How much *technical* understanding of AI and LLMs do we need to engage competently with AI and LLMs in daily life and in academic settings?

Finally, it should be acknowledged that, while the teacher education students in this study suggested many ways to deal with perceived possibilities and limitations, including e.g., using ChatGPT to practice specific language and transversal skills, as of now, we have very little knowledge on how these actions impact school students’ language competence or digital text sovereignty. In other words, which implementations of AI in EFL teaching are indeed “conducive to learning for a variety of foreign language competences” (Vogt and Flindt, 2023, p. 182) is, as of yet, unclear.

## References

- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: can language models be too big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Berthele, R., & Udry, I. (2023). Welche fremdsprachlichen Lernziele der Schulen können mit digitalen Tools aber ohne Lernen erreicht werden? Simply playing the ostrich will not help in the long run. *Zeitschrift für Interkulturellen Fremdsprachenunterricht*, 28(April), 443–461.
- Bonner, E., Lege, R., & Frazier, E. (2023). Large language model-based artificial intelligence in the language classroom: Practical ideas for teaching. *Teaching English with Technology*, 2023(1). <https://doi.org/10.56297/BKAM1691/WIEO1749>
- Buendgens-Kosten, J. (forthcoming). “Like a shell without actual content”: Large language models as a topic in EFL teacher education. In C. Blume (Ed.), *Multiliteracies-aligned Teaching and Learning in Digitally-mediated Second Language Teacher Education*. Routledge.
- Casal-Otero, L., Catala, A., Fernández-Morante, C., Taboada, M., Cebreiro, B., & Barro, S. (2023). AI literacy in K-12: A systematic literature review. *International Journal of STEM Education*, 10(1), 29. <https://doi.org/10.1186/s40594-023-00418-7>
- Frederking, V. (2022). *Digitale Textsouveränität. Personale und funktionale textuelle Bildung im Deutschunterricht im Zeichen digitaler Transformation: Eine Theorieskizze: Version 3*. [deutschdidaktik.phil.fau.de/files/2021/09/digitale-textsouveraenitaet.pdf](https://deutschdidaktik.phil.fau.de/files/2021/09/digitale-textsouveraenitaet.pdf)
- Frederking, V., & Krommer, A. (2022). Sprachliche, literarische und mediale Bildung in der digitalen Welt. In V. Frederking & R. Romeike (Eds.), *Fachliche Bildung in der digitalen Welt: Digitalisierung, Big Data und KI im Forschungsfokus von 15 Fachdidaktiken. Allgemeine Fachdidaktik Band 3* (pp. 82–120). Waxmann.
- Friedrichsen, M., & Bisa, P.-J. (Eds.). (2016). *Digitale Souveränität: Vertrauen in der Netzwerkgesellschaft*. Springer VS. <https://doi.org/10.1007/978-3-658-07349-7>
- Grünwald, A. (2019). Digitaler Wandel: Warum überhaupt noch Fremdsprachen in der Schule lernen? In E. Burwitz-Melzer, C. Riemer, & L. Schmelter (Eds.), *Das Lehren und Lernen von Fremd- und Zweitsprachen im digitalen Wandel: Arbeitspapiere der 39. Frühjahrskonferenz zur Erforschung des Fremdsprachenunterrichts* (pp. 80–89). Narr.
- Hummel, P., Braun, M., Tretter, M., & Dabrock, P. (2021). Data sovereignty: A review. *Big Data & Society*, 8(1), 205395172098201. <https://doi.org/10.1177/2053951720982012>
- Jeon, J. (2022). Exploring AI chatbot affordances in the EFL classroom: Young learners’ experiences and perspectives. *Computer Assisted Language Learning*, 1–26. <https://doi.org/10.1080/09588221.2021.2021241>

- Kohnke, L., Moorhouse, B. L., & Zou, D. (2023). ChatGPT for Language Teaching and Learning. *RELC Journal*, 54(2), 537–550. <https://doi.org/10.1177/00336882231162868>
- Koraishi, O. (2023). Teaching English in the age of AI: Embracing ChatGPT to optimize EFL materials and assessment. *Language Education & Technology*, 3(1), 55–72.
- Lee, S.-M. (2020). The impact of using machine translation on EFL students' writing. *Computer Assisted Language Learning*, 33(3), 157–175. <https://doi.org/10.1080/09588221.2018.1553186>
- Lee, Y.-J. (2021). Still taboo? Using machine translation for low-level EFL writers. *ELT Journal*, 75(4), 432–441. <https://doi.org/10.1093/elt/ccab018>
- Mahlow, C. (2023). Large language models and artificial intelligence as tools for teaching and learning writing. *Osnabrücker Beiträge zur Sprachtheorie*, 101, 175–196. <https://doi.org/10.17192/OBST.2023.101.8607>
- Mishra, P. (2019). Considering contextual knowledge: The TPACK diagram gets an upgrade. *Journal of Digital Learning in Teacher Education*, 35(2), 76–78. <https://doi.org/10.1080/21532974.2019.1588611>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record: The Voice of Scholarship in Education*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Moorhouse, B., & Kohnke, L. (2023). *The Effects of Generative AI on Initial Language Teacher Education: The Perspectives of Teacher Educators* [Preprint]. SSRN. <https://doi.org/10.2139/ssrn.4532479>
- Olari, V., & Romeike, R. (2021). Addressing AI and data literacy in teacher education: A review of existing educational frameworks. *The 16th Workshop in Primary and Secondary Computing Education*, 1–2. <https://doi.org/10.1145/3481312.3481351>
- Pack, A., & Maloney, J. (2023). Using generative artificial intelligence for language education research: Insights from using OpenAI's ChatGPT. *TESOL Quarterly*, tesq.3253. <https://doi.org/10.1002/tesq.3253>
- Vogt, K., & Flindt, N. (2023). Artificial intelligence and the future of language teacher education: A critical review of the use of AI tools in the foreign language classroom. In P. Hohaus & J.-F. Heeren (Eds.), *The future of Teacher Education: Innovations across Pedagogies, Technologies and Societies* (pp. 179–199). Brill.
- Yang, H., Kim, H., Lee, J. H., & Shin, D. (2022). Implementation of an AI chatbot as an English conversation partner in EFL speaking classes. *ReCALL*, 34(3), 327–343. <https://doi.org/10.1017/S0958344022000039>

## Appendix

### Station A: Can ChatGPT Do This?

Find a partner. Together, pick a few different tasks cards. Sort them according to how ‘hard’ you think they’d be for ChatGPT. Together, discuss how – and why – you sorted them the way you did.

**Only after you have committed to one way of sorting them**, check the flip sides. Is this what you expected? Take notes.

### Station B: Compare Your Text to the ChatGPT Version

- a) Compare the text you have written with one or two of the texts generated by ChatGPT. Which strengths and weaknesses does your text have, which strengths and weaknesses does the ChatGPT text have?
- b) Exchange your observations (...). Make sure to also discuss the different prompts, and how they relate to the different texts.

### Station C: Can You Create “Cheat-Proof” Tasks?

Look through the cards. Would these ideas make writing tasks and writing assessments “AI-cheat-proof”? “AI-cheat-proof” means, in this context, that **taking these measures makes it more likely that the work products reflect students’ writing skills and/or content knowledge, and that students are more likely to develop their writing skills and/or content knowledge with these ideas in place.**

Step 1: Look at the suggestions provided. Put a **green sticker** on the ideas you find **helpful**, and a **black sticker** on the ideas you find **not helpful**. For this activity, focus on **the context of EFL at school**.

Step 2: Find a partner. Together, discuss at least three cards in more detail. What are advantages/disadvantages of this idea? Take notes.

Step 3: Which other measures could one take? Use the **yellow cards for ideas related to TEFL teacher education**, **blue cards for EFL learning at school**. Remember: Write in GREEN (on the cards) if you are ok with your ideas being used for research.