

CHATGPT IN GERMAN EXAMINATION LAW

Dr. h.c. Dennis Zielinski¹

ABSTRACT

This paper provides an in-depth analysis of the legal and ethical challenges presented by the integration of artificial intelligence, specifically ChatGPT, in German academic settings. It thoroughly examines how AI influences examination law, distinguishing between legal and illegal uses of AI in educational contexts. Additionally, the paper discusses the evolving nature of examination practices in response to the advancements in AI technology, emphasizing the need for legal frameworks to adapt to these technological changes.

INTRODUCTION

The present qualitative content analysis according to Mayring² focuses on the legal aspects of using artificial intelligence (AI) in German examination law. The text explores the challenges arising from the progress of AI, especially through the Chatbot "ChatGPT" and similar applications, for educational institutions. The qualitative analysis is divided into four main sections, named Use of Artificial Intelligence, Sanctions, Discovery and Proof of Attempted Deceptions and Outlook³:

I. Use of Artificial Intelligence

In the legal assessment of the use of Artificial Intelligence, both the factual aspect and the legal consequences need to be considered. Therefore, it is necessary to first examine the question of the legality of AI use (see section 1. below) and then consider the potential legal repercussions of unlawful application (see section 2. below).

1. Legality

What may be considered legally as one's "own" performance? Given that distinctions in prohibitions already exist between traditional on-site examinations, such as in-person written exams (see section aa) below), and remote examinations conducted at home (see section bb) below), it is advisable to separately examine each. For other forms of assessments (see section cc) below), the specifics will be crucial.

1 The author is a business lawyer and doctoral student at the Graduate School of IIC University of Technology

2 Mayring P./ Fenzl T. (2014): Qualitative Inhaltsanalyse, S. 234.

3 Mayring P./ Fenzl T. (2014): Qualitative Inhaltsanalyse, S. 235.

a) Examinations

The use of Artificial Intelligence requires the use of technical hardware. Generally, this hardware is not integrated into the examination process unless the exams are conducted on stationary computers provided by the educational institution. Carrying and using mobile phones already constitutes deceptive behavior⁴. Therefore, if exam takers use AI to support their performance in written exams, it does not necessarily present an independent legal issue related to Artificial Intelligence. If mobile phones, considered as a "modern cheat sheet"⁵, are prohibited, any form of usage is unlawful. Whether this occurs through consulting Wikipedia or utilizing AI to generate solutions for written exams, it is not legally distinguishable. ChatGPT is fundamentally not relevant to in-person exams and does not pose a problem for written exams⁶.

b) Home Assignments

In the context of home assignments, such as essays and their variations, as well as thesis papers, it is inevitable to address the question of what is lawful and what is unlawful. Concerning illegality, a distinction must be made between the performance (question of deception, see section aa below) and the assessment of the performance (see section bb below).

aa) Performance

In principle, the use of artificial (external) intelligence in the context of examinations is not legally different from the use of "non-artificial" external intelligence. The key question is not solely about the legal evaluation of using Artificial Intelligence to assess examination performance but rather whether a performance, not exclusively generated by the examinee, should have been produced by the examinee themselves. If this is the case, deceptive behavior is to be assumed⁷.

Deceptions in home assignments are widespread even without the "ChatGPT phenomenon." The COVID-19 pandemic has shed light on the issue, particularly with "open-book exams"⁸ resembling "small home assignments."⁹ Serious awareness of wrongdoing is rarely found among examinees, and this may be a modern societal trend:

4 OVG Bautzen, Urt. v. 05.11.2019 – Az. 3 B 388/18.

5 Hoos-Leistner, 2019, S. 86.

6 Robbers, Prüfungsrecht, 6. Aufl. 2014, Rdn. 210.

7 OVG Münster, Urt. v. 16.02.2021 – Az. 6 B 1868/20.

8 Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 230.

9 OVG Bautzen, Urt. v. 16.02.2022 – Az. 2 B 274/21.

What is prohibited is deemed "not harmful" if easily achievable, widespread, and the risk of detection is low. "Academic ghostwriting" has existed as long as there have been home assignments, and it has always been perceived as a somewhat marginal phenomenon, if it ever was. Plagiarism was and remains the next level. The key advantages for exam candidates in "copy & paste" compared to "external commissioning" are cost-effectiveness and sole authorship. Fortunately, plagiarism can be detected with reasonable accuracy through the routine use of professional software, and they are well demonstrable through the principles of prima facie evidence (see subsequent II. 2.). However, the situation is different with ChatGPT - the discoverability is (still) low (see subsequent II. 1.), and the proof seems legally challenging.

The "simplest" scenario is the complete creation of an assignment by an AI. The examinee receives a task, feeds it to the chatbot, and receives a completed assignment, which is then submitted as an examination performance.¹⁰ It is understood that AI in itself is not an authorized tool; there is no need for a separate regulation in an examination statute for this. Any aids that substitute the necessary independent effort of the examinee are prohibited, constituting deception.

What if the examinee takes on the "finalization" themselves and "merely" uses the AI in the "creative process"? The candidate would "seek assistance" by having ChatGPT generate a "text proposal," which they would then verify, elaborate, and finalize. Would this be deception in a legal sense? This will initially depend on what the required examination performance entails. If, as is regularly formulated in examination statutes, a home assignment is to be produced "independently," this means without external assistance. "External" refers to any processing that may play a role in the evaluation and has not been carried out by the examinee in person¹¹. Therefore, if for example, spelling and punctuation are relevant to the evaluation or may be relevant, the editorial involvement of software is prohibited. The required performance in a home assignment encompasses not only the end result but also the steps taken to achieve it. The "intermediate" use of AI is also prohibited.

¹⁰ Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 228f..

¹¹ OVG Münster, Nvwz 1995, 800, 803 – Erste juristische Prüfung.

Otherwise, even in home assignments, the use of sources is not only common and desirable but is constitutive. Initially, a legally relevant difference may not be discernible if the examinee gains orientation on their topic through reading a textbook or a AI-generated text commissioned by them. Permissible research is not limited to generally accessible or public sources. The deception in a legal sense is more of a technical nature:

The text submitted as an examination performance must not be a "verbatim adoption"¹², directly or indirectly, of the AI-generated work.

bb) Performance Evaluation

Even the non-prohibited use of ChatGPT is conceivable. Educational institutions may handle this differently from the considerations made here and allow the use of AI as a "component" of a home assignment. However, unless the (examination regulation) rule for the examination form "home assignment" is correspondingly adjusted, this would be a contradiction in adiecto: The external processing (by the AI) is elevated as an allowed part of the "independent" processing: The product is declared independent even though it is not, at least not exclusively. In a legal sense, however, there is nothing to support such a construction, and one cannot reasonably call it coherent.

More serious difficulties will arise elsewhere: How much "intermediate step" is allowed, and at what point is the text finalized by the examinee, as opposed to the "allowed AI draft," considered "independent"? Is it sufficient for the candidate to verify the coherence of the AI-generated text; is it enough if they "polish the formulations"? Are complete sentences, complete paragraphs allowed to be derived from the AI draft? How high can the "AI contribution" be, and should the assessment be quantitative, qualitative, or both? And would (excessively) high "AI contributions" be considered deception, or should they be taken into account in the performance evaluation? The theoretically clear boundaries between deception and poor performance would be blurred. Above all: A lawful assessment presupposes that the examiner knows what the examinee has accomplished. For the examinee's own performance to be measured at all, not only the finished home assignment but also the AI draft must be submitted. Those who "authorize" AI as assistance and conduct an evaluation without knowing the original AI text passages are deceiving themselves. The performance is simply not assessable¹³.

¹² Hoos-Leistner, 2019, S. 75.

¹³ Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 231.

Deception may be "mitigated" if the candidate explicitly declares to have "used" AI, possibly even specifying which chatbot and with what task command. However, the problem persists that an assessment oriented towards the performance rendered will only be possible if the "AI reference text" is known. A prerequisite for a lawful assessment is that the examiner correctly and completely acknowledges all performance-relevant achievements. Of course, even a legally flawed assessment is capable of acquiring legal force as an administrative act¹⁴.

Assessment capability would be established if the candidate quotes the passages in the "finished" work that they have adopted from an AI version. The absence of authorship by a natural person does not hinder quotability, as deceptions in examination law are not a copyright issue. Both direct text reproductions (marked with quotation marks) and indirect – paraphrased – text reproductions are generally subject to citation requirements.¹⁵ These are the mechanisms to combat plagiarism: One does not plagiarize if one cites correctly, and if large parts of work are copied and properly identified, the minimal personal contribution to the performance may be taken into consideration in the evaluation; it is not considered deception. Now, citation requires traceability: one must be able to follow a quote, and one must be able to verify it. Quotation is rendered futile if the referenced AI text is not available. Again, the capability for assessment is lost if the reference text is not provided.

c) Other Forms of Examination

Overall, only what does not belong to the required examination performance may be entrusted to AI. The decisive factor is what the examination regulations define as the required performance. In the case of oral examination performances, a distinction must be made: the development of a presentation by AI is impermissible, as is the creation of a "draft version" by AI. The same principles apply as with home assignments. Preparations that are not intertwined with the "actual" examination performance should be legally harmless. For example, role plays, which are indeed a form of examination, may be "rehearsed" using AI-generated texts, as they may be practiced with other externally sourced texts.

¹⁴ Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 815ff..

¹⁵ VGH Mannheim, Urt. v. 15.11.2019 – Az. 9 S 307/19..

2. Sanctions

The unlawful use of AI constitutes an attempt at deception. It is sufficient if the examination regulations – as is the case with practically every examination regulation – prohibit deception; there is no need for a specific prohibition regulation for AI. The "natural" consequence of an attempted deception is the failure of the examination. In addition, a diverse range of possible sanctions is conceivable, the availability of which depends in particular on the options provided by the examination regulations.

The frequent casualness with which examination regulations are formulated stands in stark contrast to the vulnerabilities that thoughtlessly installed examination regulations expose and to the possibilities that carefully crafted examination regulations offer.

Not every attempted deception must lead to failure¹⁶. The quantitative proportion of the work and its qualitative significance are crucial. Minor irregularities can be "overlooked," and if something is relatively insignificant, a non-formal admonition may suffice. Specific regulations to refrain from sanctions for reasons of proportionality are not necessary in the examination regulations. Conversely, the question will regularly arise whether a "simple failure" should be sufficient. When considering the possible scenario that a text submitted as an examination performance was entirely created by AI, its moral reprehensibility is at least identical to "complete plagiarism," i.e., the exclusive or almost exclusive unauthorized use of non-original texts¹⁷. These cases regularly meet the criteria of a "particularly serious" attempted deception, with the consequence that a final failure of the examination is also a possible sanction. This results in the loss of the examination entitlement, even if it was not the final attempt. This constellation is evidently closely related to cases of "academic ghostwriting," which are also regularly considered particularly serious attempted deceptions. Of course, for this sanction to be imposed, there must be a corresponding regulation in the examination regulations. When determining sanctions, general preventive aspects may also play a role. The potential deterrence of imitators may be taken into account.

¹⁶ Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 241.

¹⁷ OVG Hamburg, Urt. v. 19.11.2013 – Az. 3 Bs 274/13.

The seriousness of the phenomenon of multiple and serious deceptions in academic work is not adequately addressed by the fact that the (Nordrhein-Westfalen, Germany) legislator, in § 66 IV 2 NRWG, places a "time-lapse protection shield" over the "perpetrator": The withdrawal of a degree award based on § 48 I, III VwVfG (unlawful administrative act; obtained through deception) is only allowed within five years¹⁸.

The deceitful individual is then "exempt," and this after a relatively short time, and particularly irrespective of technological developments that might one day enable the legally secure proof of the use of AI. Universities have the option to incorporate independent regulations for the revocation of degree awards into their examination regulations, which (presumably) are not affected by the statutory time-lapse regulation¹⁹.

Nestled in efficacy between "not passed" and "loss of examination entitlement" is the possibility of identifying attempted deceptions in the performance overview belonging to the study completion documents. So where there would normally be only a "np" (not passed), there could be an "np (AT)" (attempted deception), with corresponding explanations in the legend.

In cases where university laws entrust the design of certificate documents to the universities, there is nothing against implementing such regulations in examination regulations. Of course, such regulations are then required, as certificate contents are subject to substantive legislative reservation²⁰.

The level of sanctions is not limited to the actual examination legal relationship. Some university laws provide for further possible sanctions for attempted deceptions, including the deregistration of students (see, for example, §§ 42 III No. 5 HmbHG, 63 V 6 NRWG, 69 IV RhPfHochSchG), sometimes with the possibility of subsequent time-related bans (§ 30 IV 3 LSAHSG), and monetary fines based on administrative offense provisions are also possible (§§ 63 V 2, S. 3 NRWG, 51 VIII 2, 3 MVLHG). To the extent provided for by law or regulation, affidavits may be required and taken (§ 27 I 1 VwVfG). Corresponding regulations related to the independence of examination performances are found, for example, in § 21 VI BbgHG, § 63 V 1 NRWG, 51 VII 1 MVLHG. The false submission of an affidavit is a criminal offense under § 156 StGB.

18 VG Köln, Urt. v. 15.12.2005 – Az. 6 K 6285/04.

19 OVG Hamburg, NvwZ-RR 2006, 410, 411 – Vorbereitungsdienst.

20 Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 440.

II. Discovery and Proof of Attempted Deceptions

If the use of AI eludes detection, discussions about permissibility and prohibition become irrelevant in practice. So how can the examination institution discover (below 1.) and prove (below 2.) an unauthorized use of AI?

1. Discovery

Preceding the question of proof is the discovery that triggers the need for proof: the "initial suspicion," if you will. If there are no procedural routines for this, the regular case will be that the attempted deception is not discovered, and then the question of proof only arises with "random finds," especially those where the wording or the performance conspicuously deviates from the otherwise known performance profile of the candidate. These are situations that can be expected mainly at smaller educational institutions where candidates and examiners know each other personally²¹.

It is recommended that educational institutions equip themselves with software that can at least recognize the probability of AI usage. The routine use of suitable software would focus attention on the phenomenon as required.

If AI texts are not considered "minor offenses"²² from the perspective of the educational institution, it cannot afford to avoid systematically addressing the issue. Furthermore, in individual cases, the candidate may argue that factually recurring deceptive acts can never lead to final failure. Negligence in detection efforts could thus impact the effectiveness of the sanction system²³.

Another aspect could assume legal significance: If the matter is "allowed to run its course" and action is only taken in individual cases of random finds, questions about the validity of the examination system may arise. If the examination procedure related to a cohort is characterized by a significantly high number of attempted deceptions, this can affect the overall evaluability²⁴.

21 Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 297.

22 Pappmeyer, 2017, S. 44.

23 Vgl. BVerwG, NJW 2018, 1896, Rdn. 7 – Aufstiegsprüfung.

24 Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 806.

Someone who has not passed (and has not failed due to an attempted deception but due to poor performance) might argue that the high number of tolerated deceptions knowingly accepted by the educational institution has shifted the evaluation standard to their disadvantage. If the cause for the unlawful favoring of individual examinees lies within the sphere of the examination institution, the assumption of procedural error is sufficient if an influence on the evaluation standard cannot be ruled out with sufficient certainty²⁵.

2. Proof

The question arises as to whether and to what extent a "discovery software" can simultaneously be attributed with a legally valid proof function. So far, plagiarism software is limited in its mechanism to providing assistance in manually checking for plagiarism by identifying textual similarities. It serves as the first stage in a two-step verification process, with the second stage providing evidence of deception. In AI verification processes, this "second stage" must be eliminated because there are simply no reference texts known to the plagiarism software. The insidious nature of text created using AI is that it is original; it is not simply copied or paraphrased. In plagiarism cases, a presumption of deception can be derived from textual similarities. This possibility does not exist with ChatGPT, and proof must be established through the software itself if the reference text is not known. Whether this will be sufficient in doubtful cases remains an open question today. Does the "overwhelming probability," as in other attempts at deception, suffice to shift the burden of proof against the examinee?

Is it enough if a "discovery software" has a reliable and verified "hit rate" for AI-generated texts? The acid test is the legal proceedings, and a court would have to "believe" the software if it wanted to consider the attempt at deception as proven. This will, in any case, require at least a solid reliability proof for the software²⁶.

Can one infer a possible attempt at deception from the "error-free" nature of a text? It is expected that AI-generated texts will not contain inaccuracies in orthography, punctuation, and grammar—a phenomenon that is likely to be the exception in "human-made" texts. In principle, an attempt at deception cannot be "proven" by arguing that a performance is particularly good (namely, linguistically correct).

25 Stelkens/Bonk/Sachs/Kallerhoff/Fellenberg, VwVfG § 26 Rdn. 43.

26 Robbers, Prüfungsrecht, 6. Aufl. 2014, Rdn. 188.

However, if the comparison with the examinee's previous written works shows a significant deviation, this may be taken into account within the presumption of deception.

In examination procedures, typically administrative procedures, it is worth looking into the Administrative Procedure Act. This provides opportunities for the authority to establish facts. The VwVfG generally applies alongside the procedural regulations of examination ordinances, both in state universities and in examination proceedings of non-state universities acting as licensees²⁷.

Of particular interest here are the provisions for administrative fact-finding in §§ 24, 26, and 27 VwVfG²⁸. These are not affected by the partial exclusion for examination law in § 2 III No. 2 VwVfG, so it does not need to be clarified whether and to what extent this exclusion itself is subject to restrictions. The authority investigates the relevant facts *ex officio* (principle of investigation; § 24 I 1 VwVfG and uses the evidence deemed necessary at its discretion (§ 26 I 1 VwVfG)²⁹. This includes, among other things, the hearing of parties involved (§ 26 I 2 No. 2 VwVfG)³⁰.

The available regulatory framework does allow the examination institution (including in the form of the examiner) to question the examinee to determine whether he could have deceived—here, by using AI. If the candidate stumbles when answering the question about the origin of his examination performance and cannot provide a coherent answer, this may establish the appearance of deception, with the consequence of a shift in the burden of proof. Of course, limitations imposed by examination procedure law must be considered: an "additional oral examination" not provided for alongside a written assignment is prohibited³¹.

The examination of the examinee must be separate from the evaluation level, and "unconvincing" statements by the examinee should not be considered in the evaluation. To preempt this, it is recommended to first assess the work and document the evaluation before conducting any potential questioning of the examinee (specifically, to determine a possible attempt at deception)³².

27 Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 176.

28 Stelkens/Bonk/Sachs/Kallerhoff/Fellenberg, VwVfG § 26 Rdn. 43.

29 Pappmeyer, 2017, S. 25.

30 OVG Münster, Urt. v. 10.12.2015 – Az. 19 A 254/13.

31 Fischer/Jeremias/Dietrich, Prüfungsrecht, 8. Aufl. 2022, Rdn. 653.

32 Schober, 2010, S. 22.

III. Outlook

The comparison of ChatGPT with the introduction of the pocket calculator has become commonplace. Unlike what may have been predicted with the proliferation of calculators, mathematics is still examined today; it is just examined differently than before. The routines have adapted to technology, as they always do, albeit sometimes with a delayed timeframe. The same will happen with AI. Why not integrate ChatGPT into the process, for example, by instructing the examinee to generate a text draft on a topic through ChatGPT and then finalize an "original" text from it?

This would be a "new" examination performance and would need to be regulated as such. In the future, there will likely be fewer exams where the performance consists solely of preparing a written assignment. The difficulties in detecting and proving attempted deceptions are too great. Written assignments are expected to be accompanied by a supplementary oral examination component, at least for assurance that the candidate has genuinely engaged with the topic. Regular consultations between the candidate and the examiner during the development phase of a written assignment can also be a means of promoting independence. It is also not ruled out to require the candidate to submit regular "interim versions" of their work to observe the creation process. All of this necessitates adjustments to the examination regulations, and such adjustments are, in any case, required concerning the sanctioning toolkit (see above I. 2.).

REFERENCES

1. Fischer/Jeremias/Dietrich (2022). Prüfungsrecht, 8. Aufl., Springer Fachmedien Wiesbaden
2. Hoos-Leistner, H. (2019). Fragen zum Prüfungsrecht. Springer Fachmedien Wiesbaden
Juris – das Rechtsportal
3. Mayring P./ Fenzl T. (2014): Qualitative Inhaltsanalyse Neue juristische Wochenschrift – NJW
4. Neue Zeitschrift für Verwaltungsrecht – NvWZ
5. Pappmeyer, K. (2017). Coronapandemie und die universitären Prüfungen. Eine Untersuchung von Täuschungsversuchen. Springer Fachmedien Wiesbaden
6. Robbers (2014) Prüfungsrecht, 6. Aufl., Springer Fachmedien Wiesbaden
7. Schober, O. (2010). Prüfungsrecht. Schlüssel zum Plagiat. Springer Fachmedien Wiesbaden
8. Stelkens/Bonk/Sachs/Kallerhoff/Fellenberg (2022) VwVfG