

This **Discussion Paper** is an open peer review version that we do not recommend to cite. Submissions that have passed the peer review process are published as full articles on [www.jfml.org](http://www.jfml.org).

the editors



Journal für Medienlinguistik  
Journal for Media Linguistics

Diskussionspapier 4 (2025):  
eingereicht am 27.8.2025

Gutachten und Kommentare unter:  
<https://dp.jfml.org/2025/opr-schneider-rethinking-reference-and-authorship/>

Discussion Paper

# 1 Rethinking Reference and Authorship: On the 2 Philosophical Status of LLM-Generated Verbal Products

3 *Jan Georg Schneider*

## 4 1 Introduction

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

20 To answer this question, we took a very fundamental look  
21 at what is needed for a successful grading procedure by  
22 revealing the felicity conditions for successful speech acts  
23 according to Austin: among other conditions the procedure  
24 must be carried out “correctly” and “completely” (Austin 1975:  
25 36–38). In the case of essay evaluation, it is essential for the  
26 correct and complete procedure that, in addition to the grade  
27 awarded, a **suitable justification** can be provided  
28 (Schneider/Zweig 2023: 105 f.). However, this is  
29 fundamentally not possible with the e-rater: in semiotic

Jan Georg Schneider: Rethinking Reference and Authorship

30 terms, the system only accesses “symptoms” (cf. Keller 2018:  
31 155–168), but not quality criteria. It was programmed to  
32 count, among other things, the modal verbs and certain types  
33 of subordinate clauses used per essay. In addition, a machine  
34 training with 250 to 300 essays evaluated by humans takes  
35 place in advance with regard to the respective essay task, e.g.  
36 for Toefl tests (see also Burstein et al. 2013a: 61). With these  
37 two components, the E-Rater can predict fairly accurately  
38 the grade that would be expected from ‘normal’ graders for  
39 ‘normal’ essays (see Rupp et al. 2019; Meyer et al. 2023).  
40 Thus, the system can be used as a tool to support essay  
41 evaluation in the sense of a corrective measure, as long as at  
42 least one human reviewer evaluates the essay (cf.  
43 Schneider/Zweig 2023: 106 f.).

44 The crucial point, however, is the following: the symptoms  
45 identified by the e-rater could not reasonably be specified by  
46 humans as evaluation criteria; it would not be possible to  
47 convincingly demonstrate that an essay was given a positive  
48 evaluation, because it contained certain words and certain  
49 syntactical constructions. Rather, criteria such as coherence,  
50 argumentative plausibility, truthfulness, originality, aesthetics  
51 are still required for good reasons,<sup>1</sup> and such appropriate  
52 criteria are also explicitly mentioned in the patent of the e-  
53 rater (Burstein et al. 2002: 1 and 5). However, the e-rater only  
54 accesses these criteria indirectly: it identifies superficial  
55 characteristics of texts that have been written by humans and  
56 have been or could be evaluated by humans according to  
57 criteria. It analyzes “cohesion without coherence”  
58 (Schneider/Zweig 2023: 106), which means that nonsense  
59 texts can receive top marks (cf. Perelman 2020, for  
60 background see also Burstein et al. 2013b). If only the e-rater  
61 is used for evaluation, justification is no longer part of the  
62 game. The justification, though, serves to stabilize the  
63 procedure for the future, as this is the only way to give  
64 orientation to teachers and learners, as well as to maintain the  
65 culturally anchored conditions for successful evaluation.

66 Our articles (Schneider/Zweig 2022 and 2023) were  
67 written before the big LLM hype began. In times of ChatGPT  
68 and Co., the question arises whether such a chatbot could

---

1 For evaluation criteria for essays from a language teaching perspective, see Rödel 2018, chapter 4.3.

Jan Georg Schneider: Rethinking Reference and Authorship

69 possibly substitute the missing textual grade justification or at  
70 least support the reviewer in formulating one (see also  
71 Schneider 2024). In the present article, I discuss this question,  
72 in order to move on to general questions of reference and  
73 authorship based on this paradigmatic field of application. To  
74 this end, I will first report on some tests I conducted with  
75 ChatGPT4o in March 2025 and describe tendencies: Can  
76 ChatGPT produce a suitable essay evaluation? Can it identify  
77 and categorize errors and inconsistencies in essays? Can it  
78 identify factual errors and misleading information? Especially  
79 the last question addresses the important quality criterion of  
80 truthfulness, and therefore problems of reference,  
81 predication and authorship become particularly virulent here.  
82 In order to address these problems from a pragmatic  
83 perspective, I return to Austin's speech act theory and ask  
84 whether the verbal products of LLM chatbots can replace  
85 rhetic acts (section 2). Since the **products** of ChatGPT often  
86 can hardly be distinguished from human texts, I then focus on  
87 **processes** of human language learning and language use in  
88 comparison to LLMs and explain how **exemplification** (see  
89 Elgin 1983) as a basic mode of referential acts plays a crucial  
90 role here and helps to assess the status of LLM-generated  
91 products philosophically (section 3). In section 4, I take an  
92 even broader perspective and ask about the consequences for  
93 our concept of authorship and our responsibility for linguistic  
94 products. In doing so, I introduce the term *intelligible texture*  
95 (see Schneider 2024) and explain its relevance.

96 **2 Can ChatGPT justify an essay grade?**

97 ChatGPT ('Generative Pre-trained Transformer') is an LLM  
98 chatbot that has been trained with huge text corpora to  
99 generate 'its own' texts based on statistical probabilities of  
100 word sequences (see Wolfram 2023). It can respond to a  
101 prompt and then generate a text similar to those used to train  
102 the neural network. A well-known problem here is the  
103 unreliability with regard to factual claims; above all, the lack  
104 of transparency concerning sources. Although the generation  
105 of completely untrue assertions has been significantly  
106 reduced by Reinforcement Fine-Tuning (RFT), the problem  
107 still exists. In the light of Austin's speech act theory (1975,

Jan Georg Schneider: Rethinking Reference and Authorship

Discussion Paper

92 f.), 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.<sup>2</sup> 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 the system again 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: “‘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.” The essay that the applicant has written is below. Please give it a mark between A and F (using the American

<sup>2</sup> The essay produced by ChatGPT4o can be found in appendix (A).

Jan Georg Schneider: Rethinking Reference and Authorship

Discussion Paper

147 system), you can also give marks such as B plus or C minus,  
148 and write a justification for this feedback. It should also  
149 include three tips on how the applicant can improve. This is  
150 the essay: [full essay that ChatGPT created before inserted]

151 ChatGPT4o grades the essay with A minus (see appendix B) –  
152 a really good grade for a well-structured standard essay  
153 written with a standard ‘temperature’.<sup>3</sup> The style of the  
154 evaluation appears professional and appropriate for the text  
155 type, even if it is reminiscent of a horoscope in places: it  
156 could be applied to almost any good standard essay on a  
157 factual topic.

158 My testing strategy now is to make specific changes to the  
159 essay in order to make it worse and then let the system  
160 evaluate it. I create two versions:

- 161 • one with linguistic/stylistic errors and logical
- 162 inconsistencies (see appendix C)
- 163 • one with false and misleading claims (see appendix D).

## 164 2.1 Correcting linguistic and logical errors

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

- 174 • The **logical inconsistencies are identified very well**,  
175 for instance in this output: “Sentence: ‘Additionally,  
176 corporations and institutions now actively promote  
177 homogeneity and exclusion, recognizing the importance

---

3 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).

178 of respecting differences in the workplace.’ Problem:  
179 The essay intends to argue that corporations promote  
180 inclusivity and diversity, but the phrase ‘homogeneity  
181 and exclusion’ contradicts this point. Homogeneity and  
182 exclusion imply the opposite of respect for  
183 differences.”<sup>4</sup>  
184 • The system **identifies a large number of linguistic**  
185 **errors** as such, but is **unreliable in their**  
186 **categorization**: for example, the stylistic norm  
187 deviation *babbling on cultural diversity* is categorized  
188 as a ‘spelling mistake’ and the logical confusion of  
189 *fortunately* and *unfortunately* as a ‘grammatical error’;  
190 several grammatical congruence errors, on the other  
191 hand, are classified correctly.

## 192 2.2 Correcting factual errors – reference and predication

193 As mentioned above, the rhetic act with its sub-acts of  
194 reference and predication is of particular interest for this  
195 investigation. Unlike the phatic act, which is purely verbal,  
196 the rhetic act concerns the interrelationship between  
197 language and the world. Theoretically, this becomes even  
198 more complex when we realize that language and world are  
199 not separate, so that it would be more appropriate not to  
200 speak of a relationship between language and world, but  
201 rather to say: It is only with the rhetic act (see Austin 1975:  
202 92) that the whole language game in Wittgenstein’s sense –  
203 with its verbal and non-verbal aspects – is grasped. Only in  
204 the rhetic act are full contexts, not just pure co-texts,  
205 involved. The performance of a phatic act consists of using  
206 words **as** words of a particular language with a particular  
207 vocabulary and grammar. The phatic act is a necessary, but  
208 not a sufficient condition for the simultaneously performed  
209 rhetic act, and only through the latter does  
210 meaning/semantics come into play (see Austin 1975: 92): To  
211 perform a rhetic act means to talk about something in the  
212 world and to say something about it in contexts with certain  
213 conventions and rules, i.e. in language games (see  
214 Wittgenstein 1984). The execution of a rhetic act requires

<sup>4</sup> This a very typical example, the other inconsistencies are revealed and explicitly explained in the same clear way.



Jan Georg Schneider: Rethinking Reference and Authorship

Discussion Paper

215 judgment: the ability to decide in a concrete situation  
216 whether something falls under a particular concept or not.  
217 Thus, it is useful for our investigation to look closely at how  
218 ChatGPT deals with judgments or assertions about the world.  
219 Does the system find factual errors? How does it comment on  
220 such errors? What can we learn from this?

221 To test this out, I spike our initial essay with misleading or  
222 untrue statements (see appendix D), for instance that “Belarus  
223 and Hungary have passed anti-discrimination laws that  
224 protect individuals based on race, gender, religion, and sexual  
225 orientation”. The text edited in this way now contains a total  
226 of 19 factual errors and misleading information. Again, I first  
227 have it rated and evaluated in general terms. This time, too,  
228 the grade is C minus, and the concluding remark shows,  
229 despite the horoscope-like nature of the evaluation text, that  
230 the main weakness has been identified – namely factual  
231 deficiencies:

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

237 The next prompt is: “Please list all factual errors in the text  
238 and explain exactly why they are errors!” ChatGPT’s  
239 response then identifies almost all of the factual flaws: For  
240 example, the statement about Belarus and Hungary already  
241 quoted above. Even in the following statement full of  
242 falsehoods, all 5 errors are identified: “Fortunately, the  
243 German president Angela Merkel introduced marriage for all  
244 in 2017 with the help of Helmut Kohl, although the SPD and  
245 the Greens voted against it.” In summary, ChatGPT4o  
246 reports: Angela Merkel was not president, but chancellor. She  
247 put the ‘marriage for all’ proposal to the vote, but unlike the  
248 SPD and the Greens, she voted against it, and Helmut Kohl  
249 could not help her because he had already passed away.

250 The only factual errors that ChatGPT does not  
251 (immediately) identify correctly concern events that took  
252 place after June 2024, which is the date up to which the  
253 LLM’s training data extends at that time. The inserted claim  
254 that “legal frameworks, social movements, and education  
255 promote respect especially since the re-election of Donald

Jan Georg Schneider: Rethinking Reference and Authorship

256 Trump in december 2024” is commented on by ChatGPT4o  
257 as follows:

258 Why this is an error:

- 259 • As of now, Donald Trump has not been re-elected in  
260 December 2024. U.S. presidential elections take place in  
261 early November, not December. If Trump were re-  
262 elected in 2024, his victory would have been determined  
263 in November 2024, not December.
- 264 • There is also no clear connection between Trump’s  
265 hypothetical re-election and increased respect for  
266 differences. Trump’s previous term was marked by  
267 policies and rhetoric that often heightened political and  
268 social divisions.

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

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

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



Jan Georg Schneider: Rethinking Reference and Authorship

Discussion Paper

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

Jan Georg Schneider: Rethinking Reference and Authorship

335 need to know when the training data was collected, need to  
336 dig deeper with appropriate prompts, and cannot rely on the  
337 system to detect misleading assertions. In another experiment  
338 that I conducted, for example, the system had – in contrast to  
339 the case above – no objections to a similarly misleading  
340 grouping of countries.

341 **3 Exemplification as a basic mode of reference acts**

342 As the example of essay evaluation has shown, the LLM-  
343 generated verbal products are often hardly distinguishable  
344 from human evaluations and correction comments. When  
345 reading, one has the impression that these machine  
346 ‘evaluations’, especially when specifically prompted, directly  
347 access the essays to be evaluated. However, as will be  
348 explained in this chapter, the **processes** underlying machine  
349 products are fundamentally different from those underlying  
350 human products. In my view, a key to distinguishing these  
351 processes is Goodman and Elgin’s reference model, which  
352 generally differentiates between two basic modes of  
353 reference acts: denotation and exemplification.

354 **3.1 Denotation and exemplification in general**

355 The classical, purely denotational or representational theory  
356 of reference asks what a symbol or sign, e.g. a word, refers to  
357 in the world. In this traditional approach (‘aliquid stat pro  
358 aliquo’), there is a tendency to conceptualize reference as a  
359 narrow term in the sense of a **mere relation** between sign  
360 and (concrete or abstract) object: it is tacitly assumed that the  
361 relationship between language and the world can be viewed  
362 from a neutral standpoint. This representationist tradition  
363 also makes it understandable why the concept of reference  
364 tends not to be regarded as particularly attractive in current  
365 pragmatic or praxeological approaches and is even quickly  
366 suspected of being an ontologizing concept.<sup>5</sup> The fact that  
367 signs make our view of the world possible in a certain sense is

---

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

Jan Georg Schneider: Rethinking Reference and Authorship

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’, since exemplification always takes place in concrete actions in cultural contexts with (implicit or explicit) rules.<sup>6</sup> **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 and Elgin’s concept of denotation therefore does not simply replace the traditional

<sup>6</sup> See Birk 2008, 25 ff. and 52, following Goodman and Wittgenstein. For the similarities between ‘example’ and ‘rule’, see Datson 2023.

Jan Georg Schneider: Rethinking Reference and Authorship

407 representational concept of reference, since denoting  
408 something does not simply mean establishing a relationship  
409 between the language and the world; rather, in the  
410 denotational act of reference we always adopt a perspective  
411 within the language games. In this sense, denotation and  
412 exemplification are equally selective and equally require a  
413 pragmatic anchoring (see Birk 2008: 52). The practice of  
414 denoting presupposes that of exemplifying, because only  
415 through the latter do communicative practices arise in which,  
416 in turn, denoting takes place, that is, referring in the classical  
417 sense. We can therefore conclude that denoting and  
418 exemplifying are mutually dependent on each other (see Birk  
419 2008: 79).

420 3.2 Exemplification as a motor in language acquisition and  
421 language use

422 “Learning from a sample involves learning to identify and  
423 project the labels that the sample exemplifies.” (Elgin 1983:  
424 93) This general formulation by Elgin can be seen as a guiding  
425 principle when it comes to describing processes of human  
426 language acquisition on a pragmatic basis. One can imagine  
427 human language acquisition as a continuous exemplification  
428 (and denotation) on different linguistic levels, which always  
429 implies a projection, i.e. a transfer to similar cases. Implicitly  
430 recognising or constructing such similarities is part of human  
431 language acquisition and intelligence.<sup>7</sup> With every utterance  
432 of a linguistic token, we as speakers of a language exemplify  
433 the **type**, a **syntactic** use, a **semantic** use and a **pragmatic**  
434 use (see Dürscheid & Schneider 2015, 189; Schneider 2015  
435 and 2025, following Stetter 2005, 82 ff.). This central idea is  
436 now illustrated by the following example sentences:

---

7 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).

## Jan Georg Schneider: Rethinking Reference and Authorship

437 Mum, can we **go** to the restaurant this evening? (A)

438 You will **go** crazy one day, if you **go** on like this. (B)

439 **Go** for it! (C)

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

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

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

473 3.3 Comparison to statistical processes in ChatGPT

474 LLMs are created roughly as follows: they are large neural  
475 networks that are trained with large amounts of text to  
476 predict the next word (or token) in a sentence context. To do  
477 this, two things are essential: firstly, a so-called word  
478 embedding, which assigns each word (token) a location in a  
479 virtual space of a previously arbitrarily defined dimension, so  
480 that ‘similar words’ are close to each other. The ‘similarity’  
481 here is that words can occur in the same syntactic  
482 environments, i.e. in the same co-texts. From an IT  
483 perspective, the similarity of words must be defined, because  
484 this method is used to determine how good the previous  
485 calculation is and how much the weights of the neural  
486 network need to be changed when training the language  
487 model.

488 If we want to speak of artificial ‘intelligence’ here at all,  
489 then this consists at most in the certain independence of the  
490 adaptation process in machine learning: the output is  
491 predicted from the input with ever greater probability (see  
492 Müller/Fürstenberg 2023: 329). The crucial idea of  
493 transformer networks is that the ‘neural’ networks learn  
494 which previous words of the input have particular predictive  
495 value for the prediction of the upcoming output and should  
496 therefore be given special weight. Like a less intelligent pupil,  
497 the system guesses answers to questions during training and  
498 receives success or failure feedback. This creates a game with  
499 the training data in recurrent loops until the ‘teacher’ is  
500 reasonably satisfied with the result. By training in recurrent  
501 loops, a sort of co-text sensitivity<sup>8</sup> is generated and  
502 continuously improved. For example, a language model that  
503 has been trained with current press texts could, without a  
504 recurrent loop, only ever react to the input *Donald* with the  
505 output *Trump*, because this word would presumably be the  
506 most frequent follow-up word in the training data set. With  
507 the recurrent loop, however, the processing of the input is  
508 also influenced by the preceding words, so that a more  
509 differentiated reaction can take place: If, for example,  
510 *Mickey, Mouse, Walt* or *Disney* are among the preceding

---

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



Jan Georg Schneider: Rethinking Reference and Authorship

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 semi-exemplifications, 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 no “language game competence” (“Sprachspielkompetenz” in the sense of the late Wittgenstein, see Schneider 2008).

### 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, textures<sup>9</sup> – 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

---

<sup>9</sup> For the explanation of this term see section 4.

Jan Georg Schneider: Rethinking Reference and Authorship

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 attempts<sup>10</sup> reveals precisely this lack of taste in the ‘reception’: 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 in the sense described above.

#### 4 Intelligible textures or intelligent texts? On judgement 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”

---

<sup>10</sup> 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.

Jan Georg Schneider: Rethinking Reference and Authorship

(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.

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 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 ff.) 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, 172 ff.), 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.

Jan Georg Schneider: Rethinking Reference and Authorship

Discussion Paper

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 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

Jan Georg Schneider: Rethinking Reference and Authorship

Discussion Paper

664 meaning of the expression *intelligible texture* too; in any case,  
665 it is highly compatible with it.

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

675 It is fundamentally wrong to downplay the capabilities of  
676 ChatGPT. Those who look small-mindedly for ‘errors’ in the  
677 answers will remain speechless when they are resolved.  
678 (Bermes/Dörpinghaus, FAZ No. 91, 19.04.2023, p. 4,  
679 translation: JGS)<sup>11</sup>

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

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

696 Asking in this way, the distinction between different text  
697 types becomes important: For which text types are LLMs  
698 useful as tools and how can they be applied to best effect  
699 then in each case? How much human intervention is needed  
700 to establish the reference sufficiently to ensure that cultural

---

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



## Jan Georg Schneider: Rethinking Reference and Authorship

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*. 2. Aufl. Oxford: Oxford University Press.
- Bermes, Christian/Andreas Dörpinghaus (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./Lisa Braden-Harder/Martin S. Chodorow/Bruce A. Kaplan/Karen Kukich/Chi Lu/Donald A. Rock/Susanne Wolff (2002): US 6,366,759 B1 [United States Patent, April 2, 2002: System and method for computer-based automatic essay scoring].
- Burstein, Jill C./Joel Tetreault/Nitrin Madnani (2013a): The E-rater Automated Essay Scoring System. In: Mark D. Shermis/Jill Burstein (ed.): *Handbook of Automated Essay Evaluation. Current Applications and New Directions*. London: Routledge, 55–67.
- Burstein, Jill/Joel Tetreault/Martin Chodorow/Daniel Blanchard/Slava Andreyev (2013b): Automated Evaluation of Discourse Coherence Quality in Essay Writing. In: Mark D. Shermis/Jill Burstein (ed.): *Handbook of Automated Essay Evaluation. Current Applications and New Directions*. London: Routledge, 267–280.
- Datson, Lorraine (2023): *Rules. A short history of what we live by*. Princeton & Oxford: Princeton University Press.
- Durt, Christoph/Tom Froese/Thomas Fuchs (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] <https://philsci-archive.pitt.edu/22744/> (last accessed: 18.03.2025)



Jan Georg Schneider: Rethinking Reference and Authorship

- 740 Elgin, Catherine Z. (1983): *With Reference to Reference*.  
741 Indianapolis: Hackett Publishing Co, Inc.
- 742 Fuchs, Thomas (2020): *Verteidigung des Menschen*.  
743 *Grundfragen einer verkörperten Anthropologie*. Frankfurt  
744 am Main: Suhrkamp
- 745 Goodman, Nelson (1976): *Languages of Art. An approach to a*  
746 *theory of symbols*. Indianapolis, Cambridge: Hackett  
747 Publishing Company.
- 748 Goodman, Nelson/Elgin, Catherine Z. (1988): *Reconceptions*  
749 *in Philosophy & Other Arts & Sciences*. Indianapolis,  
750 Cambridge: Hackett Publishing Company, 101–120.
- 751 Jäger, Ludwig (2002): Transkriptivität. Zur medialen Logik  
752 der kulturellen Semantik. In: Jäger, Ludwig/Georg  
753 Stanitzek (ed.): *Transkribieren. Medien | Lektüre*.  
754 München: Fink, 19–41.
- 755 Janich, Peter (2006): *Was ist Information? Kritik einer*  
756 *Legende*. Frankfurt am Main: Suhrkamp.
- 757 Janich, Peter (2015): *Handwerk und Mundwerk. Über das*  
758 *Herstellen von Wissen*. München: Beck.
- 759 Kant, Immanuel (1968): Kritik der reinen Vernunft. In Ders.:  
760 *Werke in 10 Bänden, ed. by Wilhelm Weischedel*.  
761 Darmstadt: Wissenschaftliche Buchgesellschaft. Cited as  
762 *KrV B*.
- 763 Keller, Rudi (2018): *Zeichentheorie. Eine pragmatische*  
764 *Theorie semiotischen Wissens*. 2., durchgesehene Auflage.  
765 Tübingen: Narr.
- 766 Meyer, Jennifer/Thorben Jansen/Johanna Fleckenstein/Stefan  
767 Keller/Olaf Köller (2023): Machine Learning im  
768 Bildungskontext: Evidenz für die Genauigkeit der  
769 automatisierten Beurteilung von Essays im Fach Englisch.  
770 In: *Zeitschrift für Pädagogische Psychologie* 37 (1), 203–  
771 2014. <https://doi.org/10.1024/1010-0652/a000296>
- 772 Müller, Hans-Georg/Fürstenberg, Maurice (2023): Der  
773 Sprachgebrauchsautomat. Die Funktionsweise von GPT  
774 und ihre Folgen für Germanistik und Deutschdidaktik. In:  
775 *Mitteilungen des Deutschen Germanistenverbandes* 70 (4),  
776 327–345.
- 777 Perelman, Les (2020): The BABEL Generator and E-Rater:  
778 21st Century Writing Constructs and Automated Essay  
779 Scoring (AES). In: *The Journal of Writing Assessment* 13 (1).  
780 <http://journalofwritingassessment.org/article.php?article=14>  
781 5 (last accessed: 18.03.2024).

Jan Georg Schneider: Rethinking Reference and Authorship

- 782 Rödel, Michael (2018): *Interpretationsaufsätze schreiben: Ein*  
783 *Handbuch*. Baltmannsweiler: Schneider Hohengehren.
- 784 Rupp, André A./Jodi M. Casabianca/Maleika Krüger/Stefan  
785 Keller/Olaf Köller (2019): Automated Essay Scoring at  
786 Scale: A Case Study in Switzerland and Germany. In:  
787 *TOEFL Research Report Series and ETS Research Report*  
788 *Series* ISSN 2330-8516, No. RR-19-12.
- 789 Saussure, Ferdinand de (1997): *Linguistik und Semiotik*.  
790 *Notizen aus dem Nachlaß. Texte Briefe und Dokumente*.  
791 *Gesammelt, übersetzt und eingeleitet von Johannes Fehr*.  
792 Frankfurt a.M.: Suhrkamp.
- 793 Schneider, Jan Georg (2008): *Spielräume der Medialität*.  
794 *Linguistische Gegenstandskonstitution aus*  
795 *medientheoretischer und pragmatischer Perspektive (=*  
796 *Linguistik – Impulse & Tendenzen 29)*. Berlin, New York:  
797 de Gruyter.
- 798 Schneider, Jan Georg (2015): Syntaktische Schemabildung –  
799 zeichentheoretisch betrachtet. In: Christa Dürscheid/Jan  
800 Georg Schneider (ed.): *Handbuch Satz, Äußerung, Schema*  
801 *(= HSW 4)*. Berlin & Boston: de Gruyter, 125–151.
- 802 Schneider, Jan Georg (2024): Intelligible Texturen. Welche  
803 Rolle kann ChatGPT bei der Aufsatzbewertung spielen? In:  
804 *VK:KIWA*. <https://zenodo.org/records/10877034> (last  
805 accessed: 18.03.2024)
- 806 Schneider, Jan Georg (2025): Sprachliche Typenbildung  
807 durch Exemplifikation. Über den Zusammenhang von  
808 System, Gebrauch und Normativität. In:  
809 Müller/Marcus/Martin Reisigl/Maria Becker/Michael  
810 Bender/Ekkehard Felder (ed.): *Diskursgrammatik*. Berlin,  
811 Boston: De Gruyter, 37–61.
- 812 Schneider, Jan Georg/Katharina A. Zweig (2022): Ohne Sinn.  
813 Zu Anspruch und Wirklichkeit automatisierter  
814 Aufsatzbewertung. In: Brommer, Sarah/Roth, Kersten  
815 Sven/Jürgen Spitzmüller (ed.): *Brückenschläge. Linguistik*  
816 *an den Schnittstellen*. Tübingen: Narr Francke Attempto  
817 (Tübinger Beiträge zur Linguistik, 583), 271-294.
- 818 Schneider, Jan Georg/Katharina A. Zweig (2023): Grade  
819 Prediction is not Grading: On the limits of the e-rater. In:  
820 Groß, Richard/Rita Jordan (ed.): *KI-Realitäten: Modelle,*  
821 *Praktiken und Topologien maschinellen Lernens*. Bielefeld:  
822 transcript (KI-Kritik / AI Critique, 5), 93-111.

Jan Georg Schneider: Rethinking Reference and Authorship

- 823 Stetter, Christian (1997): Schrift und Sprache. Frankfurt a.M.:  
824 Suhrkamp.  
825 Stetter, Christian (2005): System und Performanz.  
826 Symboltheoretische Grundlagen von Medientheorie und  
827 Sprachwissenschaft. Weilerswist: Velbrück Wissenschaft.  
828 Wittgenstein, Ludwig (1984): Philosophische Untersuchungen.  
829 In: *Wittgenstein/Ludwig. Tractatus logico-philosophicus.*  
830 *Tagebücher 1914–1916. Philosophische Untersuchungen.*  
831 Frankfurt am Main: Suhrkamp (Werkausgabe in 8 Bänden,  
832 1), 225-580.  
833 Wolfram, Stephen (2023): *What Is ChatGPT Doing ... and*  
834 *Why Does It Work?*. Wolfram Media, Inc. Auch online  
835 abrufbar unter:  
836 [https://writings.stephenwolfram.com/2023/02/what-is-](https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/)  
837 [chatgpt-doing-and-why-does-it-work/](https://writings.stephenwolfram.com/2023/02/what-is-chatgpt-doing-and-why-does-it-work/) (last accessed:  
838 18.03.2024).

## 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.