

Hausarbeit
„Einführung in die Sozioinformatik“
WS 2023/24

Eine technikinduzierte sozioinformatische Analyse
des auf GPT-4 basierenden Feedbacktools fiete.ai

Jannik Brack

23.02.2024

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Hausarbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

Statutory declaration

I hereby declare that I have authored this essay independently, that I have not used any other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

Kaiserslautern, 23.02.2024

Jannik Brack

Inhaltsverzeichnis

1	Einleitung	1
2	Technikbeschreibung	2
2.1	fiete.ai	2
2.2	ChatGPT	3
2.3	fiete.ai-Prompting	4
2.4	Grenzen und Möglichkeiten	5
3	Akteure und deren Motivationen	6
3.1	Akteure	6
3.2	Motivationen	7
3.3	Umwelt	9
4	Erstellung der Wirkungsgefüge	10
4.1	Systemvariablen	10
4.2	Kanten	12
4.3	Abbildungen der Wirkungsgefüge	14
5	Analyse der Wirkungsgefüge	16
6	Zweite Gedanken	17
7	Fazit	18
8	Anhang	18

1 Einleitung

Mit der Veröffentlichung von ChatGPT-4 im März 2023 baute das Softwareunternehmen OpenAI seine Stellung in der Branche der Entwicklung und Erforschung künstlicher Intelligenz weiter aus. Damit verstärkte es sowohl den schon seit ChatGPT-3 bestehenden medialen Hype um den Chatbot, als auch die Attraktivität für andere Unternehmen, die das Basis Modell GPT-4 potentiell für sich nutzen wollen [?]. Ein solches gründeten Hendrik Haverkamp und Malte Hecht und entwickelten den KI-gestützten Feedback-Tutor „fiete.ai“ [?], welcher Schüler:innen automatisiert Feedback zu ihren Bearbeitungen von Aufgaben gibt, die vorher von Lehrkräften erstellt und eingepflegt wurden. Diese Art und Weise Feedback individuell zu erstellen, verspricht Entlastung für Lehrkräfte und ein fortschrittliches Unterrichtsdesign für Schüler:innen mit sich zu bringen.

Dass nicht nur die indirekte Nutzung eines neuartigen Sprachmodells wie GPT-4, sondern auch dessen Verwendung im Schulunterricht potentiell gewisse Technikfolgen, also Auswirkungen auf die Nutzer:innen und die Gesellschaft haben kann, sollte dabei zumindest bedacht werden. Deshalb bietet es sich an eine Technikfolgenabschätzung durchzuführen.

Dazu verwende ich im folgenden Text die Methode der „technikinduzierten sozioinformatischen Analyse“ (TSIA) nach Zweig et al. [?, Seite 113]. Diese beschreibt eine spezielle Vorgehensweise bei der Technikfolgenabschätzung und beruht auf der Annahme, dass es sich beim vorliegenden um ein sozioinformatisches System handelt. Systeme sind dann sozioinformatisch, wenn bei der Wechselwirkung zwischen einem sozialen System und einem technischen System „[...] der zentrale Technikteil ein Hard- und/oder Softwaresystem ist oder ein solches enthält“ [?, Seite 70]. Dies trifft für die Betrachtung von fiete.ai sicherlich zu, da sowohl der Schulunterricht mit Lehrkräften und Schüler:innen ein komplexes soziales System, als auch die Kombination von fiete.ai und GPT-4 ein komplexes Softwaresystem darstellen. Zweig et al. zitieren hier den VDI, nach dem Technikfolgenabschätzung zu betreiben bedeutet, den aktuellen Stand der vorliegenden Technik zu analysieren, technische, gesundheitliche, soziale und andere Folgen dieser abzuschätzen, diese Folgen zu beurteilen, wünschenswerte Entwicklungen zu fordern und Handlungs- und Gestaltungsmöglichkeiten herzuleiten sowie herauszuarbeiten [?, ?, Seite 10].

In dieser Analyse werden ich nach der Vorgehensweise der TSIA, zuerst das vorliegende informatische System beschreiben und verständlicher machen, dann die relevanten sozialen Akteure und deren Motivationen herausarbeiten, damit ein Wirkungsgefüge aufstellen, welches Systemvariablen und deren Wechselwirkungen modelliert, dieses Wirkungsgefüge dann analysieren und auf mögliche Technikfolgen untersuchen, und schließlich „zweite Gedanken denken“, indem ich versuche

aus der Perspektive der Akteure auf das System zu blicken [?, Seite 114].

2 Technikbeschreibung

In diesem ersten Abschnitt wird sowohl das Feedbacktool `fiete.ai`, als auch die Funktionsweise des zugrundeliegenden Sprachmodell GPT-4 von OpenAI, erklärt und deren Prozessabläufe verständlich gemacht. Dabei gehe ich auch auf die Art und Weise wie `fiete.ai` sich den Chatbot zunutze macht und mit ihm kommuniziert ein. Außerdem wird herausgestellt, welche technischen Grenzen und Möglichkeiten sich aus der Synthese beider Technikkomponenten ergeben.

2.1 `fiete.ai`

Das Feedback- und Bewertungstool `fiete.ai`, entwickelt von Hendrik Haverkamp und Malte Hecht, soll Lehrer:innen unterstützen, indem es „allen Schüler:innen ein individuelles Feedback zu ihren Texten und Aufgaben gibt“ [?, Homepage], wobei die Kriterien und Vorgaben von den Lehrkräften selbst einstellbar sind. Es folgt eine Beschreibung der „User-Experience“, sowohl aus Lehrer:innen-, als auch aus Schüler:innenperspektive.

Befindet man sich, als Lehrkraft in der Web-App [?, Web-App], der tatsächlichen Nutzer:innenoberfläche der Website, so kann man wählen, ob man eine neue Aufgabe erstellen oder eine alte einsehen und modifizieren will. Entscheidet man sich für ersteres, kann man wiederum aus Vorlagen zu gewissen Aufgabentypen wählen oder eine leere Vorlage öffnen, um seine gänzlich eigene Aufgabe zu erstellen. Unabhängig von seiner Wahl hat man dann die Möglichkeit in mehreren Texteingabefeldern den Namen der Aufgabe, die eigentliche Aufgabenstellung, bis zu sieben Feedbackkriterien und zusätzliches Material einzugeben oder abzuändern. Nachdem man die Aufgabe erstellt hat, werden ein Link und ein QR-Code generiert, welche man an seine Schüler:innen verteilen kann, um damit den Zugang zur Beantwortungsoberfläche zu gewähren. Des Weiteren hat man die Möglichkeit die Sprache einzustellen in der Fiete antworten soll, eine Rechtschreib- und Grammatikprüfung im Antwortfeld an- oder abzuschalten, die Feedbackkriterien sicht- oder unsichtbar zu machen und den erlaubten Abgabezeitraum festzulegen. Hat schon ein:e Schüler:in eine Abgabe getätigt, kann die Lehrkraft an diesem Punkt auf den Reiter „Ergebnisse“ wechseln und dort sowohl die Abgabe selbst als auch das Feedback von Fiete zu jedem/jeder Schüler:in einsehen.

Aus Schüler:innenperspektive beginnt die Nutzung mit dem Anklicken des

Links, welcher die Bearbeitungsansicht öffnet. Dort werden sowohl die Aufgabe als auch, sofern von der Lehrkraft aktiviert, die Feedbackkriterien angezeigt. Im nachfolgenden Textfeld kann die erste Abgabe getätigt werden. Alternativ zur Tastatureingabe kann auch Text abfotografiert und per Texterkennung eingegeben werden. Hat der/die Schüler:in seinen/ihren Text vollständig eingegeben und möchte diesen nun analysieren lassen, klickt er/sie auf den darunterliegenden Button „Fiete, gib mir Feedback!“ Nach kurzer Ladezeit wird das KI-generierte Feedback, geordnet nach den dazugehörigen Feedbackkriterien unter dem abgegebenen Text dargestellt. Außerdem steht darüber noch ein allgemeiner Absatz, in dem die Kernaussagen der einzelnen Feedbackteile zusammengefasst sind. Neben der ersten Abgabe kann der/die Schüler:in nun eine zweite, finale Abgabe tätigen. Nachdem er/sie die Anmerkungen des Feedbacks wahrgenommen und nach Möglichkeit in Form einer neuen korrigierten Version umgesetzt hat, kann diese nun ebenfalls von Fiete verarbeitet werden. Per Klick auf den neu erschienenen Button „Endgültig abgeben“ wird ein finales Feedback generiert und ebenfalls geordnet und mit überstehender Zusammenfassung unter der Abgabe angezeigt.

Um das Feedback passend zu den jeweiligen Kriterien generieren zu können und um dessen Qualität zu steigern, verwendet fiete.ai ein spezielles Verfahren, mit dem es die zugrundeliegende Künstliche Intelligenz anspricht. Um verstehen zu können, wie genau diese Schnittstelle zwischen beiden Technikkomponenten gestaltet ist, müssen wir uns zuerst anschauen, wie das laut Malte Hecht¹ verwendete „Language Model“ GPT-4 funktioniert.

2.2 ChatGPT

Ein Language Model (LM) ist eine Anwendung des „Natural Language Processing“ (NLP), dem interdisziplinären Teilgebiet der Informatik und Linguistik, welches sich der Verarbeitung und dem Verständnis natürlicher, von Menschen gesprochener Sprache mithilfe von Computern widmet [?].

Laut Hiemstra sind LMs „Generative Modelle, also solche die einen Mechanismus für Wahrscheinlichkeiten definieren, um natürliche Sprache zu generieren.“. Sie „weisen einem unbekanntem Textstück eine Wahrscheinlichkeit basierend auf gewissen Trainingsdaten zu“ [?].

Zu diesen LMs gehören auch die von OpenAI entwickelten GPTs. Sie bauen iterativ aufeinander auf, das ursprüngliche GPT-1 aus 2018 wurde also fortlaufend weiterentwickelt, bis schließlich im Mai 2020 GPT-3 und im März 2023 GPT-4

¹Alle technischen Informationen und Hintergründe zur Funktionsweise von fiete.ai selbst und den Einsatz von GPT-4, stammen aus einem Gespräch/Interview mit fiete.ai-Softwareentwickler Malte Hecht

veröffentlicht wurde [?, Seite 1123].

Das Kürzel „GPT“ steht für „Generative pre-trained Transformer“.

Als Transformer bezeichnet man eine gewisse Architektur eines „künstlichen neuronalen Netzes“ (KNN)[?, 3]. Sie wurde von Google Reasearchern 2017 entwickelt und beruht auf dem Konzept der „Self-Attention“. Dieses erlaubt es nach Vaswani et al. für jede einzelne Position einer Input-Sequenz einen „Attention-Score“ zu berechnen, indem ihre Repräsentation im Neuronalen Netz mit den der jeweils anderen Positionen verglichen wird. Aus allen Attention Scores einer Sequenz kann dann eine gewichtete Summe berechnet werden, welche die Anzahl und Intensität der Bezüge, einzelner Positionen zu anderen innerhalb der Sequenz beinhaltet[?, 3.2].

„Self-attention, sometimes called intra-attention, is an attention mechanism relating different positions of a single sequence in order to compute a representation of the sequence.“ [?]

Mit dieser Art von Darstellung erhofft man sich, dass das Model mehr Informationen über die Semantik des Inputs erhält [?].

„Pre-trained“ heißt das Model, weil es mit dem Verfahren des „self-supervised Learning“ auf großen Datenmengen vortrainiert wurde. Bei dieser Art von Training lässt man das Generative Modell gewissermaßen mit den Trainingsdaten alleine und lässt es sich daraus eine Grundstruktur generieren [?, ?, Seite 1124 II.A]. Darauf aufbauend wird dann ein „supervised“ fine-tuning vorgenommen, welches das Neuronale Netz auf den speziellen Anwendungsfall anpasst und „aligned“, also dazu bringt ethisch bedenkliche oder beispielsweise gar rassistische Äußerungen, die in den Trainingsdaten enthalten waren zu unterlassen.[?]

2.3 fiete.ai-Prompting

Um GPT-4 als Basis für die eigene Software nutzen zu können, verwendet fiete.ai eine API (Application-Programmer-Interface) von OpenAI mit dem der Server der fiete.ai Website über eine Schnittstelle mit dem Language Model kommunizieren kann. Wird eine Abgabe getätigt, und Feedback angefordert, so werden sowohl die Aufgabenstellung, die Feedbackkriterien, die zur Verfügung gestellten Materialien, als auch die erste Abgabe in eine vorgefertigten Promptmaske zusammengeführt. Laut Malte Hecht kommt dabei ein „Chain-of-thought-Prompting“ zum Einsatz. Bei dieser Art Prompts, also Eingaben bzw. Befehle, zu designen, wird das Sprachmodell angewiesen bei der Bearbeitung schrittweise vorzugehen und die logische Verkettung der einzelnen Lösungsschritte herauszustellen. Dadurch kann nach Wei et al. „die Fähigkeit von LLMs komplexe Argumentationen durchzuführen signifikant verbessert werden“ [?, Seite 1]. Bei fiete.ai wird dieses Verfahren genutzt um

die Qualität des generierten Feedbacks zu verbessern, die Argumentation nachvollziehbarer zu gestalten und Halluzinationen, also im Kontext unpassende und scheinbar wahllos generierte Outputs, zu reduzieren [?]. Die Funktionsweise dieses Verfahrens ist möglicherweise dadurch zu erklären, dass sich bei der schrittweisen Generierung der Antwort, die einzelnen Lösungsschritte immer wieder im Kontextfeld befinden. Dieses Kontextfeld bestimmt, was zusätzlich zum Prompt bei der Generierung jedes nächsten Tokens berücksichtigt werden soll. Laut Malte Hecht wird die Antwort auf die Prompts immer im Json-Format gegeben, was entweder erreicht werden kann, indem man den Json-Mode, eine Voreinstellung die man beim Aufruf der API treffen kann, aktiviert oder indem im Prompt selbst die Aufforderung zur Antwort in JSON steht. Das Format kann dann auf Seiten von fiete.ai validiert und laut eigenen Aussagen effizienter verarbeitet werden, als käme die Antwort im String-Format zurück. Nach der Validierung wird das generierte Feedback dann aufbereitet, und geordnet nach den Feedbackkriterien in der Web-App angezeigt.

2.4 Grenzen und Möglichkeiten

Im Sinne der Analyse stellt sich nun die Frage, wie fähig diese Art der Feedbackerstellung ist, und welche Qualität sie hat. Seine Größe und die Transformer-Architektur mit dem Self-Attention Mechanismus machen es dem Sprachmodell GPT-4 möglich komplexe Satzkonstruktionen zu produzieren, auf den Inhalt eines mehrere Seiten langen Textes einzugehen und dazu gestellte Fragen scheinbar mit Logik zu beantworten. Es erweckt den Anschein als könne es den Inhalt des Inputs tatsächlich verstehen. Tatsächlich kann von einem Verständnis wie wir es vom Menschen kennen, jedoch nicht die Rede sein. Letztlich sind die ausgegebenen Tokens doch nur das Ergebnis eines, zwar größtenteils undurchschaubaren, aber in der Funktionsweise verständlichen, statistischen Modells. Dieses ist zwar durch raffinierte Techniken wie der Attention oder dem Chain-of-thought Prompting dazu fähig Sprache zu simulieren, die vermutlich vor Zeiten der GPTs von den meisten als eindeutig menschengemacht eingeordnet worden wäre, allerdings bleiben die Outputs letztendlich genau das: Simulationen davon was auf Basis der Trainingsdaten und des fine-tunings die wahrscheinlichsten nächsten Tokens in einer Zeichenkette wären. In „Ohne Sinn- Zu Anspruch und Wirklichkeit automatisierter Aufsatzbewertungen“ leiten Zweig und Schneider her, warum eine automatisierte Bewertungsvorhersage, wie sie der „E-Rater“ durchführt, keine tatsächliche Bewertung sein kann [?]. Dazu argumentieren sie anhand der Sprechakttheorie von Austin, die besagt, dass eine Äußerung selbst zu einer Handlung werden kann. Allerdings nur dann wenn gewisse Gelingenheitsbedingungen dafür erfüllt sind[?]. In ihrem Buch „Die KI war’s!“ argumentiert Zweig ähnlich und zeigt, dass ChatGPT, beziehungsweise zeitgenössisches Maschinelles Lernen generell, keine Bewertungen

im Sinne der Sprechakttheorie vornehmen kann: „Maschinelles Lernen [. . .] kann [. . .] nicht das für die Entscheidung »übliche Verfahren« im Sinne der Sprechakttheorie lernen. Dazu müssten zum Beispiel innere Zustände der menschlichen Entscheider zum Lernen zur Verfügung stehen“ [?]. Das „übliche Verfahren“ meint hier die Gelingenheitsbedingung A.1 von Austin, die besagt: „Es muss ein übliches konventionales Verfahren mit einem bestimmten konventionalen Ergebnis geben; zu dem Verfahren gehört, dass bestimmte Personen unter bestimmten Umständen bestimmte Wörter äußern“. Da Maschinelles Lernen allerdings, innerhalb der Neuronalen Netze, Verfahren anwendet, die mit heutigen Methoden nicht vollständig verständlich gemacht werden können, schlussfolgert Zweig, dass die Bedingung nicht erfüllt ist. Auch zeigt Zweig, dass die Begründungen und Rechtfertigungen, die ChatGPT für die Bewertungen auf Nachfrage anführt, nicht konsistent sind und ebenfalls keinem eindeutigen Verfahren folgen. Wenden wir nun die dadurch gewonnen Erkenntnisse auf fiete.ai an, so kommen ich zu dem Schluss, dass es sich mit dem Feedbacktool im Wesentlichen gleich verhält. Zwar könnte man argumentieren, dass durch die von der Lehrkraft voreinstellbaren Feedbackkriterien ein gewisses „übliches Verfahren“ implementiert wird, allerdings kann sich ein Neuronales Netz, aufgrund seiner Funktionsweise nicht an solche „chrispe“, also klare und präzise, Kriterien halten. Außerdem lässt sich für fiete.ai ebenfalls zeigen, dass bei der wiederholten Eingabe des exakt selben Textes, unterschiedliche Feedbacks generiert werden können. Nun stellt sich als Abschluss der Technikbeschreibung für diese Analyse die Frage, ob das Feedback was fiete.ai generiert denn, wenn es auch die Anforderungen an einen Sprechakt nicht erfüllt, trotzdem eine gewisse, im Unterricht nutzbare Qualität hat. Die Expertenmeinung von Prof. Zweig besagt an dieser Stelle, dass ca. 80% des von fiete.ai gegebenen Feedbacks tatsächlich sinnvoll, also nicht zwingend gut und perfekt, aber jedenfalls nicht kontraproduktiv sind². Im Sinne der folgenden Analyse nehmen wir dieses 80:20- Verhältnis zwischen sinnvollem und tatsächlich kontraproduktivem, gar verwirrendem Feedback als realistisch an und arbeiten damit weiter.

3 Akteure und deren Motivationen

3.1 Akteure

An der Nutzung von Feedbacktools wie fiete.ai sind mehrere Akteure maßgeblich beteiligt. Zum einen die Lehrkräfte, die sowohl für die tatsächliche Integration der Software in ihren Unterricht als auch die Erstellung der Aufgaben zuständig und verantwortlich sind. Sie sind außerdem als Hauptansprechpartner:innen für Schüler:innen dafür verantwortlich offene Fragen nach der Benutzung von fiete.ai zu

²Diese Information stammt aus persönlichen Gesprächen mit Prof. Zweig

beantworten und gegebenenfalls unklares oder gar falsches Feedback richtigzustellen.

Zum anderen sind die Schüler:innen selbst Akteure, da sie die primären Nutzer:innen der Software sind und die Empfänger:innen des gegebenen Feedbacks darstellen.

Als Entwickler sind auch Hendrik Haverkamp und Malte Hecht Akteure mit einem Interesse an der Nutzung von fiete.ai.

Den Analysehorizont, also den Zeitraum, in dem wir das System auf Technikfolgen untersuchen wollen, setzen wir auf Wochen bis wenige Monate fest. Er entspricht also ungefähr der Dauer einer Unterrichtsreihe beziehungsweise eines Themenblocks im Schulunterricht.

3.2 Motivationen

- **Lehrkräfte**

Als Lehrbeauftragte und Organisator:innen des Unterrichts haben die Lehrkräfte ein Interesse an einer zuverlässigen Funktionsweise des Feedbacktools. Laut der „Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz“ (SWK) bedroht der Mangel an qualifiziertem Lehrpersonal die „Sicherstellung der Unterrichtsversorgung und beeinträchtigt auch die Qualität des Unterrichts. Der Lehrkräftemangel deutete sich seit Jahren in Prognosen an, steigende Zahlen von Schüler:innen bei stagnierenden bis sinkenden Zahlen von Studierenden und steigenden Zahlen von Lehrkräftepensionierungen haben dazu beigetragen“ [?]. In ihrer „Empfehlung zum Umgang mit dem aktuellen Lehrkräftemangel“ empfiehlt die Konferenz beispielsweise Maßnahmen wie die „Entlastung und Unterstützung qualifizierter Lehrkräfte durch Studierende und andere, formal nicht (vollständig) qualifizierte Personen“ oder auch eine „Flexibilisierung durch Hybridunterricht, Erhöhung der Selbstlernzeiten sowie Anpassung der Klassenfrequenz“[?]. Im Kontext des Hybridunterrichts fällt dabei auch der Begriff des „Blended Learning“, welcher die Kombination von Präsenzunterricht und E-Learning beschreibt und damit auch die Einbindung von Feedbacktools wie fiete.ai enthält [?]. Außerdem heißt es in dem Empfehlungsschreiben: „Die SWK empfiehlt, [. . .] für den Fall, dass mehr Korrekturaufgaben auf die betroffenen Lehrkräfte zukommen, diesen qualifizierte Korrekturassistent:innen an die Seite zu stellen“ [?]. Weiter empfiehlt die Konferenz, die Selbstlernzeiten zu erhöhen, beispielsweise mit Konzepten wie „Flipped Classroom“, bei dem „Lernende sich neue (Teil-)Themen zunächst über geeignete Video-, Audio- oder Textmaterialien selbstständig und im eigenen Tempo aneignen. Der anschließende Präsenzunterricht fokussiert dann auf die Klärung von Verständnisfragen und eine vertiefte und erweiterte Auseinandersetzung mit den Inhalten“ [?]. Dabei

wird außerdem angemerkt, dass „die Begleitung in kooperativen Selbstlernzeiten nicht zwangsläufig durch die Lehrkraft geleistet werden muss, sondern auch durch fachlich weniger qualifizierte Tutor:innen erfolgen kann, die eine allgemeine didaktische Qualifizierung zur Unterstützung kooperativer Lernprozesse erhalten haben (Grunze et al., 2004)“ [?, ?]. Zieht man all diese Anmerkungen in Betracht, so scheint es Bedarf zu geben, den Lehrkräften repetitive Aufgaben, wie beispielsweise Fehlerkontrollen bei den Hausaufgaben, abzunehmen oder diese zu erleichtern, sodass dann mehr Zeit für individuelle und konkrete Betreuung einzelner Schüler mit Lernschwierigkeiten oder spezifischen Fragen aufgewendet werden kann.

Wir halten diese Art der **Zeitersparnis durch fiete.ai** also als eine erste Motivation der Lehrkräfte fest. Dabei ist zu beachten, dass selbstverständlich nicht jede:r Lehrer:in dieses Ziel verfolgt, wir es allerdings im Sinne der Analyse als überstehendes Interesse der Lehrerschaft als Ganze annehmen. Des Weiteren nehmen wir als Motivation an, dass sich **möglichst viele Schüler:innen durch den Einsatz von fiete.ai im Unterricht verbessern** sollen. Diese Annahme sollte bezogen auf die gesamte Lehrerschaft trivial sein.

- **Schüler:innen**

Die Motivationen von Schüler:innen zur Nutzung von Lerngelegenheiten im Unterricht sind komplex und hängen von sozialen, kulturellen Faktoren, aber vor allem auch von persönlichen Erfahrungen und der Einschätzung der eigenen Fähigkeiten ab [?]. Zur Vereinfachung nehme ich allerdings eine generelle Motivation zur Verbesserung der eigenen Leistungen für die Schülerschaft als Ganze an. Zwar wird es in jeder Schüler:innenkonstellation einzelne geben, die von vornherein weniger Motivation zur sinnvollen Bearbeitung der Aufgaben haben, allerdings nehme über den aus Schülerperspektive recht langen Betrachtungszeitraum einer ganzen Unterrichtsreihe an, dass die Schüler:innen sich grundsätzlich durch fiete.ai verbessern wollen.

Wir halten als Motivation der Schüler also fest, dass sie sich insgesamt **durch den Einsatz von fiete.ai verbessern** und das **Lernziel der Unterrichtsreihe erreichen** wollen. [?]

- Die **Entwickler** Aus persönlichen Gesprächen mit den Entwicklern Hendrik Haverkamp und Malte Hecht geht hervor, dass sowohl die Anzahl der Anwendungsfälle als auch die Entlastung der Lehrkräfte Motivationen für sie darstellen. Steigt die Anzahl der Anwendungsfälle von fiete.ai so wird das Tool vermutlich öfter an anderen Schulen eingesetzt, die Reputation steigt und weitere Innovations- und Entwicklungsschritte werden durch das Wachstum möglich. Da für sie aber auch die Zufriedenheit der Lehrkräfte, ihrer primär

Nutzenden von Interesse ist, wollen sie auch, dass diese durch ihre Software entlastet werden.

3.3 Umwelt

Nicht zu den Akteuren zählen die Eltern der Schüler:innen, da sie bezogen auf den Analysehorizont dieselben Motivationen haben. Auch könnte man OpenAI als möglichen Akteur aufführen, da das Unternehmen die Entwicklerseite der zugrundeliegenden künstlichen Intelligenz darstellt. Allerdings scheint es unrealistisch, dass das internationale Unternehmen innerhalb des Analysezeitraums direkten Einfluss speziell auf das betrachtete System nehmen würde.

4 Erstellung der Wirkungsgefüge

In diesem Schritt werden wir die verschiedenen Handlungen und Kausalitäten möglichst kompakt in einem Wirkungsgefüge modellieren. Ein Wirkungsgefüge ist dabei lediglich „ein sehr einfaches visuelles Modell von kausalen Abhängigkeiten zwischen messbaren Eigenschaften oder messbarem Verhalten des Systems“ [?, Seite 99].

Um die Auswirkungen unserer Annahme, dass 80% des Feedbacks von fiete.ai sinnvoll ist, herauszustellen, designen wir allerdings ein Wirkungsgefüge für ein perfekt funktionierendes fiete.ai-System, bei dem 100% sinnvolles Feedback, ununterscheidbar von einer idealen, menschlichen Lehrkraft gegeben werden und ein „realistisches“, welches die 80% berücksichtigt.

4.1 Systemvariablen

Zur Modellierung des Systems „fiete.ai im Schulunterricht“ müssen gewisse Systemvariablen festgelegt werden. Mithilfe dieser quantifizierbaren Größen und den kausalen Beziehungen zwischen ihnen, den Systemkanten, können später Aussagen über die Funktionsweise des modellierten Systems getroffen werden.

Die Systemvariablen werden durch beschriftete Kästen im Wirkungsgefüge dargestellt, die Kanten entweder durch einfache Pfeile, die agonistische Beziehungen darstellen („Je höher [...], desto höher [...]/Steigt die [...], so steigt auch die [...]“), oder gestrichelte Pfeile, welche antagonistische Wirkungen symbolisieren („Je höher [...], desto niedriger [...]/Steigt die [...], so sinkt die [...]“).

Zusätzlich werden noch die relevanten Akteure in Kreisen dargestellt und deren jeweilige Motivationen nach der selben Logik (durchgezogene Linien für „Akteur will [...] erhöhen/ will, dass [...] steigt“ und gestrichelte Linien für „Akteur will [...] senken/ will, dass [...] abnimmt“).

fiete.ai wird zur Vereinfachung im Kontext des Wirkungsgefüges nur noch fiete genannt. Außerdem wird im Wirkungsgefüge selbst Schüler:innen mit S und Lehrkräfte mit L abgekürzt.

1. Anzahl der Einsätze von fiete (im Unterricht)

Diese Kenngröße gibt die Anzahl der Einsätze von fiete.ai-Aufgabenstellungen in einer Unterrichtsreihe an. Sie wird letztlich durch die Lehrkraft bestimmt, die entscheidet für wie nützlich sie den Einsatz der Feedbacksoftware hält und wie oft sie damit Aufgaben erstellen will.

2. Anzahl der gewissenhaft bearbeiteten Aufgaben

Die Anzahl der Aufgaben, die von Schüler:innen sorgsam und unter Einsatz

ihrer derzeitigen Fähigkeiten bearbeitet werden. Sie sinkt nicht, wenn ein:e Schüler:in lediglich eine schlechte Abgabe tätigt, sondern nur dann, wenn er/sie die Aufgabe ungültig, nicht im Sinne der Aufgabenstellung oder gar nicht bearbeitet. Quantifiziert werden könnte diese Größe indem man ungültige und unsinnige Abgaben zählt oder genauer indem man in einer Testreihe die Schüler:innen befragt.

3. Anzahl der S, die sich (durch fiete) verbessern

Bekommen Schüler:innen durch fiete.ai sinnvolles, oder gar pädagogisch wertvolles Feedback, so können sie sich und ihr Verständnis des Themas oder des Lernziels tatsächlich verbessern. Gemeint ist hier die Anzahl derer, die sich ohne nennenswerte nachträgliche Betreuung oder individuelles Feedback durch die Lehrkraft, durch die Nutzung von fiete.ai verbessert oder von ihr profitiert haben.

4. Anzahl der S, die individuelle Betreuung (durch L) brauchen

Diese Variable beschreibt die Anzahl der Schüler:innen, die wegen fiete.ai weitere individuelle Betreuung benötigen. Genauer sind diejenigen gemeint, die nach einem widersprüchlichen Feedback eine Klärung von der Lehrkraft fordern, oder bei denen die Anweisungen zu unverständlich oder sprachlich komplex waren. Jegliches schlicht falsches Feedback kann auch dazu führen, dass skeptische Schüler:innen dies bei der Lehrkraft melden und dann individuell betreut werden müssen.

5. Zeitersparnis durch fiete

Entsprechend kann allerdings durch erfolgreiche Nutzung von fiete.ai potentiell Zeit gespart werden, da das Tool im Gegensatz zur Lehrkraft, allen Schüler:innen gleichzeitig zur Seite stehen kann. In einem Unterrichtsaufbau ohne KI-Feedbacktutor wäre ein so individuelles Feedback entweder nicht oder nur mit großem Zeitaufwand möglich.

6. Erfolgsquote des Einsatzes (von fiete)

Dieses Verhältnis gibt an wie viele der Schüler:innen am Ende der Unterrichtsreihe eine Verbesserung durch fiete.ai erzielen konnten. Man könnte diese Größe messen, indem man eine Testgruppe an Schüler:innen eine erste Aufgabenstellung bearbeiten lässt, diese von Menschen Aus- und Bewerten lässt und die Schüler:innen dann an weiteren ähnlichen Aufgaben mit fiete.ai üben lässt. Schließlich würde man einen letzten Test schreiben, bei dem dieselben Fähigkeiten und Lernziele abgefragt werden. Die Anzahl an Schüler:innen die sich im Vergleich zum ersten Test verbessert hätten, geteilt durch die Anzahl der Größe der Testgruppe würde vermutlich grob dieser Erfolgsquote entsprechen.

4.2 Kanten

1. **Kante I** stellt die agonistische Beziehung zwischen der *Anzahl der Einsätze von fiete* und der *Anzahl der gewissenhaft bearbeiteten Aufgaben* dar. Sieht die Lehrkraft vor, dass mehr fiete-Aufgaben im Unterricht zum Einsatz kommen, so werden auch mehr Aufgaben insgesamt von den Schüler:innen bearbeitet und ein gewisser Anteil davon, die *Anzahl der gewissenhaft bearbeiteten Aufgaben* nimmt ebenfalls zu.
2. **Kante II** ist diejenige, die im ersten Wirkungsgefüge eine 100% sinnvolle Feedbackrate symbolisiert. „Sinnvoll“ bedeutet in diesem Fall, dass die Empfehlungen die fiete gibt auf jeden Fall nicht falsch, irreführend oder kontraproduktiv für den/die Schüler:in sind. Folglich können wir annehmen, dass sich die *Anzahl der S, die sich verbessern* wenn auch nicht unbedingt nach jedem einzelnen Feedback, insgesamt erhöhen wird.
3. **Kante III** stellt die direkte Konsequenz davon dar, also dass sich durch mehr profitierende Schüler:innen die Erfolgsquote des Einsatzes erhöht.
4. **Kante IV** steht für die agonistische Beziehung zwischen der *Anzahl S, die sich verbessern* und der *Zeitersparnis durch fiete*. Diese Abhängigkeit ergibt sich dadurch, dass wenn mehr Schüler:innen sich durch fiete verbessern, die Lehrkräfte für diese Schüler:innen keine zusätzliche Betreuung leisten müssen.
5. **Kante V** Wenn die *Anzahl gewissenhaft bearbeiteter Aufgaben* steigt, und wir von einer „sinnvolles Feedback“-Rate von 100% ausgehen, so sinkt auch die *Anzahl S, die individuelle Betreuung brauchen*, da fiete selbst den Großteil dieser Betreuung übernehmen kann. Falls ein/e Schüler:in trotzdem noch Schwierigkeiten hat, das Lernziel zu erreichen, so kann er/sie eine weitere Aufgabe mit fiete bearbeiten und sich darauf verlassen, weiteres sinnvolles Feedback zu bekommen.
6. **Kante VI** Steigt jedoch die *Anzahl S, die individuelle Betreuung brauchen* an, so reduziert sich die *Zeitersparnis durch fiete*, da dann wieder die Lehrkraft persönlich aktiv werden muss.
7. **Kante VII** stellt eine Rückwirkung der *Zeitersparnis durch fiete* auf die Anzahl der Einsätze von fiete dar. Sie beruht schlicht auf der Annahme, dass wenn die Lehrkraft durch fiete entlastet wird, ein Anreiz geboten ist das Tool auch weiterhin, wenn nicht sogar öfter, im Unterricht einzusetzen.

8. **Kante VIII** Diese letzte Kante besagt, dass die *Anzahl S, die individuelle Betreuung brauchen* sich antagonistisch auf die *Anzahl gewissenhaft bearbeiteter Aufgaben* auswirkt. Dies lässt sich mit negativen Erfahrungen der Schüler:innen begründen. Haben einige Schüler:innen (mehrmals) kontraproduktives, widersprüchliches oder falsches Feedback von fiete bekommen und mussten deshalb die Lehrkraft um Hilfe fragen, so liegt es nahe, dass ihre Vertrauen und ihre Motivation weiterhin mit fiete zu arbeiten sinkt.

Besonderheiten im “80%- Wirkungsgefüge“

1. **Kante II*** Da die “sinnvolles Feedback“-Rate hier nur 80% beträgt, wird die Wirkung dieser Kante zwar immer noch agonistisch, aber vermutlich nicht mehr genauso „stark“ sein. Die *Anzahl der Schüler:innen, die sich verbessern*, steigt bei gleichem Anstieg der *Anzahl der gewissenhaft bearbeiteten Aufgaben* nicht mehr genauso stark an.
2. **Kante V*** Diese Kante jedoch wird von einer antagonistischen zu einer agonistischen. Da jetzt 20% des von fiete gegebenen Feedbacks eher kontraproduktiv für die Schüler:innen ist, wird die *Anzahl der S, die individuelle Betreuung brauchen* dadurch eher ansteigen. Zwar werden sich immer noch tendenziell mehr Schüler:innen durch die Nutzung verbessern, allerdings werden nichtsdestotrotz mehr Schüler:innen Feedback benötigen, wenn die „Anzahl gewissenhaft bearbeiteter Aufgaben“ steigt.

4.3 Abbildungen der Wirkungsgefüge

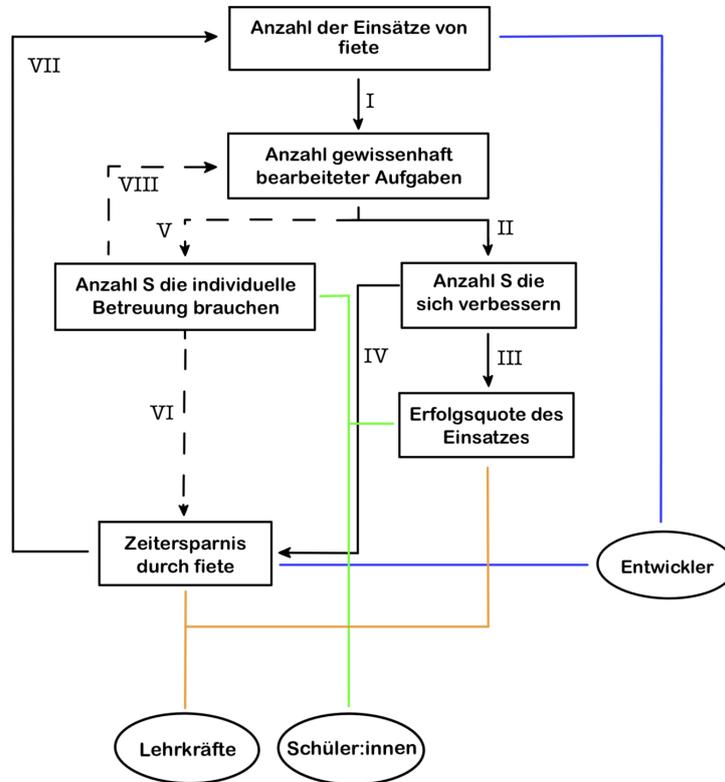


Abbildung 1: Wirkungsgefüge des Systems unter Idealbedingungen.

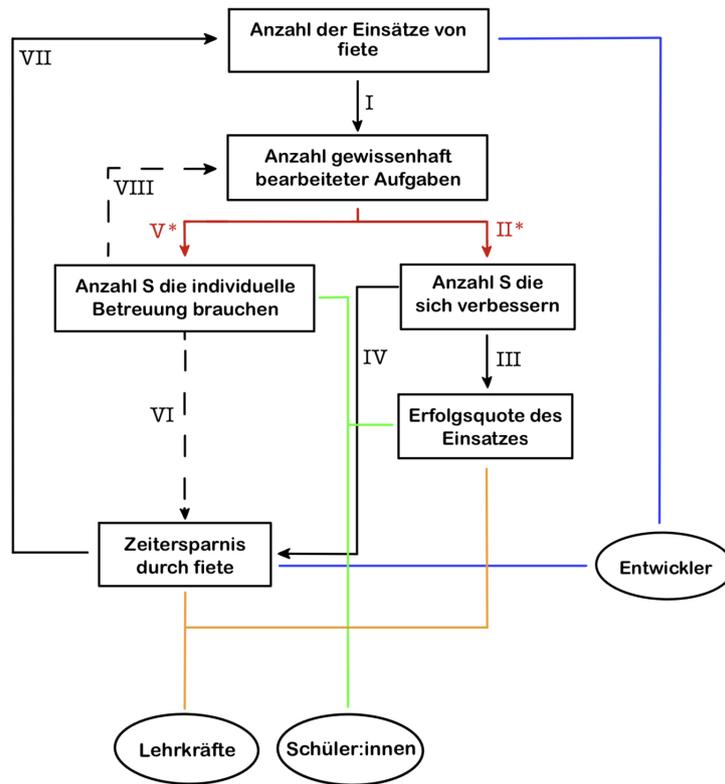


Abbildung 2: Wirkungsgefüge mit 80% sinnvollem Feedback.
 Die Unterschiede sind rot gekennzeichnet

5 Analyse der Wirkungsgefüge

In dieser Phase der Analyse untersuchen wir die erstellten Wirkungsgefüge, um mehr über die Dynamik des zugrundeliegenden Systems zu erfahren. Wir gehen der Frage nach, „wie die sozialen Akteure basierend auf ihren eigenen Motiven und der Anreizstruktur, die die Software in ihrer genauen Verwendung bietet, reagieren könnten“ [?, Seite 119]. Im ersten Schritt suchen wir dabei nach bestimmten Mustern, sogenannten Feedbackzyklen. Diese können entweder positiv, also „selbstverstärkend“ oder negativ, „gedämpft“ sein. Sie stellen eine Rückkopplung einer Variable (wenn auch über Umwege) auf sich selbst dar [?, Seite 120].

“100%-Wirkungsgefüge“

Einen solchen positiven Feedbackzyklus finden wir in Form der Kanten

„ $I \rightarrow II \rightarrow IV \rightarrow VII \rightarrow I$ “

Dass er positiv ist, erkennen wir daran, dass er ausschließlich aus agonistischen Kanten besteht [?, Seite 120]. Er bedeutet, dass wenn *fiete* eingesetzt wird, sich dadurch Schüler:innen verbessern, folglich eine Zeitersparnis entsteht und ein Anreiz für die Lehrkraft gegeben ist, *fiete* noch öfter einzusetzen. Diese selbstverstärkende Wirkung ist natürlich nicht tatsächlich unendlich, sondern endet dann, wenn die *Anzahl der Einsätze von fiete*, die Zeitersparnis oder die *Anzahl S die sich verbessern* ihr Maximum erreichen.

Wir können also daraus schließen, dass in einer Konstellation in der ein „perfektes“ Feedbacktool im Unterricht eingesetzt wird, die Tendenz besteht, die zur individuellen Betreuung von der Lehrkraft aufgewendete Zeit möglichst stark zu reduzieren. Je nachdem wie stark diese Entwicklung tatsächlich voranschreiten würde, ergäbe sich daraus vermutlich eine starke Veränderung des Unterrichtsalltags, wenn die Lehrkraft tatsächlich maximal viel Zeit einsparen und Unterricht durch *fiete* gestalten wollte. Ungeachtet anderer Konsequenzen und Technikfolgen, die eine so umfassende Nutzung des digitalen Dienstes hätte, wären im vorliegenden Wirkungsgefüge alle „Wünsche“ in Form der Motivationen der Akteure „erfüllt“. Auch möglich ist, dass die Lehrkraft ab einem bestimmten Punkt die weitere Erhöhung der *Anzahl der Einsätze von fiete* einstellt, um ebendiese strukturelle Veränderung des Unterrichts zu verhindern.

“80%-Wirkungsgefüge“

Denselben positiven Feedbackzyklus „ $I \rightarrow II^* \rightarrow IV \rightarrow VII \rightarrow I$ “ gibt es auch im „realistischeren“ Wirkungsgefüge noch. Dazu kommt allerdings ein kleinerer, negativer Feedbackzyklus zwischen den Kanten „ $V^* \rightarrow VIII$ “. Er ist negativ, weil er aus einer agonistischen und einer antagonistischen Kante besteht und stellt die

negative Rückkopplung von kontraproduktivem, nicht sinnvollem Feedback auf die *Anzahl gewissenhaft bearbeiteter Aufgaben* dar [?, Seite 120]. Als “Brückenvariable“ dient dabei die *Anzahl S, die individuelle Betreuung brauchen*. Außerdem entsteht durch die Kanten „ $I \rightarrow V^* \rightarrow VI \rightarrow VII \rightarrow I$ “ eine weiterer negativer Feedbackzyklus, bestehend aus drei agonistischen und einer antagonistischen Kante. Dieser wirkt auf die Variable *Zeitersparnis durch fiete* inhibitorisch (hemmend) gegen den positiven „ $I \rightarrow II^* \rightarrow IV \rightarrow VII \rightarrow I$ “ Zyklus. Je nachdem wie stark die einzelnen Beziehungen sind, inwieweit die Lehrkraft die Anzahl der Einsätze fortschreitend erhöht und damit die Feedbackzyklen ermöglicht, wird die *Zeitersparnis* eher höher oder niedriger als im Ursprungszustand ausfallen. Ich gehe allerdings davon aus, dass das Verhältnis von 80% zu 20% noch genug Schüler:innen ermöglichen wird sich zu verbessern, sodass insgesamt doch der positive Zyklus wirkmächtiger und damit die *Zeitersparnis* bestehen bleibt. Die negative Technikfolge die bestehen bleibt, ist, dass bei falscher Einschätzung der Systemdynamik durch die Lehrkraft, die *Zeitersparnis durch fiete* gegen Null gehen, und die *Anzahl S, die individuelle Betreuung brauchen* steigen kann. Dies würde für die Lehrkraft nicht eine Entlastung, sondern sinnlosen Mehraufwand durch das Aufgabendesign und erhöhten Bedarf nach Betreuung, also ebenfalls Aufwand, bedeuten. Die Schüler:innen wären dann womöglich stärker als zuvor gespalten in diejenigen, die auch ohne viel sinnvolles Feedback durch fiete gut mit dem Unterrichtsstoff zurechtkommen und diejenigen die durch kontraproduktives Feedback einen höheren Aufwand haben und mehr Betreuung brauchen um wieder aufzuholen.

6 Zweite Gedanken

Dies ist die zweite Phase der Analyse des Wirkungsgefüge und zugleich die letzte der TSIA. Hier sollen durch Hineinversetzen in die Perspektive der Akteure konstruktive Handlungsvorschläge entwickelt werden, welche die negativen Technikfolgen bestmöglich umgehen, verhindern oder erträglich machen. Betrachten wir zuerst die Position der Lehrkräfte. Für sie ist es essenziell ein wirksames Unterrichtskonzept zur Integration des Feedbacktutors zu entwickeln. Sie müssen genügend Zeit einplanen, um bei durch die Nutzung von fiete entstandene Rückfragen beantworten zu können. Unter dieser eingeplanten Zeit darf jedoch nicht die Qualität des Aufgabendesigns leiden, falls noch weitere Aufgaben erstellt werden.

Aus Sicht der Schüler:innen empfiehlt es sich die Aufgaben möglichst gewissenhaft zu bearbeiten, damit sie auch tatsächlich zu 80% sinnvolles Feedback bekommen und sich dadurch verbessern können. Sobald sie allerdings merken, eher irreführende oder unsinnige Anweisungen und Korrekturen zu bekommen, sollten sie dies der Lehrkraft melden und sich nach besserem Feedback erkundigen.

Die Entwickler haben ein Interesse daran, dass der positive Feedbackzyklus

aufrecht erhalten bleibt, damit die *Anzahl der Einsätze von fiete* weiter steigt. Sie können dies erreichen, indem sie die Lehrer:innen belehren wie fiete am Besten in den Unterricht implementiert werden kann und ihnen die Nutzung so leicht und verständlich wie möglich machen. Außerdem ist es in ihrem Interesse die Quote von 80% weiter zu erhöhen und dabei die Utopie der 100% anzustreben.

7 Fazit

Diese TSIA hat die Technikfolgen des Einsatzes von fiete.ai, einem KI-gestützten Feedbacktutor, im Schulunterricht untersucht. Es hat sich gezeigt, dass ein Verständnis, sowohl von den Fähigkeiten zeitgenössischer generativer Künstlicher Intelligenz als auch deren Schwächen und Unzulänglichkeiten, wertvoll im Umgang mit durch KI entstehenden Dynamiken und emergenten Phänomenen ist. Durch die Erstellung zweier Wirkungsgefüge und der Differenzierung beider Fälle, konnten Hypothesen über die zugrundeliegende Logik des Systems aufgestellt werden und in der Folge, an deren Motivationen orientiert, Handlungsempfehlungen für die einzelnen relevanten Akteure ausgesprochen werden.

8 Anhang