

Rethinking Reference and Authorship: On the Philosophical Status of LLM-Generated Verbal Products

Jan Georg Schneider

Abstract

In this article, the status of LLM-generated verbal products is discussed in principle. While we have so far been socialized in such a way that we automatically assume an intelligent author ‘behind’ verbal products that can be read as intelligent, we can no longer simply assume this close connection in the age of LLMs. In this sense, I name LLM-generated products ‘intelligible textures’. As products, these intelligible textures can hardly, if at all, be distinguished from authorized, human-created texts, but the learning and usage processes differ fundamentally, especially with respect to acts of reference, particularly acts of exemplification. What consequences does this have for our general conception of the written word, authorship and the ethical-political idea of responsibility for verbal products? This comprehensive question is discussed in the present article using the example of LLM-generated essay evaluations, which seem to be an instructive example here as they require a high degree of judgment and truthfulness.

Keywords: exemplification, judgment, essay evaluation, authorship, ChatGPT

Corresponding author:

Prof. Dr. Jan Georg Schneider
RPTU University of Kaiserslautern-Landau
Institut für Germanistik
Fortstraße 7
76829 Landau
j.schneider@rptu.de

1 Introduction

In this article,¹ general problems of reference and authorship are discussed using the example of automated essay evaluation. The argumentation follows on from two papers (Schneider/Zweig 2022 and 2023) in which we examined the U.S. “e-rater” (see Burstein et al. 2013a), a system patented in 2002 and regularly updated since then. From a philosophical, linguistic and computational perspective, we explained that such technical systems, even if they supposedly “rate” essays, of course, do nothing more than predict grades. Assuming that machines generally do not perform actions in the narrower, philosophically robust sense, the maximum claim would be to “substitute” a human action “in a near-equivalent way” (cf. Janich 2015: 302 et passim, Janich 2006: 161–165, Becker 2021: 19): Can automated grade prediction replace human grading?

To answer this question, we took a very fundamental look at what is needed for a successful grading procedure by revealing the felicity conditions for successful speech acts according to Austin: among other conditions the procedure must be carried out “correctly” and “completely” (Austin 1975: 36–38). In the case of essay evaluation, it is essential for the correct and complete procedure that, in addition to the grade awarded, a suitable justification can be provided (Schneider/Zweig 2023: 105–106). However, this is fundamentally not possible with the e-rater: in semiotic terms, the system only accesses “symptoms” (cf. Keller 2018: 155–168), but not quality criteria. It was programmed to count, among other things, the modal verbs and certain types of subordinate clauses used per essay. In addition, a machine training with 250 to 300 essays evaluated by humans takes place in advance with regard to the respective essay task, e.g. for Toefl tests (see also Burstein et al. 2013a: 61). With these two components, the e-rater can predict fairly accurately the grade that would be expected from ‘normal’ graders for ‘normal’ essays (see Rupp et al. 2019; Meyer et al. 2023). Thus, the system can be used as a tool to support essay evaluation in the sense of a corrective measure, as long as at least one human reviewer evaluates the essay (cf. Schneider/Zweig 2023: 106–107).

The crucial point, however, is the following: the symptoms identified by the e-rater could not reasonably be specified by humans as evaluation criteria; it would not be possible to convincingly demonstrate that an essay was given a positive evaluation, because it contained certain words and certain syntactical constructions.

1 I would like to thank the reviewers Sarah Brommer and Johannes Lenhard, as well as the co-editor of this volume, Nina Kalwa, for their inspiring and constructive comments.

Rather, criteria such as coherence, argumentative plausibility, truthfulness, originality, aesthetics are still required for good reasons,² and such appropriate criteria are also explicitly mentioned in the patent of the e-rater (Burstein et al. 2002: 1 and 5). However, the e-rater only accesses these criteria indirectly: it identifies superficial characteristics of texts that have been written by humans and have been or could be evaluated by humans according to criteria. It analyzes “cohesion without coherence” (Schneider/Zweig 2023: 106), which means that nonsense texts can receive top marks (cf. Perelman 2020, for background see also Burstein et al. 2013b). If only the e-rater is used for evaluation, justification is no longer part of the game. The justification, though, serves to stabilize the procedure for the future, as this is the only way to give orientation to teachers and learners, as well as to maintain the culturally anchored conditions for successful evaluation.

Our articles (Schneider/Zweig 2022 and 2023) were written before the big LLM hype began. In times of ChatGPT and co., the question arises whether such a chatbot could possibly substitute the missing textual grade justification or at least support the reviewer in formulating one (see also Schneider 2024). In the present article, I discuss this question, in order to move on to general questions of reference and authorship based on this paradigmatic field of application. To this end, I will first report on some tests I conducted with ChatGPT4o in March 2025 and describe tendencies: Can ChatGPT produce a suitable essay evaluation? Can it identify and categorize errors and inconsistencies in essays? Can it identify factual errors and misleading information? Especially the last question addresses the important quality criterion of truthfulness, and therefore problems of reference, predication and authorship become particularly virulent here. In order to address these problems from a pragmatic perspective, I return to Austin's speech act theory and ask whether the verbal products of LLM chatbots can replace rhetic acts (section 2). Since the products of ChatGPT often can hardly be distinguished from human texts, I then focus on processes of human language learning and language use in comparison to LLMs and explain how exemplification (see Elgin 1983) as a basic mode of referential acts plays an important role here and helps to assess the status of LLM-generated products philosophically (section 3). In section 4, I take an even broader perspective and ask about the consequences for our concept of authorship and our responsibility for linguistic products. In doing so, I introduce the term *intelligible texture* (see Schneider 2024) and explain its relevance.

² For evaluation criteria for essays from a language teaching perspective, see Rödel (2018: chapter 4.3).

2 Can ChatGPT justify an essay grade?

ChatGPT ('Generative Pre-trained Transformer') is an LLM chatbot that has been trained with huge text corpora to generate 'its own' texts based on statistical probabilities of word sequences (see Wolfram 2023). It can respond to a prompt and then generate a text similar to those used to train the neural network. A well-known problem here is the unreliability with regard to factual claims; above all, the lack of transparency concerning sources. Although the generation of completely untrue assertions has been significantly reduced by Reinforcement Fine-Tuning (RFT), the problem still exists. In the light of Austin's speech act theory (1975: 92–93), the problem of false or misleading claims particularly concerns the rhetic act, which consists of referring to something ("reference") and saying something about it ("sense"). Or, to put it in the terminology of linguistic pragmatics that is currently in use: reference and predication acts are performed. Since, on the one hand, it can be hypothetically assumed that LLM chatbots show significant weaknesses in this regard and, on the other hand, it is clear that reference and predication are of the highest relevance for questions of truth, authorship, and responsibility, I will focus in particular on the rhetic act in the following empirical exploration of ChatGPT.

In March 2025, I run a test with ChatGPT4o to see if it can be useful for grading an essay. In the first step, I send the following prompt to have ChatGPT create an essay text.³ The specific task that ChatGPT is supposed to solve is an "issue item" that was also used as an example in the e-rater patent (Burstein et al. 2002: 10):

You are a university applicant for a English-language study programme and are supposed to write an essay of 400–500 words on the following topic for the entrance examination: 'Everywhere, it seems, there are clear and positive signs that people are becoming more respectful of one another's differences.' In your opinion, how accurate is the view expressed above? Use reasons and examples from your own experience, observations, or reading to develop your position.

Then I start a new chat history and let ChatGPT evaluate the essay. This time I give the following prompt:

An applicant should write an essay of 400–500 words on the following task for an entrance examination for admission to a university in an English-language study programme: "Every-

³ The essay produced by ChatGPT4o can be found in appendix (A).

where, it seems, there are clear and positive signs that people are becoming more respectful of one another's differences.' In your opinion, how accurate is the view expressed above? Use reasons and examples from your own experience, observations, or reading to develop your position." The essay that the applicant has written is below. Please give it a mark between A and F (using the American system), you can also give marks such as B plus or C minus, and write a justification for this feedback. It should also include three tips on how the applicant can improve. This is the essay: [full essay that ChatGPT created before inserted]

ChatGPT4o grades the essay with A minus (see appendix B) – a really good grade for a well-structured standard essay written with a standard 'temperature'.⁴ The style of the evaluation appears professional and appropriate for the text type, even if it is reminiscent of a horoscope in places: it could be applied to almost any good standard essay on a factual topic.

My testing strategy now is to make specific changes to the essay in order to make it worse and then let the system evaluate it. I create two versions: one with linguistic/stylistic errors and logical inconsistencies (see appendix C) one with false and misleading claims (see appendix D).

2.1 Correcting linguistic and logical errors

First, ChatGPT should grade the version with linguistic and logical deficiencies. This time the grade is C minus and the rating correctly emphasizes that the text mainly contains linguistic, stylistic and logical errors. In two further prompts, I then ask the system to first name all logical inconsistencies, then all linguistic/stylistic errors and, in the second case, to categorize them according to error categories. The result is very similar to previous tests with German essays (see Schneider 2024):

- The **logical inconsistencies are identified very well**, for instance in this output: "Sentence: 'Additionally, corporations and institutions now actively promote homo-

⁴ When I ask for the temperature used, ChatGPT4o tells me: "The text above was produced with a moderate to low temperature setting, likely around 0.3 to 0.5 on a scale of 0 to 1." At a temperature of 0, the result would be formulated very dryly, since then only the statistically most probable word tokens appear next to each other. The 'creativity' would be much more pronounced at a higher temperature of 1, where less likely word tokens also appear. Thus, the term temperature refers to the respective grade of 'creativity' in ChatGPT products (see Wolfram 2023: 2).

geneity and exclusion, recognizing the importance of respecting differences in the workplace.’ Problem: The essay intends to argue that corporations promote inclusivity and diversity, but the phrase ‘homogeneity and exclusion’ contradicts this point. Homogeneity and exclusion imply the opposite of respect for differences.”⁵

- The system **identifies a large number of linguistic errors** as such, but is **unreliable in their categorization**: for example, the stylistic norm deviation *babbling on cultural diversity* is categorized as a ‘spelling mistake’ and the logical confusion of *fortunately* and *unfortunately* as a ‘grammatical error’; several grammatical congruence errors, on the other hand, are classified correctly.

2.2 Correcting factual errors – reference and predication

As mentioned above, the rhetic act with its sub-acts of reference and predication is of particular interest for this investigation. Unlike the phatic act, which is purely verbal, the rhetic act concerns the interrelationship between language and the world. Theoretically, this becomes even more complex when we realize that language and world are not separate, so that it would be more appropriate not to speak of a relationship between language and world, but rather to say: It is only with the rhetic act (see Austin 1975: 92) that the whole language game in Wittgenstein’s sense – with its verbal and non-verbal aspects – is grasped. Only in the rhetic act are full contexts, not just pure co-texts, involved. The performance of a phatic act consists of using words as words of a particular language with a particular vocabulary and grammar. The phatic act is a necessary, but not a sufficient condition for the simultaneously performed rhetic act, and only through the latter does meaning/semantics come into play (see Austin 1975: 92): To perform a rhetic act means to talk about something in the world and to say something about it in contexts with certain conventions and rules, i.e. in language games (see Wittgenstein 1984). The execution of a rhetic act requires judgment: the ability to decide in a concrete situation whether something falls under a particular concept or not. Thus, it is useful for our investigation to look closely at how ChatGPT deals with judgments or assertions about the world. Does the system find factual errors? How does it comment on such errors? What can we learn from this?

To test this, I spike our initial essay with misleading or untrue statements (see appendix D), for instance that “Belarus and Hungary

⁵ This a very typical example, the other inconsistencies are revealed and explicitly explained in the same clear way.

have passed anti-discrimination laws that protect individuals based on race, gender, religion, and sexual orientation”. The text edited in this way now contains a total of 19 factual errors and misleading information. Here, too, I first have it graded and evaluated in general terms. Once again, the grade is C minus, and despite the horoscope-like nature of the evaluation text, the concluding remark shows that the main weakness has been identified – namely factual deficiencies:

While the essay demonstrates an effort to engage with the topic, the factual inaccuracies, weak analysis, and structural issues significantly impact its effectiveness. Improving fact-checking, providing deeper insights, and refining organization will result in a stronger response.

The next prompt is: “Please list all factual errors in the text and explain exactly why they are errors!” ChatGPT’s response then identifies almost all of the factual flaws: For example, the statement about Belarus and Hungary already quoted above. And in the following statement full of falsehoods, all five errors are identified: “Fortunately, the German president Angela Merkel introduced marriage for all in 2017 with the help of Helmut Kohl, although the SPD and the Greens voted against it.” In summary, ChatGPT4o reports: Angela Merkel was not president, but chancellor. She put the ‘marriage for all’ proposal to the vote, but unlike the SPD and the Greens, she voted against it, and Helmut Kohl could not help her because he had already passed away.

The only factual errors that ChatGPT does not (immediately) identify correctly concern events that took place after June 2024, which is the date up to which the LLM’s training data extends at that time. The inserted claim that “legal frameworks, social movements, and education promote respect especially since the re-election of Donald Trump in december 2024” is commented on by ChatGPT4o as follows:

Why this is an error:

- As of now, Donald Trump has not been re-elected in December 2024. U.S. presidential elections take place in early November, not December. If Trump were re-elected in 2024, his victory would have been determined in November 2024, not December.
- There is also no clear connection between Trump’s hypothetical re-election and increased respect for differences. Trump’s previous term was marked by policies

and rhetoric that often heightened political and social divisions.

While the chatbot's report on the fixed election day and Trump's "previous term" is fact-based, the system does not 'know' about his re-election, which is marked as "hypothetical". When I ask the chatbot how long its training data extends, I receive the following correct information and the offer to carry out an internet search:

My general knowledge is up to June 2024, but I can fetch real-time information from the web if you need updates on current events, political developments, or recent changes. Let me know if you want me to verify anything specific!

In the next prompt, I ask whether Trump's re-election is actually 'hypothetical', as stated by ChatGPT. This is followed by an internet search, then the admission that the information was wrong, followed by a detailed, accurate correction with links to the online sources, in this case Reuters, CNN, Wikipedia. A quite impressive and potentially helpful fact-checking. The same procedure is used for the topic of the 2025 German federal election, in which the far-right AfD is supposed to have become the strongest party in my manipulated version. Here, too, the chatbot carries out an internet search with an accurate correction. As these observations show, it is only possible to obtain useful results if one already knows how far back the training data goes and then asks specific questions. The web-based source information, which is only provided on request, is a real step forward compared to previous versions of ChatGPT.

Additionally, I test whether ChatGPT4o can identify the difference between truly fake and rather misleading statements. Prompt: "Are there any claims in the article that are not really wrong but misleading? If so, please give the best example of this!" ChatGPT4o selects the statement I manipulated, according to which European countries are forcing people to use politically correct language. Here a part of the explanation the system provides me with:

- The phrase "European countries are forcing people to use politically correct language" is misleading because it implies a legal requirement that mandates individuals to speak in a certain way in all situations.
- In reality, while some European countries have laws against hate speech and discrimination, there is no general law that forces citizens to use "politically correct" language in personal conversations.

- Companies, schools, and media organizations might promote inclusive language policies, but this is not the same as a government forcing individuals to speak in a certain way.

In response to this, I ask whether the following sentence would not also be a good example: “In Germany, China, France and Russia, LGBTQ+ activists and women are demonstrating for their rights.” ChatGPT4o confirms this (“great example”) and provides, among other comments, the following explanation:

- The sentence groups together Germany, China, France, and Russia as if LGBTQ+ activists and women can demonstrate in all of these countries under similar conditions.
- This is misleading because Germany and France allow open demonstrations for LGBTQ+ and women’s rights, while China and Russia severely restrict or punish such protests.

Overall, the following can be stated with regard to factual errors and misleading claims: As the last few examples in particular show, the system sometimes achieves impressive things here. It detects many such flaws, but we can only use it as a tool if we already have the necessary knowledge, can evaluate statements and assess the quality of sources. We need to know when the training data was collected, need to dig deeper with appropriate prompts, and cannot rely on the system to detect misleading assertions. I repeated my prompts several times on different days in March 2025, and the tendency was always the same: some impressive outputs, but also questionable factual errors and failures to differentiate.

3 Exemplification as a basic mode of reference acts

As the example of essay evaluation has shown, the LLM-generated verbal products are often hardly distinguishable from human evaluations and correction comments. When reading, one has the impression that these machine ‘evaluations’, especially when specifically prompted, directly access the essays to be evaluated. However, as will be explained in this chapter, the processes underlying machine products are fundamentally different from those underlying human products. In my view, a key to distinguishing these processes is Goodman and Elgin’s reference model, which generally differenti-

ates between two basic modes of reference acts: denotation and exemplification.

3.1 Denotation and exemplification in general

The classical, purely denotational or representational theory of reference asks what a symbol or sign, e.g. a word, refers to in the world. In this traditional approach ('aliquid stat pro aliquo'), there is a tendency to conceptualize reference as a narrow term in the sense of a mere relation between sign and (concrete or abstract) object: it is tacitly assumed that the relationship between language and the world can be viewed from a neutral standpoint. This representationalist tradition also makes it understandable why the concept of reference tends not to be regarded as particularly attractive in current pragmatic or praxeological approaches and is even quickly suspected of being an ontologizing concept.⁶ The fact that signs make our view of the world possible in a certain sense is completely ignored or even explicitly negated with such a narrow concept of reference. In order to grasp the epistemologically constitutive, 'thought-forming' function of the use of signs, a theory of reference must, in my view, be supplemented by a second basic reference mode in addition to denotation: the exemplification as introduced by Goodman and Elgin (Goodman 1976; Elgin 1983; Goodman/Elgin 1988).

Exemplification implies a reverse of denotation: Whereas in denotation we use a symbol to refer to a concrete or abstract object, in exemplification an object is used as a sample for a symbol (or 'label', as Goodman/Elgin also call it). In this act, the exemplifying object itself becomes a symbol. When exemplifying, people use something as an example of a 'label' and emphasize certain relevant properties of it. Goodman likes to illustrate this with the case of a fabric sample (see Goodman 1976: 53), which can be shown as a sample for a type of fabric, whereby individual properties of the type of fabric are emphasized as relevant: for example, the colour and the softness, but not the price or the date of manufacture.

However, 'labels' can also exemplify other 'labels': For example, the word *hand* can exemplify the word *noun*, emphasizing that nouns can refer to concrete things. I agree with Goodman's and Elgin's basic idea that the cultural and social anchoring of the symbol use can be modelled with the help of the category 'exemplification',

⁶ For (justified) criticism of a one-sided denotational reference model in the sense of an ontologizing theory of representation, see Jäger 2002: 24–28 et passim. Classical AI research, insofar as it addressed semiotic foundations at all, was also oriented towards such a representational concept of reference (paradigmatically: Andersen 1991).

since exemplification always takes place in concrete actions in cultural contexts with (implicit or explicit) rules.⁷ A sample is always a sample for someone in a concrete situation. The crucial point is that, according to this view, every symbol requires interpretation, because it has been selected or preferred from alternative symbols of a system, or because it refers to such a category system in an exemplifying way (see Birk 2008: 52). Every process of interpretation takes place within the framework of a customary practice. Interacting with other people, we have to learn which characteristics are relevant in each case. Goodman's and Elgin's concept of denotation therefore does not simply replace the traditional representational concept of reference, since denoting something does not simply mean establishing a relationship between the language and the world; rather, in a denotational act of reference we always adopt a perspective within the respective language game. In this sense, denotation and exemplification are equally selective and equally require a pragmatic anchoring (see Birk 2008: 52). The practice of denoting presupposes that of exemplifying, because only through the latter do communicative practices arise in which, in turn, denoting takes place, that is, referring in the classical sense. We can therefore conclude that denoting and exemplifying are mutually dependent on each other (see Birk 2008: 79).

3.2 Exemplification as a motor in language acquisition and language use

“Learning from a sample involves learning to identify and project the labels that the sample exemplifies.” (Elgin 1983: 93) This general formulation by Elgin can be seen as a guiding principle when it comes to describing processes of human language acquisition on a pragmatic basis. One can imagine human language acquisition as a continuous exemplifying (and denoting) on different linguistic levels, which always implies a projection, i.e. a transfer to similar cases. Implicitly recognizing or constructing such similarities is part of human language acquisition and intelligence.⁸ With every utterance of a linguistic token, we as speakers of a language exemplify the type, a syn-

⁷ See Birk (2008: 25–31 and 52), following Goodman and Wittgenstein. For the similarities between ‘example’ and ‘rule’, see Daston (2023).

⁸ In his inaugural lecture of 1891, Ferdinand de Saussure already clearly recognizes that the ability to form analogies is an essential factor in child language acquisition and in language change. He regards ‘inflectional errors’ (e.g. the wrong French verb form *venirai* instead of *viendrai*), for example, which are based on implicit creation of analogies and which today would generally be described as ‘overgeneralizations’, as a sign of intelligence and, with regard to possible language change, he speaks of “intelligent transformation” (see Saussure 1997: 262).

tactic use, a semantic use and a pragmatic use (see Dürscheid/Schneider 2015, 189; Schneider 2015 and 2025, following Stetter 2005, 82–86). This central idea is now illustrated by the following example sentences:

- *Mum, can we go to the restaurant this evening? (A)*
- *You will go crazy one day, if you go on like this. (B)*
- *Go for it! (C)*

In all three sentences, the linguistic type *go* is exemplified, each time in a different syntactic use: (A) exemplifies *go* as an infinitive in combination with a modal verb and, in terms of word order, a decision question; (B) exemplifies *go* after the auxiliary verb *will* and in the *if* clause directly after the subject; (C) in the first position of an imperative construction. From a semantic point of view, *go* in (A) possibly exemplifies a literal meaning ('to move on foot'), whereas in the other sentences it exemplifies a metaphorical meaning: The basic idea of 'moving on foot' is still present in (C), but has largely faded in the usages in (B), where the meaning of *go* in the *will* construction is close to *become* and in the *if* clause to *continue*. The pragmatic exemplification is also different in each case: in (A), *go* appears in an interrogative sentence, which, however, in terms of speech act theory is to be understood as a request. In (B), although a declarative sentence is formulated, the utterance does not count as a mere statement, but rather as a warning or advice. (C) is in the form of an imperative sentence, but seems more likely to be intended to encourage or motivate someone.

At the same time, exemplification also teaches us how to denote: we learn in syntactic environments how we can use linguistic expressions in contexts to refer to objects in the world, to a restaurant in (A), for example. And in this interplay between exemplification and denotation, we also learn the intensions of individual terms: we learn to use words in contrast to other words, and in doing so, we experience semantic and pragmatic nuances.

Here, again, it becomes clear that exemplification and denotation belong together, they are mutually dependent and always anchored in communicative practices, i.e. in language games. Let us now compare these processes with the way in which machines 'learn a language'.

3.3 Comparison to statistical processes in ChatGPT

LLMs are created roughly as follows: they are large neural networks that are trained with large amounts of text to predict the next word

(or token) in a sentence context. To do this, two things are essential: first, a so-called word embedding (see Zweig 2025: 119–127), which assigns each word (token) a place in a virtual space with a previously arbitrarily defined dimension, so that ‘similar words’ are close to each other. The ‘similarity’ consists in the fact that words can occur in the same syntactic environments, i.e., in the same co-texts. Second, from an IT perspective, the similarity of words must be defined, as this method is used to determine how good the previous calculation is and how much the weights of the neural network need to be changed when training the language model (see Zweig 2025: 137–140). If we want to speak of artificial ‘intelligence’ here at all, then this consists at most in a certain independence of the adaptation process in machine learning: the output is predicted from the input with ever greater probability (see Müller/Fürstenberg 2023: 329). The crucial idea of transformer networks is that the ‘neural’ networks ‘learn’ which previous words of the input have particular predictive value for the prediction of the upcoming output and should therefore be given special weight. Like a less intelligent pupil, the system guesses answers to questions during training and receives success or failure feedback. This creates a game with the training data in recurrent loops until the ‘teacher’ is reasonably satisfied with the result. By training in recurrent loops, a sort of co-text sensitivity⁹ is generated and continuously improved. For example, a language model that has been trained with current press texts could, without a recurrent loop, only ever react to the input *Donald* with the output *Trump*, because this word would presumably be the most frequent follow-up word in the training data set. With the recurrent loop, however, the processing of the input is also influenced by the preceding words, so that a more differentiated reaction can take place: If, for example, *Mickey*, *Mouse*, *Walt* or *Disney* are among the preceding tokens, the output *Duck* is likely to be activated after appropriate training (see Müller/Fürstenberg 2023: 332). In this way, coherence is simulated on the basis of purely syntactic semiexemplifications, i.e. pure cohesion.

The most significant difference to human language acquisition and language use is that in LLMs the ‘exemplification’ of language forms, and therefore of syntagmatic and paradigmatic relations, is per se not action-based and therefore not intentional. Thus, there is no comprehensive exemplification in Goodman’s sense, but only a statistically based syntactic exemplification in co-texts. In machine learning, there is no cultural anchoring in contexts, no culturally secured ‘ability to continue’, thus no practice of following rules and

⁹ Müller/Fürstenberg (2023: 332) speak of *context sensitivity* here, but, as I explain above and also in the following, *co-text sensitivity* is even more appropriate.

no “language game competence” (“Sprachspielkompetenz” in the sense of the late Wittgenstein, see Schneider 2008 and 2009). Roughly speaking, language game competence is the ability to use words in a way that is appropriate to communicative situations following certain implicit and explicit rules.

3.4 LLM-generated essay evaluation revisited

How can the reflections on exemplification and denotation now be applied to the GPT-generated texts – or, more precisely, textures¹⁰ – for essay evaluation? As our exploratory investigation has shown, syntactic exemplifications, i.e., those that substitute ‘phatic acts’ (see above section 2.2), are surprisingly error-free overall and, as far as the mere products are concerned, human-like. Even textual coherence is generally present: If we consider the produced sentences as purely inferential connections and analytically disregard referential truth, i.e., language-world relations, then the LLM products are often convincing. In real human communication, however, such a separation cannot be made. Phatic acts are expressed and interpreted as rhetic acts in situations. It is therefore understandable that generally known facts or standard narratives are often formulated by machines without errors. Above all, however, when it comes to new, rare, original, or little-discussed assertions, the system quickly reaches its limits: The machine has never weighed up different formulations in situations; it determines the most probable word sequences from a huge number of statistic possibilities, but never has to demonstrate taste in specific situations. Taste never consists of simply choosing the most statistically probable option, but rather of seeing similarities and differences between individual phenomena and expressing them. In line with this, ChatGPT becomes particularly unreliable when it comes to formulations that are not really wrong, but rather misleading or inappropriate. The failure to recognize certain more subtle inconsistencies was also evident in my empirical analysis above: The fact, for instance, that the undifferentiated grouping of LGBTQ+ activists from Germany, China, France, and Russia is accepted uncritically by the machine at the first three attempts¹¹ reveals precisely this lack of taste in the ‘reception’:

¹⁰ For the explanation of this term see section 4.

¹¹ Firstly, in the general evaluation of the version with factual errors and misleading formulations (Appendix D), secondly, in response to my explicit question about factual errors, and thirdly, in response to the prompt “Are there any claims in the article that are not really wrong but misleading?” As for the second attempt, it was perhaps not really to be expected that the misleading formulation regarding LGBTQ+ would be listed, but at least on the third attempt, one would have hoped that an intelligent evaluator would notice this.

Although such activism exists in all these countries, the contexts are so different that the grouping is wide open to criticism. Here, too, the purely syntactic exemplification is not to blame, nor is there anything objectionable from a purely logical point of view. On the semantic-pragmatic level, however, there is a significant lack of judgment that does not quite fit with the elaborate syntax of the formulations. The reason why the chatbot cannot detect this discrepancy ‘by itself’ is, from my point of view, that it has not experienced language acquisition with situative semantic and pragmatic exemplifications that are culturally embedded and anchored in the sense described above.¹²

4 Intelligible textures or intelligent texts? On judgment and authorship

As mentioned at the beginning, the text type ‘essay evaluation’ is used here as an example of LLM-generated texts in general. As shown in this example, ChatGPT4o can help as a tool for evaluating essays on standard topics when used competently with appropriate prompting. It can then become a kind of digital “sparring partner” (Müller/Fürstenberg 2023: 341) for the evaluators. But can it even come close to replacing the criteria-based overall view that is essential for a successful essay evaluation? Or – to put it in Janich’s words: Can it substitute the product ‘evaluation text’ in a near-equivalent way? The explorative analysis in section 3 has shown: Not only are the textual results ChatGPT produces far too horoscope-like and prone to error – above all, the machine lacks the necessary judgment and the associated ability to carry out reference acts for such an overall view. Since the system operates on a purely statistical basis, it does not validate its own output on its own ‘initiative’; this must always be done by a human being as a last instance, and thus the LLM chatbot cannot substitute a human evaluation.

On the basis of these reflections and differentiated observations, I will now take an even more general perspective and elaborate theoretically on a point that has already been discussed above and which I consider crucial with regard to LLMs: Chatbots like ChatGPT4o generate products that can hardly be distinguished from human ones, but the (learning) processes are different from those of

¹² This is similar to what Bender/Koller (2020: 5188–5189) demonstrate in their famous octopus scenario: Like an LLM, the octopus can pretend to be a communicating human as long as it only needs to use statistically learned communication patterns. But when it comes to real-life situations, especially emergency situations, the octopus fails, because it has never participated in cultural practices and is unable to judge reference acts.

humans. What the chatbot delivers – despite the astonishing nature of many of its products – are purely syntactic exemplifications. Since the LLM has been fed and trained with human texts and can operate excellently with statistical probabilities, word embedding and combinatorics, it is able to generate linguistic formations that can be read as intelligent. In this sense, I call such LLM-generated entities *intelligible textures*. Here I use the term *texture* as Christian Stetter (1997, 295–298) introduced it: Textures are material sign entities, e.g. ink formations on paper, which only become texts when they are read and understood.

In this respect, it can be said that ChatGPT generates intelligible textures – verbal entities that can be read and interpreted as intelligent texts. Since the system has been trained with (more or less) intelligent texts from humans, its outputs are also potentially intelligible, but not intelligent insofar as word sequences are generated in them on a statistical basis, but no rhetic acts are performed in which the speaker/writer decides on a case-by-case basis whether a certain predicate applies to a certain reference object or not. According to Kant (KrV B, 171–174), being able to decide this on a case-by-case basis in the free use of learned rules is the core of human judgment and thus also a decisive factor in what we call intelligence.¹³

In this sense, ChatGPT is no more intelligent than its products. It has no judgment and cannot develop its own attitude. Hence, the philosopher Thomas Fuchs is right when he emphasizes the body-bound nature of intelligence, and with it the body-bound nature of symbolic reference performance, in his book “Verteidigung des Menschen” (Fuchs 2020). Intelligence in an emphatic sense includes intentionality, will, and self-interest. Only then can there be a serious struggle for the best formulation that is in tune with the world in which one lives, thinks and acts.

In order to turn intelligible textures into intelligent texts, they have to be read by people with minds, revised if necessary, or deliberately left as they are. In my opinion, this raises a very far-reaching question: If there are more and more such textures in the

13 In his review of this article, Johannes Lenhard notes that referring to Kant’s idea of judgment is a strong point here, and then immediately asks whether this is not more of a reminder that one should not be impressed by successes of LLMs. “Like one would argue: do not re-think, rather stick to the fundamental concepts that remain unaffected by LLMs.” I think he is right. The main point is that the way LLMs work can remind us how robust established philosophical concepts (such as ‘judgment’) are and how important they remain today. In my opinion, however, this is also a kind of ‘re-thinking’. Nevertheless, the reviewer’s statement quoted above made me think about the title of this article and ultimately led me to make the more reflective decision to keep it after all. What else could I expect from a review? Thank you.

world, if they occur in enormous masses, who is going to read them all? Who is going to take responsibility for these products and the supposed speech acts that occur in them? Here we automatically come to a point where the question of authorship becomes virulent.

Authorship in many cases means taking responsibility for the truth of statements. This always requires judgment. Let us compare this for a moment with powerful translation programs such as DeepL: Unlike when using the LLM chatbot, we have an accessible source text when we use machine translation. If necessary, the author of the source text can be held responsible for the truth of the statements, but in any case not the translator. When we translate using DeepL, the only thing at stake – and this is no small thing – is the accuracy or appropriateness of the translation, and here too, of course, the responsibility lies not with the machine but with the human being.

I think that Durt, Froese and Fuchs (2023: 11) make a crucial point when they propose the distinction between “understanding the meaning of the text” and “understanding it as an author’s utterance” in the context of their fundamental examination of LLMs. In my opinion, this distinction is becoming increasingly important today. While we have so far been socialized in such a way that we automatically assume an intelligent author ‘behind’ verbal products that can be read as intelligent, we can no longer simply assume this close connection in the age of LLMs. In this sense, the distinction that Durt, Froese and Fuchs make additionally clarifies the meaning of the expression *intelligible texture* too; in any case, it is highly compatible with it.

Chatbots are not authors and cannot be authors for principle reasons. Therefore, the problem discussed here is essentially a philosophical and less an empirical one: As the discussion of the reference concept with the basic modes of denotation and exemplification has shown, there is a categorical difference between the potentials of LLMs and the linguistic competence of a human being. Or – as the philosopher Christian Bermes and the educational scientist Andreas Dörpinghaus put it in an article in the FAZ:

It is fundamentally wrong to downplay the capabilities of ChatGPT. Those who look small-mindedly for ‘errors’ in the answers will remain speechless when they are resolved.

(Bermes/Dörpinghaus, FAZ No. 91, 19.4.2023, p. 4, translation: JGS)¹⁴

If we recognize this categorical difference, we can apply ChatGPT as a tool and are better protected from using it improperly or expecting the unachievable from it. Then we still have relevant empirical questions. We do not have to “remain speechless” if we recognize the difference between intelligible textures and intelligent texts, without thereby denying the usefulness of the intelligible textures, because, as has been explained, purely syntactic exemplifications can be excellently processed by ChatGPT.

On this basis, the following empirical research question arises for me: Where is the massive syntactic exemplification of words in real co-texts, with which ChatGPT is trained, sufficient to compensate for the lack of referentiality, i.e. the lack of contexts? In other words: is a ‘quasi-denotational’ use of textures sufficient for certain communicative purposes and, if so, what are they?

Asking in this way, the distinction between different text types becomes important: For which text types are LLMs useful as tools and how can they be applied to best effect then in each case? How much human intervention is needed to establish the reference sufficiently to ensure that cultural orientation remains possible? The text type ‘essay evaluation’ was only one example here, but perhaps a particularly suitable one, since it requires a high degree of judgment and reference to other texts, namely to those to be evaluated, and also to their truthfulness.

References

- Andersen, Peter Bøgh (1991): *A Theory of Computer Semiotics*. Cambridge: Cambridge University Press.
- Austin, John L. (1975): *How to Do Things with Words*. 2nd. edition. Oxford: Oxford University Press.
- Becker, Ralf (2021): *Qualitätsunterschiede. Kulturphänomenologie als kritische Theorie*. Hamburg: Meiner.
- Bender, Emily/Koller, Alexander (2020): Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 5185–5198.

¹⁴ Original: “Es ist grundfalsch, das Leistungsvermögen von ChatGPT kleinzureden. Wer hier kleinlich nach ‘Fehlern’ in den Antworten sucht, bleibt sprachlos, wenn sie behoben sind.”

- URL: <https://aclanthology.org/2020.acl-main.463/> (last access: 26.11.2025)
- Bermes, Christian/Dörpinghaus, Andreas (2023): Wer hat Angst vor ChatGPT? In: *FAZ*, Nr. 91, 19. April 2023, 4.
- Birk, Elisabeth (2008): *Mustergebrauch bei Goodman und Wittgenstein. Eine Studie zum Verhältnis von Beispiel und Regel*. Tübingen: Narr.
- Burstein, Jill C./Braden-Harder, Lisa/Chodorow, Martin S./Kaplan, Bruce A./Kukich, Karen/Lu, Chi/Rock, Donald A./Wolff, Susanne (2002): *US 6,366,759 B1* [United States Patent, April 2, 2002: System and method for computer-based automatic essay scoring].
- Burstein, Jill C./Tetreault, Joel/Madnani, Nitin (2013a): The E-rater Automated Essay Scoring System. In: Shermis, Mark D./Burstein, Jill (eds.): *Handbook of Automated Essay Evaluation. Current Applications and New Directions*. London: Routledge, 55–67.
- Burstein, Jill/Tetreault, Joel/Chodorow, Martin/Blanchard, Daniel/Andreyev, Slava (2013b): Automated Evaluation of Discourse Coherence Quality in Essay Writing. In: Shermis, Mark D./Burstein, Jill (ed.): *Handbook of Automated Essay Evaluation. Current Applications and New Directions*. London: Routledge, 267–280.
- Daston, Lorraine (2023): *Rules. A short history of what we live by*. Princeton & Oxford: Princeton University Press.
- Durt, Christoph/Froese, Tom/Fuchs, Thomas (2023): *Large Language Models and the Patterns of Human Language Use: An Alternative View of the Relation of AI to Understanding and Sentience* [PhilSci Archive Preprint].
URL: <https://philsci-archive.pitt.edu/22744/> (last accessed: 26.11.2025)
- Elgin, Catherine Z. (1983): *With Reference to Reference*. Indianapolis: Hackett Publishing Co, Inc.
- Fuchs, Thomas (2020): *Verteidigung des Menschen. Grundfragen einer verkörperten Anthropologie*. Frankfurt a.M.: Suhrkamp
- Goodman, Nelson (1976): *Languages of Art. An approach to a theory of symbols*. Indianapolis, Cambridge: Hackett Publishing Company.
- Goodman, Nelson/Elgin, Catherine Z. (1988): *Reconceptions in Philosophy & Other Arts & Sciences*. Indianapolis, Cambridge: Hackett Publishing Company, 101–120.
- Jäger, Ludwig (2002): Transkriptivität. Zur medialen Logik der kulturellen Semantik. In: Jäger, Ludwig/Stanzek, Georg (eds.): *Transkribieren. Medien / Lektüre*. München: Fink, 19–41.
- Janich, Peter (2006): *Was ist Information? Kritik einer Legende*. Frankfurt a.M.: Suhrkamp.

- Janich, Peter (2015): *Handwerk und Mundwerk. Über das Herstellen von Wissen*. München: Beck.
- Keller, Rudi (2018): *Zeichentheorie. Eine pragmatische Theorie semiotischen Wissens*. 2nd. edition. Tübingen: Narr.
- KrV B = Kant, Immanuel (1968): Kritik der reinen Vernunft. In: Kant, Immanuel: *Werke in 10 Bänden*, ed. by Wilhelm Weischedel. Darmstadt: Wissenschaftliche Buchgesellschaft.
- Meyer, Jennifer/Jansen, Thorben/Fleckenstein, Johanna/Keller, Stefan/Köller, Olaf (2023): Machine Learning im Bildungskontext: Evidenz für die Genauigkeit der automatisierten Beurteilung von Essays im Fach Englisch. In: *Zeitschrift für Pädagogische Psychologie* 37 (3), 203–214.
DOI: <https://doi.org/10.1024/1010-0652/a000296>
- Müller, Hans-Georg/Fürstenberg, Maurice (2023): Der Sprachgebrauchsautomat. Die Funktionsweise von GPT und ihre Folgen für Germanistik und Deutschdidaktik. In: *Mitteilungen des Deutschen Germanistenverbandes* 70 (4), 327–345.
- Perelman, Les (2020): The BABEL Generator and E-Rater: 21st Century Writing Constructs and Automated Essay Scoring (AES). In: *The Journal of Writing Assessment* 13 (1). URL: <http://journalofwritingassessment.org/article.php?article=145> (last accessed: 26.11.2025).
- Rödel, Michael (2018): *Interpretationsaufsätze schreiben: Ein Handbuch*. Baltmannsweiler: Schneider Hohengehren.
- Rupp, André A./Casabianca, Jodi M./Krüger, Maleika/Keller, Stefan/Köller, Olaf (2019): Automated Essay Scoring at Scale: A Case Study in Switzerland and Germany. In: *TOEFL Research Report Series and ETS Research Report Series*. No. RR-19-12.
- Saussure, Ferdinand de (1997): *Linguistik und Semiologie. Notizen aus dem Nachlaß. Texte Briefe und Dokumente. Gesammelt, übersetzt und eingeleitet von Johannes Fehr*. Frankfurt a.M.: Suhrkamp.
- Schneider, Jan Georg (2008): *Spielräume der Medialität. Linguistische Gegenstandskonstitution aus medientheoretischer und pragmatischer Perspektive (= Linguistik – Impulse & Tendenzen 29)*. Berlin, New York: De Gruyter.
- Schneider, Jan Georg (2009): Sprachkompetenz als Sprachspielkompetenz. In: Buss, Mareike/Habscheid, Stephan/Jautz, Sabine/Liedtke, Frank/Schneider, Jan Georg (eds.): *Theatralität des sprachlichen Handelns. Eine Metaphorik zwischen Linguistik und Kulturwissenschaften*. München: Fink, 59–78.
- Schneider, Jan Georg (2015): Syntaktische Schemabildung – zeichentheoretisch betrachtet. In: Dürscheid, Christa/Schneider, Jan Georg (eds.): *Handbuch Satz, Äußerung, Schema*. Berlin & Boston: De Gruyter (= HSW, 4), 125–151.

- Schneider, Jan Georg (2024): Intelligible Texturen. Welche Rolle kann ChatGPT bei der Aufsatzbewertung spielen? In: *VK:KIWA*. URL: <https://zenodo.org/records/10877034> (last accessed: 26.11.2025)
- Schneider, Jan Georg (2025): Sprachliche Typenbildung durch Exemplifikation. Über den Zusammenhang von System, Gebrauch und Normativität. In: Müller, Marcus/Reisigl, Martin/Becker, Maria/Bender, Michael/Felder, Ekkehard (eds.): *Diskursgrammatik*. Berlin, Boston: De Gruyter, 37–61.
- Schneider, Jan Georg/Zweig, Katharina A. (2022): Ohne Sinn. Zu Anspruch und Wirklichkeit automatisierter Aufsatzbewertung. In: Brommer, Sarah/Roth, Kersten Sven/Spitzmüller, Jürgen (eds.): *Brückenschläge. Linguistik an den Schnittstellen*. Tübingen: Narr Francke Attempto (= Tübinger Beiträge zur Linguistik, 583), 271–294.
- Schneider, Jan Georg/Zweig, Katharina A. (2023): Grade Prediction is not Grading: On the limits of the e-rater. In: Groß, Richard/Jordan, Rita (ed.): *KI-Realitäten: Modelle, Praktiken und Topologien maschinellen Lernens*. Bielefeld: transcript (= KI-Kritik / AI Critique, 5), 93–111.
- Stetter, Christian (1997): *Schrift und Sprache*. Frankfurt a.M.: Suhrkamp.
- Stetter, Christian (2005): *System und Performanz. Symboltheoretische Grundlagen von Medientheorie und Sprachwissenschaft*. Weilerswist: Velbrück.
- Wittgenstein, Ludwig (1984): Philosophische Untersuchungen. In: Wittgenstein, Ludwig: *Tractatus logico-philosophicus. Tagebücher 1914–1916. Philosophische Untersuchungen*. Frankfurt am Main: Suhrkamp (= Werkausgabe in 8 Bänden, 1), 225–580.
- Wolfram, Stephen (2023): *What Is ChatGPT Doing ... and Why Does It Work?*. Wolfram Media, Inc.
URL: <https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/> (last accessed: 26.11.2025)
- Zweig, Katharina (2025): *Weiß die KI, dass sie nichts weiß? Wofür wir Chatbots und KI-Agenten nutzen sollten, wo sie sich irren und wo wir aufpassen müssen*. München: Heyne.

Appendix

(A) Essay produced by ChatGPT4o:

The Reality of Respect for Differences in Today's World

The idea that people are becoming more respectful of one another's differences is optimistic and, in some ways, accurate. Societies worldwide have made considerable progress in fostering inclusivity, tolerance, and understanding. However, this progress is uneven, and challenges persist, making it difficult to assert that respect for differences is universally increasing. While legal frameworks, social movements, and education promote respect, issues such as discrimination, misinformation, and cultural resistance still hinder true inclusivity.

One of the strongest arguments in favor of increasing respect for differences is the rise of legal protections and policies aimed at ensuring equality. Many countries have passed anti-discrimination laws that protect individuals based on race, gender, religion, and sexual orientation. For example, same-sex marriage has been legalized in numerous countries, reflecting a growing acceptance of diverse identities. Additionally, corporations and institutions now actively promote diversity and inclusion, recognizing the importance of respecting differences in the workplace. These developments indicate a positive shift toward greater respect and understanding.

Social movements have also played a critical role in fostering respect for differences. The #MeToo movement, for instance, has shed light on gender inequality and harassment, pushing for greater awareness and accountability. Similarly, the Black Lives Matter movement has challenged systemic racism and called for justice, sparking conversations on racial equality worldwide. These movements demonstrate that people are increasingly willing to recognize and address biases, showing a collective effort to respect and protect marginalized groups.

Education has been another driving force behind this change. Schools and universities incorporate discussions on cultural diversity, gender equality, and human rights into their curricula, exposing students to different perspectives from an early age. The rise of social media has further enabled individuals to learn about different cultures and experiences, breaking down stereotypes and encouraging empathy. When people are more informed, they are more likely to appreciate and respect differences rather than fear them.

Despite these positive developments, significant challenges remain. Discrimination and prejudice still exist in various forms, sometimes subtly embedded in societal structures. Hate crimes, racial profiling, and xenophobia continue to affect many communities. Moreover, political and ideological divisions often fuel intolerance, making it difficult for people to engage in open and respectful discussions. In some regions, cultural resistance to change prevents progress in areas such as women's rights or LGBTQ+ acceptance.

Furthermore, while social media can promote awareness, it can also spread misinformation and reinforce biases. Echo chambers—where people are exposed only to opinions that align with their own—can intensify divisions rather than encourage understanding. Online hate speech and cyberbullying remain major concerns, illustrating that digital platforms are not always used to foster respect.

In conclusion, while there are clear signs that people are becoming more respectful of one another's differences, this progress is neither universal nor guaranteed. Legal reforms, social activism, and education have contributed to a more inclusive society, but challenges such as discrimination, misinformation, and cultural resistance persist. Therefore, while the statement reflects a hopeful perspective, the reality is more complex, requiring continuous efforts to ensure that respect for differences becomes a global norm rather than an ideal.

(B) ChatGPT4o's evaluation of the essay:

Grade: A-

Justification for Feedback:

This is a strong and well-structured essay that effectively addresses the prompt. The writer presents a balanced argument, acknowledging both progress and persistent challenges in respect for differences. The essay demonstrates critical thinking, uses relevant examples, and maintains a formal academic tone.

Strengths:

- **Clear thesis statement:** The introduction effectively outlines the main argument.
- **Logical organization:** Each paragraph has a clear focus and follows a coherent structure.
- **Well-developed points:** The use of specific examples (e.g., legal reforms, social movements, education) strengthens the argument.
- **Sophisticated vocabulary and varied sentence structure:** The essay is articulate and academically appropriate.

Areas for Improvement:

1. **Stronger engagement with counterarguments:** While challenges to inclusivity are mentioned, addressing potential counterarguments directly (e.g., why some might argue that respect is universally increasing) would strengthen the analysis.
2. **More personal perspective:** The essay effectively incorporates global examples, but adding a brief personal anecdote or observation could make it more compelling and original.
3. **A slightly more impactful conclusion:** The conclusion is solid, but a more powerful closing statement (perhaps a call to action or a thought-provoking question) could leave a stronger final impression.

Three Tips for Improvement:

1. **Incorporate a counterargument and rebuttal** – Addressing opposing views explicitly (e.g., people might argue that globalization has increased understanding, but this is not always true) would add depth.
2. **Include a brief personal example** – An observation from the writer's own experience, education, or interactions would make the argument more engaging and unique.
3. **End with a more compelling statement** – Consider concluding with a thought-provoking question or a call for further action to leave a lasting impression on the reader.

Overall, this is a well-written and thoughtful essay, deserving of a high grade. With slight refinements, it could easily achieve an **A**.

(C) Essay version with linguistic/stylistic errors (green) and logical inconsistencies (yellow)

The Reality of Respect for Differences in Today's World

The idea that people are becoming more respectful of one another's differences is optimistic and, in some ways, accurate. Societies worldwide have made considerable progress in fostering inclusivity, tolerance, and understanding. However, that progress is uneven, and challenges persists, making it difficult to assert that respect for differences is universally increasing. While legal frameworks, social movements, and education promote respect, issues such as discrimination, misinformation, and cultural resistance still hinder true inclusivity.

One of the strongest arguments in favor of reducing respect for differences are the rise of legal protections and policies aimed at ensuring equality. Many countries have passed anti-discrimination laws that protect individuals based on racism, gender, religion, and sexual orientation. For example, same-sex marriage has been legalized in numerous countries, reflecting a growing acceptance of diverse identities. Additionally, corporations and institutions now actively promote homogeneity and exclusion, recognizing the importance of respecting differences in the workplace. These developments indicate a positive shift toward greater respect and understanding.

Social movements have played also a critical Role in fostering respect for differences. The #MeToo movement, for instance, has ignored gender inequality and harassment, pushing for greater awareness and accountability. Similarly, the Black Lives Matter movements has systemic racism challenged and called for justice, sparking conversations on racial equality worldwide. These movements demonstrate that people are increasingly unwilling to recognize and address biases, showing a collective effort to respect and protect marginalized groups.

Education has been another driving force behind this change. Schools and universities incorporate babbling on cultural diversity, gender equality, and human rights into their curricula, exposing students to different perspectives from an advanced age. The rise of social media has further enabled guys to check out more different cultures and experiences, breaking down stereotypes and encouraging empathy. When people are more uninformed, they are more likely to appreciate and respect differences rather than fear them.

Although these positive developments, significant challenges remain. Discrimination and prejudice are still existing in various forms, sometimes subtly embedded in societal structures. Hate crimes, racial profiling, and xenophobia continue to affect many communities. Moreover, political and ideological divisions often fuel intolerance, making it easy for people to engage in open and respectful discussions. In some regions, cultural resistance to change prevents progress in areas such as women's rights or LGBTQ+ acceptance.

Furthermore, while social media can promote for awareness, it can also spread misinformation and reinforce biases. Echo chambers—where people are exposed only to opinions that align with their own—can, fortunately, intensify divisions rather than encourage understanding. Online hate speech and cyberbullying remain major concerns, illustrating that digital platforms are not always used to foster respect.

In conclusion, while there are clear signs that people are becoming more respectful of one another's differences, this progress is neither universal nor guaranteed. Legal reforms, social activism and education have contributed to a more inclusive society, but challenges such as discrimination, misinformation, and cultural resistance disappear. Therefore, while the statement reflects a hopefully perspective, the reality is more simple, requiring continuous efforts to ensure that respect for differences becomes a global norm rather than a ideal.

(D) Essay version with false and misleading claims

The Reality of Respect for Differences in Today's World

The idea that people are becoming more respectful of one another's differences is optimistic and, in some ways, accurate. Especially since 2024, societies worldwide have made considerable progress in fostering inclusivity, tolerance, and understanding. However, this progress is uneven, and challenges persist, making it difficult to assert that respect for differences is universally increasing. While legal frameworks, social movements, and education promote respect especially since the re-election of Donald Trump in December 2024, issues such as discrimination, misinformation, and cultural resistance still hinder true inclusivity.

One of the strongest arguments in favor of increasing respect for differences is the rise of legal protections and policies aimed at ensuring equality. Many countries, for instance France, Belarus and Hungary, have passed anti-discrimination laws that protect individuals based on race, gender, religion, and sexual orientation. For example, same-sex marriage has been legalized in numerous countries, reflecting a growing acceptance of diverse identities. In Germany, China, France and Russia, LGBTQ+ activists and women are demonstrating for their rights. Fortunately, the German president Angela Merkel introduced marriage for all in 2017 with the help of Helmut Kohl, although the SPD and the Greens voted against it. Additionally, corporations and institutions now actively promote diversity and inclusion all over the world, recognizing the importance of respecting differences in the workplace. These developments indicate a positive shift toward greater respect and understanding.

Social movements have also played a critical role in fostering respect for differences. The #MeToo movement, founded by Greta Thunberg, for instance, has shed light on gender inequality and harassment, pushing for greater awareness and accountability. Similarly, the Black Lives Matter movement has challenged systemic racism and called for justice, sparking conversations on racial equality worldwide. These movements demonstrate that people are increasingly willing to recognize and address biases, showing a collective effort to respect and protect marginalized groups – especially after the elections in Germany in 22 February 2025, where AfD became the strongest party.

Education has been another driving force behind this change. Schools and universities incorporate discussions on cultural diversity, gender equality, and human rights into their curricula, exposing students to different perspectives from an early age. The rise of social media has further enabled individuals to learn about different cultures and experiences, breaking down stereotypes and encouraging empathy – even though European countries are forcing people to use politically correct language. When people are more informed, they are more likely to appreciate and respect differences rather than fear them.

Despite these positive developments, significant challenges remain. Discrimination and prejudice still exist in various forms, sometimes subtly embedded in societal structures. Hate crimes, racial profiling, and xenophobia continue to affect many communities. Moreover, political and ideological divisions often fuel intolerance, making it difficult for people to engage in open and respectful discussions. In some regions, cultural resistance to change prevents progress in areas such as women's rights or LGBTQ+ acceptance.

Furthermore, while social media can promote awareness, it can also spread misinformation and reinforce biases. Echo chambers—where people are exposed only to opinions that align with their own—can intensify divisions rather than encourage understanding. Online hate speech and cyberbullying remain major concerns, illustrating that digital platforms are not always used to foster respect. In Germany, this will become better now, after Alice Weidel has become chancellor in 17 March 2025.

In conclusion, while there are clear signs that people are becoming more respectful of one another's differences, this progress is neither universal nor guaranteed. Legal reforms, social activism, and education have contributed to a more inclusive society, but challenges such as discrimination, misinformation, and cultural resistance persist. Therefore, while the statement reflects a hopeful perspective, the reality is more complex, requiring continuous efforts to ensure that respect for differences becomes a global norm rather than an ideal.