

Transparenzpflicht für AI-Tools? Bitte nicht!

Alexander Ecker

Institute of Computer Science & Campus Institute Data Science • University of Göttingen
Max Planck Institute for Dynamics and Self-Organization • Göttingen

Meine drei Punkte heute

1 Eine Transparenzpflicht für KI-Tools ist sinnlos und gefährlich

2 Der Einsatz von KI-Tools im Studium muss auf freiwilliger Basis reguliert werden

3 Hausarbeiten als Prüfungsform sind (vermutlich) hinfällig

Wo wir gerade stehen

AI-Tools wie ChatGPT, Google Bard etc. entwickeln sich rasend schnell und werden besser

Zugrunde liegende Technologie (LLMs) wird in naher Zukunft praktisch überall eingebaut sein und uns unterstützen

Bing nutzt bereits ChatGPT

Google Bard kann Emails und Docs nutzen, auswerten, zusammenfassen

Integration in Textverarbeitungssoftware Frage der Zeit

...

Es folgt: Wir können die generelle Nutzung überhaupt nicht verbieten!

Sonst müssten wir Suchmaschinen (Google, Bing) und Textverarbeitung (Word) verbieten

Die Diskussion kann also nur darum gehen:

Wie und wann nutzen unsere Studierenden die Tools?

Wie schaffen wir es, dass Studierende trotzdem etwas lernen? (Bsp. Taschenrechner)

Muss die Nutzung von AI-Tools transparent gemacht werden?

Umgang mit KI-Modellen wie Handlungsempfehlungen für

Version 1.0 vom 12.04.2023, Beschluss des Präsidiums vom

Allgemeine Einschätzung

ChatGPT ist eine besonders populäre Version eines KI-basierten Dienstes, der unter <https://chat.openai.com/chat> verfügbar ist. Für die Zukunft ist dies besonders bedeutsam, dass der Dienst im Frage-Antwort-Dialog hochwertiger Art und Weise generieren kann. Wir nutzen ChatGPT, die sich rasch entwickelnde Gruppe von (Text- und Bild-) Generatoren, die die Zukunft Grundlage vieler unterschiedlicher Programme und Dienste sein werden. Man kann ChatGPT z.B. wie folgt fragen, sich selbst vorzustellen

W Was ist ChatGPT, allgemeinverständlich erklärt?

 ChatGPT ist ein KI-gestütztes Textmodell, das von OpenAI entwickelt wurde. "KI" steht für Künstliche Intelligenz, also Computerprogramme, die menschenähnliche Fähigkeiten, wie das Verstehen und Verarbeiten von Sprache, besitzen. ChatGPT steht für "Conversational Generative Pre-trained Transformer" und ist ein Modell, das darauf spezialisiert ist, menschenähnliche Unterhaltungen und Texte zu generieren.

In **Prüfungen ohne Aufsicht** sollte Folgendes gelten:

1. ChatGPT sollte in Prüfungen ohne Aufsicht grundsätzlich ein zulässiges Hilfsmittel sein, auch weil nicht kontrolliert werden kann, ob eine Nicht-Zulassung eingehalten würde.
2. Die Zulässigkeit von ChatGPT als Hilfsmittel endet da, wo Teile der abgegebenen Leistung **von ChatGPT stammen und nicht als solche transparent** gemacht werden. Dies ist gleichzusetzen mit einem Täuschungsversuch.
3. **Der Einsatz von ChatGPT muss also vollständig transparent erfolgen. Eine explizite Kennzeichnung von mit ChatGPT erzeugten Passagen ist erforderlich.** Studierende sollten zudem erklären, in welcher Art und Weise ChatGPT für die Erstellung der Prüfungsleistung genutzt wurde. Eine Beispielvorlage befindet sich im Anhang.
4. Eine solche vollständig transparente Nutzung von ChatGPT soll **weder positive noch negative Auswirkungen auf die Bewertung** der Prüfungsleistung haben.
5. Prüfende sollten Fragestellungen so anpassen, dass die Eigenständigkeit einer Prüfungsleistung auch dann nachvollzogen werden kann. [...]



Meine Universität plant, die Nutzung von AI-Tools wie ChatGPT generell zu erlauben, möchte aber eine Transparenzpflicht einführen. Konkret steht folgende Regelung im Raum:

1. ChatGPT sollte in Prüfungen ohne Aufsicht grundsätzlich ein zulässiges Hilfsmittel sein, auch weil nicht kontrolliert werden kann, ob eine Nicht-Zulassung eingehalten würde.
2. Die Zulässigkeit von ChatGPT als Hilfsmittel endet da, wo Teile der abgegebenen Leistung von ChatGPT stammen und nicht als solche transparent gemacht werden. Dies ist gleichzusetzen mit einem Täuschungsversuch.
3. Der Einsatz von ChatGPT muss also vollständig transparent erfolgen. Eine explizite Kennzeichnung von mit ChatGPT erzeugten Passagen ist erforderlich. Studierende sollten zudem erklären, in welcher Art und Weise ChatGPT für die Erstellung der Prüfungsleistung genutzt wurde.
4. Eine solche vollständig transparente Nutzung von ChatGPT soll weder positive noch negative Auswirkungen auf die Bewertung der Prüfungsleistung haben.

Wörtlich zitiert aus
Handlungsempfehlungen

Welche Probleme könnten sich aus dieser Transparenzpflicht ergeben?



Eine Transparenzpflicht bei der Nutzung von AI-Tools wie ChatGPT in universitären Kontexten kann sowohl Vorteile als auch Probleme mit sich bringen. Im Folgenden sind einige potenzielle Probleme dargestellt, die sich aus einer solchen Regelung ergeben könnten:

1. **Schwierigkeit bei der Abgrenzung:** Es kann herausfordernd sein, festzulegen, was genau als "von ChatGPT stammend" gilt. Wenn ein Student beispielsweise eine allgemeine Frage zu einem Thema stellt und die Antwort dann in eigenen Worten umschreibt, ist dies dann eine "von ChatGPT stammende" Passage?
2. **Mögliche Übernutzung:** Wenn ChatGPT als zulässiges Hilfsmittel in Prüfungen ohne Aufsicht erlaubt ist, könnten einige Studierende sich zu stark auf das Tool verlassen, was das eigentliche Lernen und Verstehen des Stoffs beeinträchtigen könnte.
3. **Ungleichheit:** Nicht alle Studierenden haben gleichermaßen Zugang zu oder Erfahrung mit fortschrittlichen AI-Tools. Dies könnte zu einem ungleichen Spielfeld führen, auf dem einige Studierende einen potenziellen Vorteil gegenüber anderen haben.
4. **Datenschutz:** Je nachdem, wie die Transparenzpflicht umgesetzt wird, könnten Bedenken hinsichtlich des Datenschutzes und der Vertraulichkeit der Studierendeninformationen aufkommen, insbesondere wenn sie detaillierte Protokolle darüber vorlegen müssen, wie sie ChatGPT verwendet haben.
5. **Zeitaufwand:** Die Erfordernis, jede Verwendung von ChatGPT detailliert zu dokumentieren, könnte für Studierende zusätzlichen Aufwand bedeuten und sie von der eigentlichen Aufgabe oder dem eigentlichen Lernziel ablenken.
6. **Bewertungsschwierigkeiten:** Dozenten könnten Schwierigkeiten haben, die Qualität der Arbeit eines Studierenden zu bewerten, wenn sie nicht genau wissen, welchen Anteil ChatGPT an der Arbeit hatte.
7. **Integrität:** Selbst mit einer Transparenzpflicht könnten einige Studierende versucht sein, die genaue Nutzung von ChatGPT nicht offenzulegen, was zu weiteren Täuschungsproblemen führen könnte.
8. **Potential für Overhead:** Für die Universitätsverwaltung könnte es herausfordernd sein, die Einhaltung solcher Regeln zu überwachen und zu verifizieren.
9. **Komplikationen bei Gruppenarbeiten:** Wenn in Gruppenarbeiten AI-Tools verwendet werden, könnte es schwierig sein, festzustellen, welcher Studierende welchen Beitrag geleistet hat und inwieweit ChatGPT involviert war.
10. **Definition von "Nutzung":** Es könnte schwierig sein, genau zu definieren, was als "Nutzung von ChatGPT" gilt. Beispielsweise könnte ein Student ChatGPT verwenden, um ein Konzept besser zu verstehen, ohne direkt Inhalte aus dem Tool in seine Arbeit einzufügen.



Die geplante Regelung zur Zulassung von ChatGPT und anderen AI-Tools in Prüfungen zielt darauf ab, Transparenz und Fairness zu gewährleisten. Während dies durchaus lobenswerte Ziele sind, könnten sich dennoch einige Probleme und Herausforderungen ergeben:

1. **Definition von "Stammen":** Die Formulierung "wo Teile der abgegebenen Leistung von ChatGPT stammen" ist möglicherweise zu vage. Was genau bedeutet "stammen"? Reicht es aus, wenn eine Idee oder ein Konzept von ChatGPT inspiriert wurde, oder muss ein Text wortwörtlich von der AI übernommen werden, um als solcher zu gelten?
2. **Bewertung:** Selbst wenn die Verwendung von ChatGPT keinen direkten Einfluss auf die Bewertung haben soll, **könnte es dennoch subjektive Wahrnehmungen bei Prüfern geben**. Einige könnten die Fähigkeit der Studierenden, Informationen effizient zu nutzen und zu synthetisieren, schätzen, während andere der Meinung sein könnten, dass eine zu starke Abhängigkeit von AI-Tools die individuellen Fähigkeiten des Studierenden verdeckt.
3. **Uneinheitliche Nutzung:** Während einige Studierende die Vorschriften befolgen und die Verwendung von ChatGPT transparent machen, könnten andere versuchen, ihre Interaktionen mit der AI zu verbergen, insbesondere wenn sie denken, dass dies ihre Bewertung beeinflussen könnte.

4. **Arbeitsaufwand:** Das Anfordern einer detaillierten Erläuterung darüber, wie ChatGPT verwendet wurde, kann für Studierende zusätzlichen Arbeitsaufwand bedeuten und sie davon abhalten, sich auf den eigentlichen Prüfungsinhalt zu konzentrieren.
5. **Kontext und Tiefgang:** Eine Frage, die sich stellt, ist, inwieweit AI-Tools wie ChatGPT den Studierenden wirklich helfen. Während sie schnelle Antworten liefern können, fehlt ihnen oft der Kontext und der Tiefgang, den ein tiefgehendes Studium eines Themas bieten kann.
6. **Förderung von Abhängigkeit:** Eine potenzielle Gefahr ist, dass Studierende zu abhängig von solchen Tools werden und ihre eigenen kritischen Denkfähigkeiten und Forschungskompetenzen vernachlässigen.
7. **Technische Herausforderungen:** Einige Studierende haben möglicherweise keinen einfachen Zugang zu AI-Tools oder sind nicht so vertraut mit ihrer Verwendung. Dies könnte zu technischen oder Bildungsungleichheiten führen.
8. **Authentizität:** Wenn Teile einer Antwort von ChatGPT stammen, könnte es schwierig werden, die Originalität und Authentizität der gesamten Arbeit zu bewerten.

Während die Einführung von Transparenzpflichten sinnvoll sein kann, ist es wichtig, die potenziellen Herausforderungen und Unklarheiten zu berücksichtigen und gegebenenfalls Anpassungen vorzunehmen, um die Fairness und Effektivität des Prüfungsprozesses zu gewährleisten.



Die folgenden Probleme könnten sich aus der Transparenzpflicht für die Nutzung von ChatGPT ergeben:

- **Schwierigkeiten bei der Kennzeichnung:** Es kann schwierig sein, mit ChatGPT erzeugte Texte zu identifizieren. ChatGPT kann Texte generieren, die von menschlichen Texten nicht zu unterscheiden sind. Dies kann dazu führen, dass Studierende Passagen, die sie mit ChatGPT generiert haben, nicht als solche kennzeichnen.
- **Möglichkeit des Missbrauchs:** Die Transparenzpflicht könnte von Studierenden missbraucht werden, um Täuschungsversuche zu verschleiern. Studierende könnten Passagen, die sie mit ChatGPT generiert haben, als selbstgeschrieben kennzeichnen.
- **Erhöhte Prüfungsvorbereitungszeit:** Die Studierenden müssen sich zusätzlich mit den Anforderungen der Transparenzpflicht auseinandersetzen. Dies kann zu einer erhöhten Prüfungsvorbereitungszeit führen.

Eher nicht so
nützlich

Im Folgenden werden diese Probleme näher erläutert:

Schwierigkeiten bei der Kennzeichnung

ChatGPT kann Texte generieren, die von menschlichen Texten nicht zu unterscheiden sind. Dies liegt daran, dass ChatGPT auf einem riesigen Datensatz aus Text und Code trainiert ist. Dieser Datensatz enthält auch Texte, die von Menschen geschrieben wurden. ChatGPT kann diese Texte lernen und daraus neue Texte generieren, die dem Stil menschlicher Texte entsprechen.

...

Problem 1: Abgrenzung KI vs. Nicht-KI

Frage: Welche Tools fallen unter eine Regelung?

Nur ChatGPT?

- Zu eng! Bard, Llama, ...

Positivliste

- Zu unflexibel, langsam, aufwändig, komplex

Alle KI-Tools?

- Was ist überhaupt KI? Weder Kognitionswissenschaft noch Informatik verfügen über eine allgemein anerkannte Definition von „Intelligenz“
- Für Nutzende nicht ersichtlich, ob KI am Werk ist



Die Frage, welche Systeme
kennzeichnungspflichtig
sind, ist quasi nicht zu
beantworten

Problem 2: Abgrenzung von Einsatzformen

Welche Einsatzformen fallen unter eine Regelung?

Beispiel: Hausarbeit in Philosophie

Thema:

Der freie Wille bei Immanuel Kant: Zwischen praktischer Vernunft und Kausalität

Fragestellung:

Inwiefern lässt sich der freie Wille in Kants Philosophie als Vermittler zwischen der phänomenalen und der noumenalen Welt verstehen, und welche Rolle spielt die praktische Vernunft in diesem Kontext?

Entwurf sprachlich „glätten“



ENTWURF (ICH)

Die Debatte um den freien Willen ist ein zentrales Thema in der Philosophiegeschichte, und besonders die Sichtweise von Immanuel Kant bekommt dabei viel Aufmerksamkeit. Einer der wichtigsten Aspekte, die man bei Kants Meinung zum freien Willen beachten sollte, ist der Dualismus zwischen der phänomenalen Welt der Erscheinungen und der noumenalen Welt des Dings an sich. Wenn Kant über den freien Willen spricht, findet er diesen nicht in dem, was wir mit unseren Sinnen wahrnehmen können, sondern im Bereich des Intelligiblen. Ein weiterer Punkt ...

ÜBERARBEITETE VERSION (CHATGPT)

Die Debatte um den freien Willen ist ein zentrales Thema in der Philosophiegeschichte, wobei die Position Immanuel Kants eine besondere Aufmerksamkeit verdient. Ein Schlüsselement in Kants Ansatz ist der Dualismus zwischen der phänomenalen Welt der Erscheinungen und der noumenalen Welt des Dings an sich. Interessanterweise verortet Kant den freien Willen nicht im empirisch Wahrnehmbaren, sondern im Bereich des Intelligiblen. Eine bemerkenswerte Facette in Kants Theorie ...

Stichpunkte zu Fließtext



OUTLINE (ICH)

Hauptthema: Debatte um freien Willen; Fokus auf Kants Position.

Zentraler Gedanke: Dualismus zwischen phänomenaler und noumenaler Welt.

Freier Wille: Lokalisierung? Nicht im Empirischen, sondern im Intelligiblen.

Verbindung: Freier Wille und praktische Vernunft.

Hauptargument: Rolle der praktischen Vernunft im Kontext freien Willens.

ABSTRACT (CHATGPT)

Die Debatte um den freien Willen ist ein zentrales Thema in der Philosophiegeschichte, wobei die Position Immanuel Kants eine besondere Aufmerksamkeit verdient. Ein Schlüsselement in Kants Ansatz ist der Dualismus zwischen der phänomenalen Welt der Erscheinungen und der noumenalen Welt des Dings an sich. Interessanterweise verortet Kant den freien Willen nicht im empirisch Wahrnehmbaren, sondern im Bereich des Intelligiblen. Eine bemerkenswerte Facette in Kants Theorie ist die enge Verknüpfung zwischen dem freien Willen und der praktischen Vernunft. Das Hauptargument dieser Arbeit wird sein, die entscheidende Rolle der praktischen Vernunft im Kontext des freien Willens hervorzuheben und zu untersuchen, wie sie sowohl zur Ethik als auch zur Metaphysik Kants beiträgt.

OK? Nicht OK?
Kommt darauf an!

Gliederung (ChatGPT)

1. Einleitung (2 Seiten)

- 1.1 Hintergrund und Bedeutung des Themas
- 1.2 Zielsetzung der Arbeit und zentrale Fragestellung
- 1.3 Aufbau und Struktur der Arbeit

2. Historischer Kontext und Grundlagen (3 Seiten)

- 2.1 Philosophiegeschichtlicher Hintergrund des freien Willens
- 2.2 Kants Stellung in der Debatte und kurze Einführung in seine Philosophie

3. Kants Dualismus: Phänomenale und Noumenale Welt (5 Seiten)

- 3.1 Definition und Charakterisierung der phänomenalen Welt
- 3.2 Definition und Charakterisierung der noumenalen Welt
- 3.3 Die Grenzen menschlicher Erkenntnis und ihre Auswirkungen auf den Dualismus

4. Der freie Wille in Kants Philosophie (5 Seiten)

- 4.1 Definition und Bedeutung des freien Willens bei Kant
- 4.2 Der freie Wille in Bezug auf den Dualismus: Wie er die phänomenale und noumenale Welt vermittelt
- 4.3 Kausalität und ihre Verbindung zum freien Willen

5. Die Rolle der praktischen Vernunft (5 Seiten)

- 5.1 Definition und Charakterisierung der praktischen Vernunft
- 5.2 Die Verknüpfung von freiem Willen und praktischer Vernunft
- 5.3 Praktische Vernunft als Vermittler zwischen Phänomenalität und Noumenalität

6. Implikationen und Kritik (3 Seiten)

- 6.1 Konsequenzen von Kants Ansichten für Ethik und Metaphysik
- 6.2 Gängige Kritikpunkte und Einwände gegen Kants Theorie des freien Willens und der praktischen Vernunft
- 6.3 Relevanz von Kants Ansichten in der heutigen Debatte

7. Schlussfolgerung (1 Seite)

- 7.1 Zusammenfassung der zentralen Erkenntnisse
- 7.2 Ausblick und mögliche weitere Forschungsrichtungen

Literaturverzeichnis

Problem 2: Abgrenzung von Einsatzformen

Welche Einsatzformen fallen unter eine Regelung?

- Rechtschreib-/Grammatikkorrektur
- Sprachliche Verbesserung eines Entwurfes
- Ausformulierung einer Stichpunktliste von Gedanken
- Erzeugen einer kurzen Definition eines Fachbegriffs
- Erzeugen eines gesamten Absatzes mit „Standardwissen“
- Erzeugen der Gliederung einer Arbeit
- Synthese einer Zusammenfassung für eine ansonsten fertige Arbeit



Transparenz entweder nutzlos (weil unspezifisch) oder unverhältnismäßig viel Arbeit

Was folgt?

Nahezu beliebiger Einsatz von Tools genau dann sinnvoll und nützlich wenn Nutzende in der Lage sind den Output zu bewerten, verifizieren und ggf. verbessern

- Für Studierende zu Beginn des Studiums nicht gegeben
- Im Verlauf des Studiums jedoch (hoffentlich) zunehmend

→ Erlaubter Einsatz muss vom Lernziel abhängen



Transparenz löst das Problem nicht!
(vgl. Taschenrechner)

Problem: Einsatz kann nur überprüft werden, wenn Studierende vor Ort sind

Vertrauen
(honor code)

oder

Hausarbeit als *Prüfungsform* längerfristig ungeeignet

- Als *Übung* möglicherweise weiterhin sinnvoll und wichtig
- Vgl. Naturwissenschaften
 - Übungsblätter *Prüfungsvorleistung*
 - Note durch schriftl. oder mündl. Prüfung

Warum eine Transparenzpflicht gefährlich ist

Ohne weitere Spezifikation, wie Transparenz auszusehen hat, große Unsicherheit auf Seiten der Studierenden

Meine Vorhersage: Eine solche Regelung wird in der Praxis nicht umsetzbar sein und von der Realität überholt

→ Rechtlich problematisch, wenn bspw. Abschlussarbeiten (Dissertationen?) in Zukunft überprüft werden

Es folgt: genau das Gegenteil!

→ i.A. **keine** Transparenzpflicht

→ Prüfungsformen müssen angepasst werden

Mögliche Lösungsansätze

Hausarbeit als „Übung“

- Studierende nutzen Tools ihren eigenen Fähigkeiten und Lernzielen entsprechend freiwillig (nicht)

Note durch Prüfung

- Mündliche Prüfung (evtl. Gruppenprüfung)
- Per ChatGPT erzeugte Fragen zur abgegebenen Arbeit müssen vor Ort schriftlich beantwortet werden
- ...

Meine drei Punkte heute

1 Eine Transparenzpflicht für KI-Tools ist sinnlos und gefährlich

2 Der Einsatz von KI-Tools im Studium muss auf freiwilliger Basis reguliert werden

3 Hausarbeiten als *Prüfungsform* sind (vermutlich) hinfällig