

Abstrakt

STRATEGISCHE BRIEFINGS FÜR DIE GESTALTUNG DES 21. JAHRHUNDERTS

NEVER ALONE

ZUR GESTALTUNG VON WIRTSCHAFT UND GESELLSCHAFT
IM ZEITALTER DER ALLGEGENWÄRTIGEN GENERATIVEN
KÜNSTLICHEN INTELLIGENZ.



FUTURE SOCIETY ASSOCIATION

W.I.R.E.

EXECUTIVE SUMMARY

Nach einer Dekade dystopischer und utopischer Prognosen zur Rolle von künstlicher Intelligenz (KI) hat sich der Hype mit dem Fokus auf die Corona-Pandemie, den Ukraine-Krieg und die globale Rezession abgekühlt. Zurecht: Studien zeigen, dass kaum eines der angepriesenen KI-Tools einen wirklichen Nutzen bei der Bekämpfung von Corona gebracht hatte.¹ Auch der Traum des autonomen Fahrens, im ursprünglichen Business Plan von UBER als Standard bis 2020 angekündigt, ist nicht mehr fixer Bestandteil der offiziellen Zukunftserwartung: Nachdem Elon Musk über mehrere Jahre wiederholt den Einzug von selbstfahrenden Autos für das «nächste Jahr» angekündigt hatte, ohne dass sich die Prognosen bewahrheitet hatten haben viele Hersteller haben die Entwicklung von Robocars angesichts von Kosten und Komplexität de-priorisiert.² Zwar sehen viele Unternehmen – auch in der Schweiz – künstliche Intelligenz als zentralen Innovations- und Effizienztreiber.³ Konkrete, nutzbare Anwendungsfelder für den Alltag gab es jedoch eher selten, und die Qualität von KI-basierten Tools wie Chatbots im Alltag entpuppte sich als überschaubar – bis am 22. November 2022.

Die Lancierung des ChatGPT-Bots, der vom US-Unternehmen OpenAI als umfangreiches Sprachmodell entwickelt wurde, markiert nun einen vermeintlichen Wendepunkt in der Techwelt mit weitreichenden Folgen für Wirtschaft und Gesellschaft. Scheinbar aus dem Nichts steht ein künstlich intelligentes Assistenzsystem zur Verfügung, das als Wunderwaffe bei Unternehmenskommunikation, Schulaufsätzen, Liebesbriefen oder der Gestaltung von Bildern oder Produkten Aufgaben übernimmt, die bisher den Menschen vorbehalten waren. Die angekündigte Prognose – die Auslagerung von Schlüsselaufgaben an Maschinen – scheint in Reichweite und hat zu einer enormen Dynamik geführt, die den Hype um künstliche Intelligenz neu – und in noch grösseren Dimensionen – entfacht hat.

Schätzungen zufolge handelt es sich bei ChatGPT um den am schnellsten wachsenden Internetdienst aller Zeiten, der im Januar 2023 nur zwei Monate nach dem Start, bereits 100 Millionen Nutzer erreichte. Das «fear of missing out»-Narrativ ist auf Bühnen und Business-Blogs längst wieder allgegenwärtig und prognostiziert Unternehmen, die nicht sofort mit dem KI-Schnellzug mitziehen, eine Komplet-Disruption, sprich, das Verpassen des Anschlusses an die Zukunft.

Selbst das Monopol der mächtigen Google-Suchmaschine scheint durch die Integration von generativer KI in das Microsoft-Tool Bing bedroht. So hat der gescheiterte Versuch einer Schnell-Lancierung eines Konkurrenzprodukts zu Börsen-Kursschwankungen in Milliardenhöhe geführt. Auch die Politik ist im ChatGPT-Fieber. So hatte der rumänische Ministerpräsident im Frühjahr 2023 angekündigt, eine künstliche Intelligenz als Beraterin zu nutzen, die Wünsche der Bevölkerung in sozialen Medien analysiert und Handlungsempfehlungen abgibt. Umgekehrt hat Italien den Zugang zum Tool gesperrt, nachdem eine Gruppe von Tech-Experten rund um Elon Musk eine Entwicklungspause von 6 Monaten gefordert hat. Dies ist insofern erstaunlich, als dass künstliche Intelligenz ja eine Technologie ist, die seit Jahrzehnten in Entwicklung steht und regelmässig durch Experten von Unternehmen, Wissenschaft und Politik analysiert wird. Die möglichen gesellschaftlichen Risiken und darauf ausgerichtete Strategien müssten reflektiert und vorformuliert sein. Der neu entflammte, hektische Wettbewerb um die Vormacht im Internet zeigt indes, dass viele der Grundsatzfragen weder in den Teppichetagen der grossen Internetkonzerne noch in den Strategiezentralen der Regierungen systematisch durchdacht sind.

Mit Blick auf die enormen Erwartungen, Hoffnungen und Ängste, die durch die Lancierung des praxistauglichen KI-Tools geschürt werden, gilt es, kritische Einschätzungen vorzunehmen, die Organisationen und leitende Personen zu einem gewinnbringenden und verantwortungsvollen Umgang mit der nächsten Generation von KI befähigen. Klar ist, KI ist seit 2022 ein fester Bestandteil des Alltags aller Menschen und wir müssen jetzt festlegen, welche Rolle die Technologie dabei spielen soll, akut beim Einsatz in Schulen, für Marketing und Kommunikation – und langfristig in Bezug auf die grundlegende Rolle von lernenden Maschinen in Wirtschaft und Gesellschaft.

Dieses Briefing basiert auf einem Austausch mit Partnern der «Future Society Association» und einer vertieften Analyse durch den Think Tank W.I.R.E. Es verschafft ein Basiswissen, aufgrund dessen Stellung bezogen werden kann, zeigt Chancen und Herausforderungen von KI und ChatGPT auf und liefert konkrete Vorschläge für die Diskussion darüber, was zu tun ist – heute wie für die langfristige Zukunft.

¹ <https://www.technologyreview.com/2021/07/30/1030329/machine-learning-ai-failed-covid-hospital-diagnosis-pandemic/>

² <https://theprint.in/economy/ford-volkswagen-pop-the-automated-vehicle-bubble-with-exit-from-self-driving-startup-argo/1183780/>

³ https://www.thewire.ch/data/files/Artificial_Intelligence_W.I.R.E._2022_Digital.pdf

GRUNDLAGEN

FUNKTIONSWEISE GENERATIVER KI UND CHATGPT

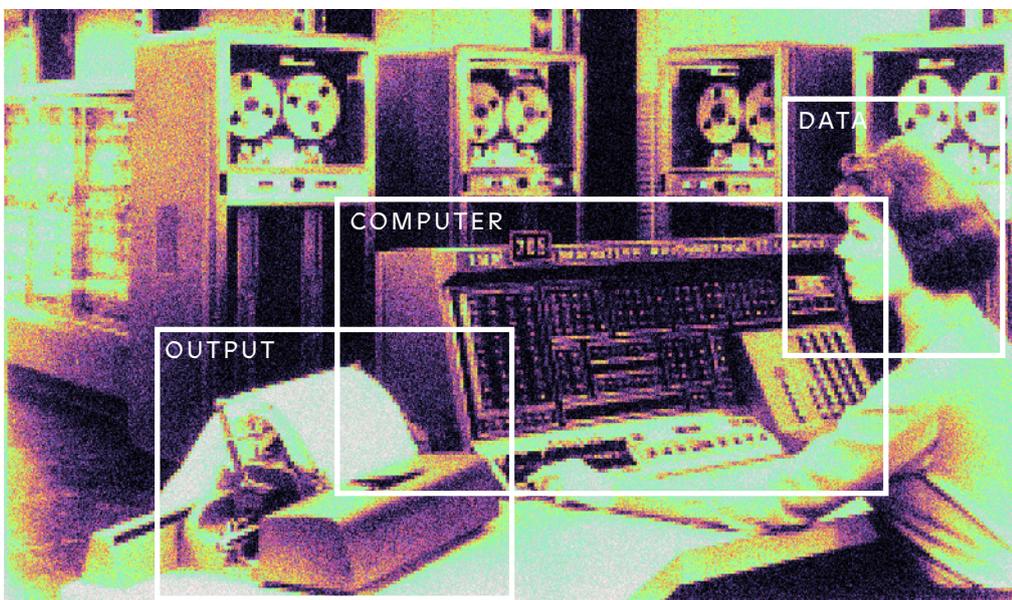
Die so genannte «Generative Künstliche Intelligenz» ermöglicht es, Inhalte wie Texte, Bilder oder Töne zu generieren, anstatt sie lediglich zu erkennen oder zu klassifizieren. Der mittlerweile bekannteste der generativen Algorithmen ist GPT (Generative Pre-trained Transformer), ein von OpenAI aus den USA entwickeltes Sprachmodell, das 2018 erstmals vorgestellt und laufend verbessert wurde.

Die Technologie ist nicht wirklich neu. Generative KI wurde in den 1960er Jahren mit Chatbots wie Eliza entwickelt. Aber erst 2014, mit der Einführung von «generativen adversarischen» Netzwerken oder GANs – einer Art maschineller Lernalgorithmen – konnte generative KI überzeugend authentische Bilder, Videos und Audios von echten Menschen erstellen. Die neueste Version von GPT und seine Konversationsversion – ChatGPT –, ist in der Lage, sehr plausible und kohärente Texte zu «generieren», die oft nur schwer von menschlich geschriebenen zu unterscheiden sind. Das Modell wurde mit dem Ziel trainiert, das nächste Wort vorherzusagen, und so Sätze oder ganze Textstücke zu produzieren. Die KI merkt sich entsprechend Muster und Strukturen der menschlichen Sprache und baut Inhalte Wort für Wort auf.

Um das System zu trainieren, wurde durch «Webscraping»⁴ ein enormer Datensatz aus Büchern, Wikipedia-Artikeln und anderen Textquellen erhoben und genutzt. Das aktu-

elle System verwendet solche Trainingsdaten bis 2021 und weiss entsprechend nichts über neuere Ereignisse. Der Bot ist in der Lage, Sachverhalte in einer neuen Form zu kombinieren, kann aber kein neues Wissen erzeugen. Das GPT-3 wurde mit Daten bis 2021 gefüttert; das Programm wusste entsprechend nichts über neuere Ereignisse. Mit GPT-4 liegt bereits eine erweiterbare Variante vor, deren Trainingsgrundlagen allerdings bislang unter Verschluss gehalten wurden. Zentral für eine Einschätzung der langfristigen Chancen und Risiken ist aber nicht primär der effektive Umfang der aktuell berücksichtigten Daten und Bilder, sondern die grundsätzlichen Potentiale und Grenzen der Technologie.

Dem GPT-Modell liegen auch die so genannten «DALL-E»⁵ Anwendungen zugrunde. Sie wurden mit Millionen im Internet verfügbaren Bildern trainiert und können basierend auf Textbeschreibungen Bilder generieren. Dies tun sie, indem Wörter aus einer Aufgabestellung in eine Anordnung von Pixeln übersetzt werden, die den statistischen Wahrscheinlichkeiten entsprechen, auf denen das System basiert. So können Bildelemente neu kombiniert und in beliebigen Illustrationsstilen visualisiert werden. Ein singendes Auto im Stil von Salvador Dalí, als 8-Bit Darstellung oder als fotorealistisches Portrait, kann so per Klick in Sekundenschnelle gestaltet werden.



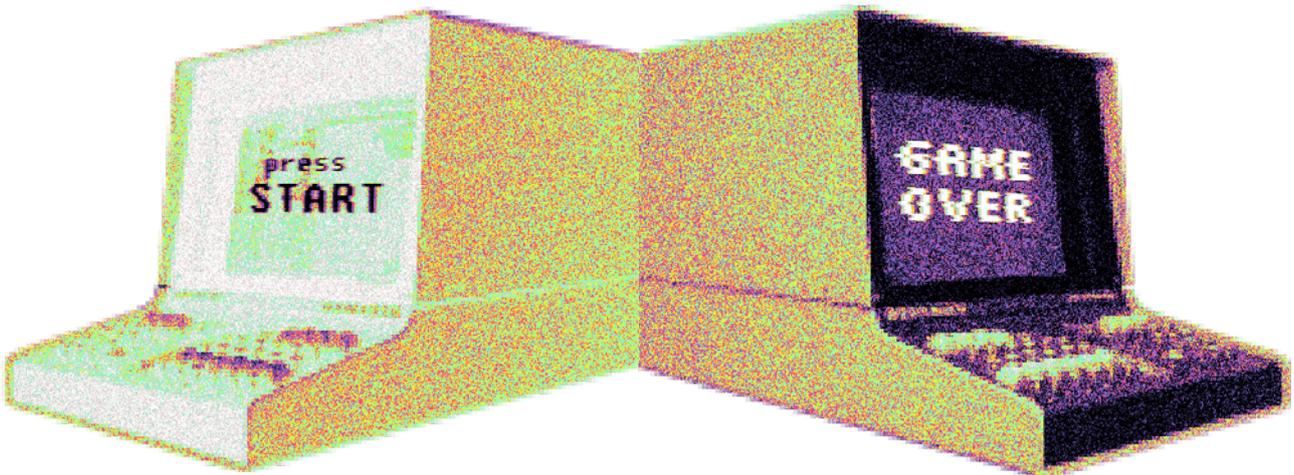
⁴ das automatische Auslesen von Inhalten einer Website.

⁵ Inspiriert durch den Roboter Wall-E aus dem gleichnamigen Pixar-Film und dem spanischen Surrealisten Salvador Dalí.

EINORDNUNG

ANWENDUNGSFELDER UND GRENZEN GENERATIVER KI

Das Einsatzspektrum von GPT-Systemen ist industriübergreifend und betrifft so gut wie alle Lebensbereiche. Um ein realistisches Verständnis zu schaffen, in welchen Aufgabenfeldern generative KI-Systeme effektiv einen nachhaltigen Mehrwert liefern und welche Herausforderungen dabei berücksichtigt werden müssen, braucht es eine differenzierte Analyse der Potenziale und Limitationen.



POTENZIALE

Erstellen von Texten

Generative KI kann unterschiedliche Textarten erstellen, vom Nachrichtenartikel über den Schulaufsatz und Gedichte bis zum Blogbeitrag. Ebenso können allgemeine Such- oder Kundenanfragen beantwortet werden. Zudem sind diese Bots in der Lage, Texte zusammenzufassen, zu korrigieren oder zu übersetzen.

Definieren von Prozessen

Die Technologie ermöglicht das Definieren von Abläufen für unterschiedliche Aufgaben in Organisationen, z.B. für Kommunikation, Kundendienst oder Bewerbungsprozesse. Ebenso können Reise- oder Trainingspläne zusammengestellt werden.

Unterstützen beim Programmieren

ChatGPT kann beim Erstellen von Computerprogrammen helfen, indem Basiscodes entwickelt werden, die später durch Menschen ausdetailliert werden. Dabei kann das Programm auch Fehler in bestehenden Programmen identifizieren.

Erzeugen von Bildern und Designs

Ein wesentliches Anwendungsfeld ist das Erstellen von photorealistischen Bildern als so genannte «deep fakes», die für künstliche Bildwelten, Avatare oder Marketingmaterial genutzt werden können. Ferner kann generative KI basierend auf bestehenden Illustrationsstilen Bilder erstellen oder den Stil von Künstlern imitieren. Darüber hinaus können Musikstücke komponiert oder vielfältige

LIMITATIONEN

Grenzen des Referenzierens

Weil ChatGPT Antworten basierend auf statistischen Wahrscheinlichkeiten angibt, die auf grossen Datenquellen beruhen, können einzelne Quellen, die für die Ausgabe verwendet wurden, nicht referenziert werden. Dies führt zu Intransparenz und verunmöglicht die Einhaltung des Urheberrechts. Zwar sind neuere Versionen in der Lage gewisse Referenzen abzubilden, eine vollständige Darstellung von Quellen dürfte aber auch langfristig beim Training auf nicht vollständig strukturierten Daten limitiert sein.

Unwahrheiten und Verzerrungen

Die Qualität der Ausgaben von ChatGPT basiert auf den Trainingsdaten, die zur Entwicklung des Modells verwendet werden. Wenn die Grundlagendaten falsch sind oder einseitige Sachverhalte abbilden, entstehen «verzerrte» Aussagen. Dies kann zu geschlechterdiffamierenden oder rassistischen Aussagen führen, vor allem aber Kulturräume bevorzugen, von denen die Daten stammen. Zwar werden hierfür Filter und menschliche Verifizierungen eingesetzt, eine inhärente Einschränkung bleibt jedoch bestehen.

Einschränkungen bei komplexen Aufgaben

ChatGPT liefert fehlerbehaftete Ausgaben, z.B. beim Zusammenfassen von Texten aber auch beim Interpretieren von Hintergrundinformation, beim Referenzieren auf subtile oder nicht unmittelbar in Bezug stehende Entwicklungen oder beim Verstehen

Varianten für Produktdesigns erstellt werden, von denen das passendste ausgewählt werden kann.

Vermitteln von Wissen und Kompetenzen

Der Einsatz von Bots kann dazu genutzt werden, Wissen zu vermitteln – für Bildung oder für journalistische Zwecke. Dabei können beispielsweise lokale Geschehnisse abgedeckt werden, über die aus wirtschaftlichen Gründen nicht berichtet wird. Oder strukturschwache Regionen können mangelnde Fachkräfte ersetzen. Nicht zuletzt kann ChatGPT auch als massgeschneidertes Lernsystem genutzt werden.

oder Erzeugen von mehrdeutigen Aussagen im ganzen Spektrum von Ironie bis Zynismus. Auch das Zusammenfassen von anspruchsvollen Texten die das Verstehen mehrerer Perspektiven erfordern stösst an Grenzen und erzeugt systematische Fehlerarten, die mit der Zeit natürlich besser werden, allerdings dennoch kontrolliert werden müssen, wenn es sich um relevante Entscheidungsgrundlagen handelt

Grenzen der Personalisierung

Die Ausgaben der generativen KI basieren auf standardisierten Inputdaten und können nur so weit personalisieren, als dass entsprechende Referenzen vorhanden sind, die statistisch ausreichen. So können Gestaltungsstile des Impressionismus oder Surrealismus interpretiert werden, nicht aber eine wirklich individualisierte Art des Formulierens oder Zeichnens.

Kosten für Energie

Der Rechenaufwand für künstliche Intelligenz ist hoch. Während die aktuellen Versionen von ChatGPT auf Testbasis die Kosten erst ansatzweise an Nutzer weiterverrechnen, dürften flächendeckende Anwendungen zu entsprechend höheren Ausgaben und Energiekosten führen.

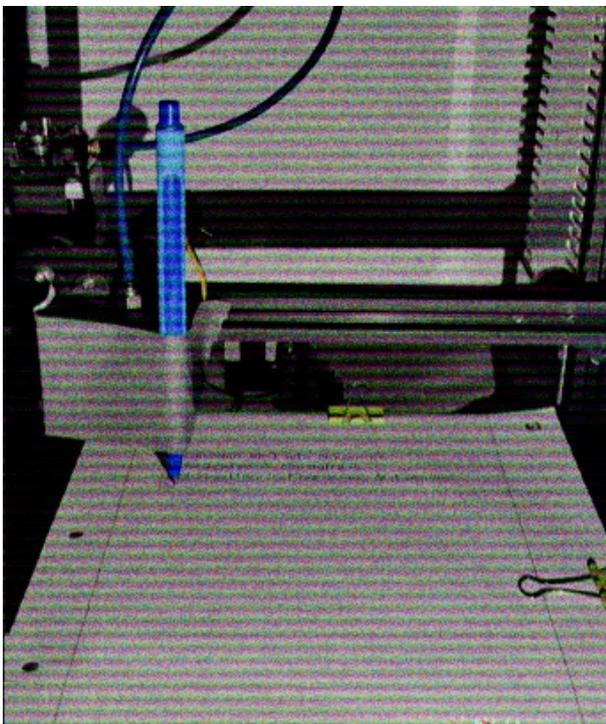


ERSTELLEN VON KÜNSTLICHEN FOTOS MENSCHEN MIT HOMOGENEN GESICHTSZÜGEN, UND ANATOMISCHEN FEHLERN

https://mymodernmet.com/ai-fake-party-photos/?fbclid=IwAR1DPyZnCLz1VYqYsX60H1EL_LZmjWJIRq0I2zOW7_yH7x55EyKLMhdovNo



**EINSATZ FÜR DAS ERSTELLEN EINER «UNENDLICHEN» SEINFELD SERIE
ENDLOSSCHLAUFEN VON ÄHNLICHEN KONVERSATIONEN**
<https://screenrant.com/seinfeld-ai-twitch-stream-jerry-seinfeld-roast-response/>



**AUSLAGERN EINES SCHULAUFSATZES MIT
VERSCHRIFTLICHUNG DURCH PLOTTEREND
PERFEKTIONIEREN VON PLAGIATEN**
<https://mymodernmet.com/chatgpt-3d-printer-homework/>



**AUTOMATISCH ERZEUGTE GESTALTUNGS-
LÖSUNGEN FÜR PRODUKTE AUSWÄHLEN
ANSTATT SELBST ENTWICKELN VON IDEEN**
<https://formlabs.com/ch/blog/generative-design/>

WIRTSCHAFT UND GESELLSCHAFT IM ZEITALTER DER GENERATIVEN INTELLIGENZ

Basierend auf den Potenzialen und den Limitationen lassen sich Prognosen ableiten, die die künftigen wirtschaftlichen und gesellschaftlichen Rahmenbedingungen prägen werden und als Planungsgrundlage dienen. Diese Prognosen sind nicht einseitig, sondern durch übergeordnete Spannungsfelder geprägt, die ein differenziertes Verständnis erfordern. Mit Blick auf die langfristige und grundlegende Gestaltung der Arbeitswelt und des privaten Lebens gilt es, Grundsatzfragen zu klären, die nicht mehr länger zu Gunsten von kurzfristigen, reaktiven Verhaltensanpassungen herausgeschoben werden dürfen.



1. Übernahme von bislang menschlichen Aufgaben – gleichzeitige Entstehung neuer Job profile und Verantwortungen

ChatGPT und künftige generative Bot-Systeme ordnen sich in die Charakteristik industrieller Revolutionen ein, die seit dem 19. Jahrhundert dazu geführt haben, dass Aufgaben, die einst durch Menschen ausgeführt wurden, teilweise oder vollständig an Maschinen ausgelagert werden. Mit dem Fokus auf Kommunikation, Prozessgestaltung und Design erfasst dieser Prozess nun eine breite Palette von Aufgaben, die bislang Menschen vorbehalten waren. Das Wegfallen von traditionellen Jobs, die durch das neue Spektrum der Automatisierbarkeit abgedeckt werden, ist höchst wahrscheinlich. Angesichts der Fehlerhaftigkeit und Grenzen beim Ausführen komplexer Aufgaben zeichnet sich aber gleichzeitig eine Aufwertung menschlicher Aufgaben im «Fine-Tuning», Personalisieren oder der Qualitätskontrolle ab. Mit der Verbreitung von generativer KI wird auch die Nachfrage nach diesen Steuerungsaufgaben zunehmen. Zudem werden sich Ergebnisse von KI immer an statistischen Mittelwerten orientieren, unerwartete, neuartige Ideen und Lösungen sind nicht oder nur bedingt Teil der Potentiale von lernenden Maschinen.

Konsequenz: Aufbau der künftigen Kompetenzen abseits von repetitiven und prozessualen Aufgaben,

Klärung der sinnvollen Einsatzorte für generative KI und Definieren von Kontrollinstanzen bei systemrelevanten Aufgaben.

2. Flutung des Internets mit künstlich generierten Inhalten – Vertrauensverlust in digitale Inhalte und Abspaltung eines qualitätsgeprüften Webs

Die breite Verfügbarkeit von generativer künstlicher Intelligenz wird den Umgang mit Wissen – und damit auch das Internet – weitreichend verändern. Automatisch generierte Texte und Bilder werden nicht nur für Schulessays oder journalistische Aufgaben genutzt, sondern künftig für einen wesentlichen Bestandteil der Inhalte des World Wide Web verantwortlich sein. Als Folge wird es für Nutzer immer schwieriger zu unterscheiden, welche Inhalte oder Social-Media-Profile von Menschen erstellt und welche Bot-generiert sind. Es wird bei Texten, genauso wie bei Bildern, mehr Aufwand erfordern um zu unterscheiden, was real und was künstlich ist. Die konstante Suche nach Authentizität führt in der Folge zu einem weitreichenden Vertrauensverlust in die virtuelle Welt und erhöht den Aufwand des Suchens um ein Vielfaches. Auch automatisierte Plagiatsdetektoren werden die Problematik nicht lösen, da sie selbst

fehlerbehaftet sind. Als Reaktion darauf werden qualitätsgeprüfte und authentische Inhalte künftig in einem kostenpflichtigen Teil des Internets verfügbar sein. Dies unterläuft allerdings den ursprünglichen basisdemokratischen Grundsatz des World Wide Webs.

Konsequenz: Der Zugang zu brauchbaren und glaubwürdigen Inhalten wird aufwändiger. Rückkehr zu physischem Austausch als Grundlage für Vertrauen. KI-erzeugte Inhalte und Bilder müssen transparent erkennbar sein.

3. Erweiterte Grundlagen für Kreativität – aber noch mehr Homogenisierung mit Verlust von Margen

Die Fähigkeit, Inhalte mit Textarten und Sprachstilen oder Bildelemente mit Illustrationsstilen oder Referenzen von Künstlern beliebig zu kombinieren, eröffnet neue – und effiziente – Möglichkeiten für kreative Anwendungen, bei denen Texte, Illustrationen oder auch Produktdesigns auf Knopfdruck erzeugt werden können. Aufgrund des Aufbaus auf bestehenden Inhalten oder Illustrationen und der Ausrichtung auf statistische Mittelwerte bringt der Einsatz von generativer KI allerdings eine Homogenisierung mit sich, bei der sich die Ergebnisse für Unternehmenskommunikation, Bebilderung, Produktdesign oder Architektur immer mehr angleichen. Die Folge ist ein wachsender Druck auf Margen und die Bezahlbereitschaft für künstlich generierte Inhalte, was die Notwendigkeit für (globale) Skalierung und plattformbasierte Geschäftsmodelle weiter in die Höhe treibt. Als Gegenpol wird echte Differenzierung folglich noch stärker von menschlichem Input abhängig. Aufgrund der hohen Relevanz der nutzbaren Daten kann sich daraus ein neue Wettbewerbsgrundlage oder ein Standortfaktor ergeben: Geschäftsmodelle oder Rechtsräume, die über qualitativ hochstehende und nutzbare Trainingsdaten verfügen, sind in einer KI-basierten Wirtschaft im Vorteil.

Konsequenz: Die zunehmende Angleichung von KI-basierten Resultaten führt zu Gegentrends und der Suche nach Vielfalt und echter Authentizität – auch als Grundlage für künftige Margen.

4. Mehr Convenience und Chancengleichzeit – aber eine rückläufige Entscheidungsfähigkeit und Risiko von Fremdbestimmung

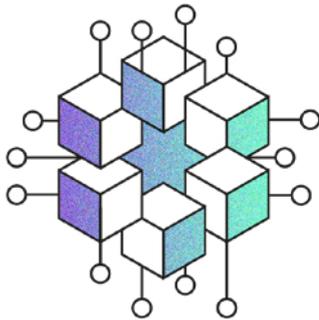
Der freie Zugang zu generativen KI-Assistenzsystemen erlaubt es immer mehr Menschen, einfache Aufgaben auszulagern: vom Schreiben eines Briefs, über Social-Media-Posts bis zum Erstellen eines ersten Brainstormings. Dies führt zu einer enormen Erleichterung beim Überwinden von alltäglichen mühsamen Tätigkeiten. KI-basierte Bots ermöglichen es zudem, Menschen ohne entsprechende Ausbildung oder Kompetenzen dazu zu befähigen, strukturierte Recherchen durchzuführen, Texte zu formulieren oder Illustrationen zu erstellen. Diese «Demokratisierung» erhöht die Chancengleichheit in der Bevölkerung. Darüber hinaus können KI-basierte Assistenzsysteme Leistungen in strukturschwachen Gegenden erbringen, die bislang nicht erschwinglich waren. Mit einer konstanten Abgabe dieser Tätigkeiten an digitale Assistenzsysteme kommt es jedoch mittel- bis langfristig zu einem Verlust von Kompetenz und Selbstwirksamkeit respektive zu mehr Fremdbestimmung. Ohne regelmässige Übung geht die Fähigkeit des Formulierens, des Strukturierens von Inhalten und damit auch kritisches oder kreatives Denken über breite Bevölkerungsschichten verloren.

Konsequenz: Die «Kosten von Convenience» in Form von Kompetenzverlust und rückläufiger Entscheidungsfähigkeit sind mehrheitlich nicht bewusst. Organisationen und Bildungsinstitutionen sind gefordert, diese Fähigkeiten zu stärken und bei wichtigen Entscheidungen bewusst auf die Steuerung durch KI zu verzichten.

SCHLÜSSELFRAGEN

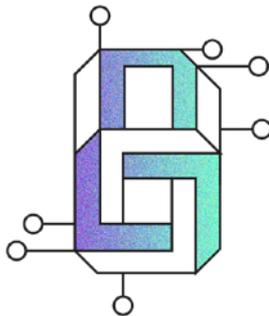
KÜNFTIGER UMGANG MIT KÜNSTLICHER INTELLIGENZ

Mit Blick auf die langfristigen Rahmenbedingungen von Wirtschaft und Gesellschaft gilt es, mehrere Schlüsselfragen zu klären, die nicht primär als reaktive Positionierung gegenüber der wachsenden Leistungsfähigkeit von KI zu beantworten sind, sondern darauf aufbauen müssen, wie wir künftige Interaktionen, unser eigenes Selbstverständnis und die Gestaltung einer wünschenswerten Gesellschaft definieren.



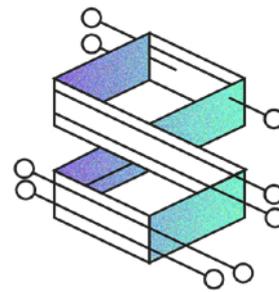
Wie differenzieren?

Die Macht der Standardisierung wird mit generativen KI-Bots weiter zunehmen. Für grosse, plattformbasierte Geschäftsmodelle stellt sich die Frage, wodurch sich Dienstleistungen und Produkte wirklich definieren, wenn künstliche Intelligenz flächendeckend eingesetzt wird.



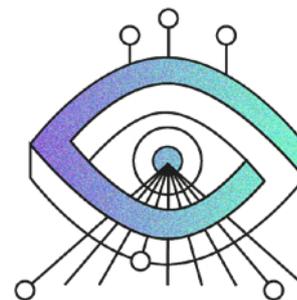
Wie Vertrauen bilden?

Mit der rasanten Verbreitung von KI-generierten Inhalten und Bildern wird es zunehmend schwierig zu erkennen, ob ein Foto real oder ein Text oder eine Nachricht von einem Menschen verfasst wurden. Der Aufwand, Täuschungen zu erkennen, nimmt zu; Vertrauensbildung im virtuellen Raum erfordert neue Grundlagen, die entweder durch Echtheitszertifikate oder einen Gegentrend zu realen Begegnungen definiert werden.



Wie geistiges Eigentum schützen und nutzbar machen?

ChatGPT nutzt aktuell das gesammelte Wissen des westlichen Kulturraums basierend auf Wikipedia-Einträgen und anderen Informationen, die aufbereitet werden, allerdings ohne Quellen zu referenzieren oder Urheberrechte zu entschädigen. Mit den ersten regulatorischen Einschränkungen wegen Datenschutz wie in Italien oder Klagen von Rechteinhaltern wird die zentrale Frage lauten, ob wir es als Gesellschaft erlauben, dass KI kostenloses Wissen nutzt und (zunehmend kostenpflichtig) weitergibt.



Wie Kompetenzen und Selbstwert erhalten?

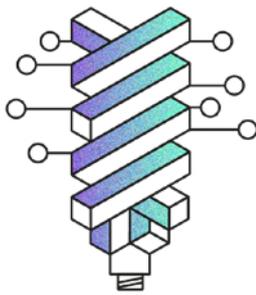
Mit der Verlockung des einfachen Auslagerns droht ein schleicher Verlust von Kompetenzen und Selbstwert. Um dem vorzubeugen gilt es, aus gesellschaftlicher Perspektive aktiv zu definieren, wo menschliche Kompetenz entscheidend ist und wie wir die Selbstwirksamkeit langfristig erhalten wollen.

HANDLUNGSFELDER

EMPFEHLUNGEN IM UMGANG MIT GENERATIVER KI

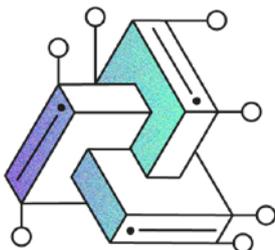
Organisationen müssen Potenziale und Limitationen generativer KI systematisch evaluieren und Massnahmen ergreifen, um wettbewerbsfähig zu bleiben und dabei das Vertrauen von Mitarbeitern und Kunden zu stärken. Gleichzeitig gilt es, die Fundamente zu schaffen, um generative KI-Lösungen produktiv und sicher in den Alltag zu tragen. Die Erkenntnisse der bisherigen digitalen Transformation zeigen, dass der klassische, von Mark Zuckerberg propagierte Ansatz «move fast and break things» nicht zu nachhaltig erfolgreichen Lösungen oder profitablen Geschäftsmodellen führt. Das Berücksichtigen von gesellschaftlichen Anforderungen und Herausforderungen muss die Grundlage sein, auf der künftige Innovationen in den Markt getragen werden. Dafür braucht es Unternehmen, die auf ein gesamtheitliches Verständnis von Innovation setzen, es braucht Universitäten, die die Grundlage für entsprechende Technologien legen, und es braucht vor allem vorausschauende Regulierungen, die technologieneutral aber mit Fokus auf Fairness und Vielfalt die rechtlichen Grundlagen schaffen.

Hierfür können sechs Aufgabenfelder definiert werden, die angegangen werden müssen.



1. EIGENTUMSRECHTE KLÄREN:

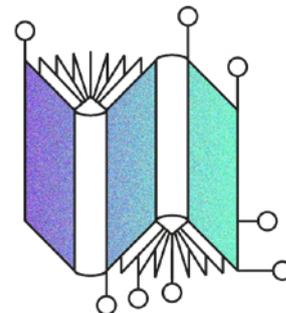
Festlegen einer Regelung zum Schutz von aber gleichzeitigem Zugang zu geistigem Eigentum, die für das Training von KI-Systemen genutzt werden kann. Voraussetzung wären technologische Lösungen, die Quellen klar referenzieren können, alternativ ein Pool von Inhalten, deren Rechthehalter für eine breite Nutzung pauschal abgegolten werden.



2. DATENGRUNDLAGEN AUFBAUEN:

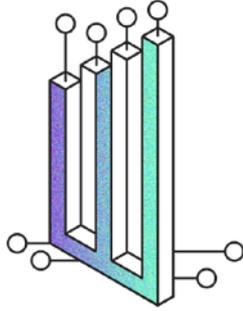
Schaffen einer breiten Basis aus relevanten, qualitätsgeprüften vielfältigen Datenquellen als Grundlage für das Trainieren von generativer KI, die von

Unternehmen und öffentlichen Organisationen oder der Öffentlichkeit genutzt werden können. Zentral dabei ist auch der Schutz von Persönlichkeitsrechten, die im besten Fall von Nutzern in «Datatrust»-Lösungen für definierte finanzielle oder andere Gegenwerte freigegeben werden. Hierbei können künftige Aufgaben eines datenorientierten «Service Public» entstehen.

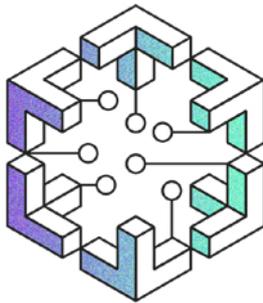


3. NUTZERKOMPETENZEN TRAINIEREN:

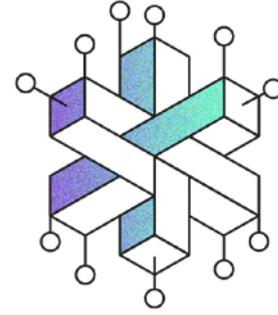
Fördern von Kompetenzen, die Nutzer in eine Position bringen, KI-Tools sinnvoll einzusetzen und Risiken in Bezug auf Falschaussagen, Diskriminierung, Unterlaufen von IP-Rechten für sich selbst oder für Dritte zu minimieren. Vermitteln von «Computational Thinking» und erklären, wie Algorithmen funktionieren, sowie Erkennung von künstlich erzeugten Bildern oder Texten üben. Job-Profil entwickeln, Aufbau von Test-Umgebungen und Trainings für kritisches Denken und quellenbasiertes Recherchieren. Schulen und Universitäten können generative KI als Arbeitstool nutzen und im Rahmen von Ausbildungen testen und lernen welche Ergebnisse brauchbar sind und welche nicht.



4. KRITISCHE OUTPUTS REGULIEREN:
Identifizieren von ethisch kritischen Ergebnissen von KI, zum Beispiel Diskriminierung von Konversationen, Texten oder bei HR-Prozessen. Rechenschaftspflicht der Unternehmen für die Entscheidungen ihrer KI-Systeme. Dabei gilt es nicht auf der Ebene von Technologien oder Software zu regulieren und Freiheitsgrade einzuschränken, sondern die Resultate in Bezug auf kritische Outputs zu analysieren und die Algorithmen daraufhin einzuschränken oder die Prozesse zu verbessern.



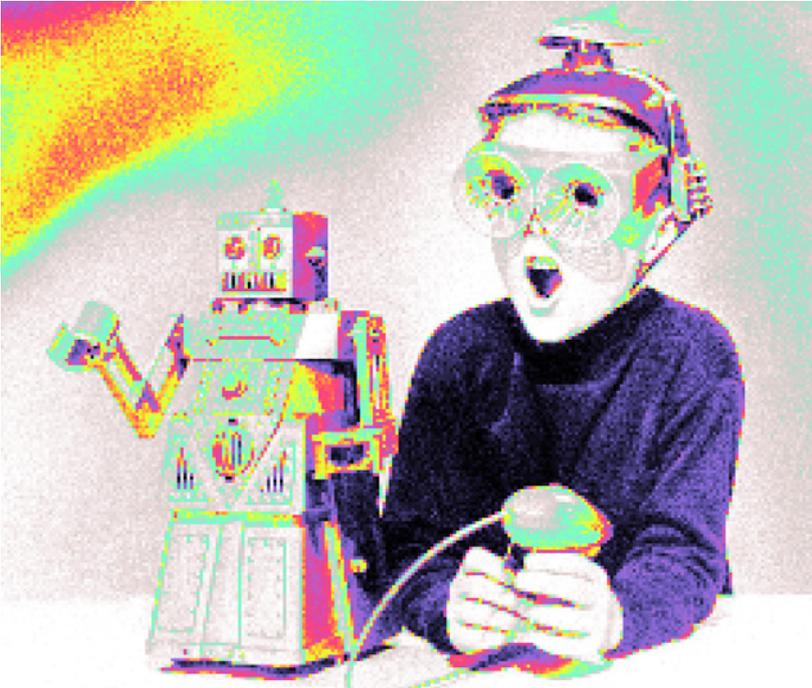
5. TRANSPARENZ SCHAFFEN:
Deklarieren von KI-Anwendungen bei der Veröffentlichung von Software, Texten und Visualisierungen, insbesondere bei Anwendung von Foto-Filtern in sozialen Medien, bei wissenschaftlichen Arbeiten und Marketing. Chance für Unternehmen durch Fördern von Selbstdeklaration und proaktive First-Mover-Transparenz. Hier braucht es eine vorausschauende Regulierung, gleichzeitig haben Organisationen die Chance durch eigene Richtlinien Vertrauen aufzubauen.



6. MENSCHLICHE KONTROLL- UND DIFFERENZIERUNGSINSTANZEN EINSETZEN:
Um einseitige und generische Ergebnisse von KI-Systemen zu verhindern, anzupassen und Vielfalt als Innovationsgrundlage zu erhalten, braucht es menschliche Intervention an neuralgischen Stellen von systemrelevanten Prozessen, aber auch für das Sicherstellen von neuen Ideen abseits der statistischen Mittelwerte. Unternehmen können sich durch eine Incentivierung von unerwarteten Ideen und Lösungen vor einer zunehmenden Angleichung schützen. Gleichzeitig stellen menschliche Interaktionen sowohl Innovationsfähigkeit als auch Vertrauen von R&D bis zu Kundenberatung sicher.

BOTS BLEIBEN – FÜR IMMER

Die Gestaltung der Zukunft, nicht die Einordnung in maschinelle Prozesse, wird zur Grundvoraussetzung für die Gestaltung einer humanen Zukunft.



deren Ergebnisse nicht mehr durch Menschen, sondern von anderen Bots beantwortet werden. So werden Vermarktungsbots künftig auf persönliche „Shopping“-Bots treffen, die ihre Nutzer vor der Flut von Angeboten schützen. Durch die Zunahme von solchen «Machine-Machine-Interactions» verschieben sich gleichzeitig die Rolle und das Kompetenzspektrum von Menschen, die mehr Kontroll- und Steuerungsaufgaben wahrnehmen. Der Einsatz von ChatGPT zieht bereits heute neue Software mit sich, die beispielsweise hilft, Plagiate zu identifizieren. Diese sind allerdings gleichermaßen fehlerbehaftet. Jüngst wurden Werke von Shakespeare oder die Bibel als KI-Erzeugnis angezeigt.⁷

Es ist klar, dass eine mächtige Technologie wie KI nicht ohne stringente Regulierung auskommt. Um so erstaunlicher und verantwortungsloser ist es, dass die zentralen Fragen und

Ansätze erst mit dem GPT-Hype in eine breite Öffentlichkeit – und wie es scheint – auf die Agenden der Politik gelangt sind. Die Technologie-Unternehmen haben die Chance hier genauso verpasst, im Hintergrund frühzeitig Konzepte zu entwickeln, die im richtigen Moment aus der Schublade gezogen werden können.

Nun gilt es, aus diesen Versäumnissen zu lernen und endlich eine systematische Früherkennung zu installieren, die sowohl für die Politik wie auch die Verantwortlichen von privaten und öffentlichen Organisationen verbindlich in konkrete Handlungsfelder übersetzt werden muss.

Final entscheidend dabei ist aber nicht die Leistungsfähigkeit der Maschinen, sondern das aktive Definieren der Rolle, die wir uns Menschen selbst geben – und zutrauen wollen. Dazu gehört ein aktiver Entscheid über das Ausmass an Irrationalität oder Emotionen, die wir zulassen, weil sie nicht nur oft für aussergewöhnliche Innovationen, sondern vor allem für kulturelle Errungenschaften verantwortlich sind. Die Qualität, die daraus entsteht, ist nicht auf eine Maximierung der Produktivität ausgerichtet, sondern auf Selbstwirksamkeit und weitere nichtquantifizierbare Grössen, die den Wert des Lebens und die Zukunft unserer Zivilisation ausmachen. Die eigentliche Antwort auf die Zukunft der künstlichen Intelligenz ist somit durch die Rolle definiert, die wir ihr geben.

Das Leistungsspektrum generativer KI wird in den nächsten Monaten und Jahren zweifelsohne weiter zunehmen. Der «Tipping Point» ist mit der breiten Verfügbarkeit der Technologie aber bereits jetzt erreicht. Die zentralen Fragen zum Umgang mit KI sind deshalb nicht abhängig von den nächsten Entwicklungsschritten, sondern betreffen das grundlegende Verhältnis zwischen Menschen und lernenden Maschinen. Klar ist: Auch in ferner Zukunft bleiben die Erzeugnisse von Maschinen «künstlich» und werden nicht von lebenden Menschen mit Gefühlen und sozialer Einbettung erzeugt. Sie basieren auf statistischen Wahrscheinlichkeiten. Bots verfügen nicht über ein «menschliches» Verständnis der Welt.

Künstliche Intelligenz wird darum weder heute noch in Zukunft ohne das Menschliche auskommen. Selbst wenn die Resultate noch verblüffender und effizienter sein werden, gilt es, über den Einsatz nicht allein auf Basis der technischen Machbarkeit, sondern auch der gesellschaftlichen Wünschbarkeit zu entscheiden. In einer Studie wurden Menschen befragt, ob sie Liebesbriefe zum Valentinstag durch ChatGPT erstellen lassen würden. 42% der Männer und 20% der befragten Frauen hatten geplant, dies zu tun, allerdings wären 50% aller Befragten verletzt, wenn diese Art Inhalt künstlich für sie selbst erzeugt würde.⁶

Mit der wachsenden Leistungsfähigkeit der KI-Systeme dürften weitere Anwendungen folgen,

⁶ <https://www.fastcompany.com/90846284/chatgpt-ai-love-letter-chatbot-valentines-day-poll>

⁷ <https://www.indiatoday.in/technology/news/story/oh-god-open-ai-tool-that-identifies-text-written-chatgpt-believes-bible-was-written-by-ai-2329163-2023-02-01>

ABSTRAKT ist eine Schriftenreihe des Think Tanks W.I.R.E., die vorausschauend Entwicklungen, Trends oder neue Technologien mit Relevanz für Wirtschaft und Gesellschaft kompakt aber differenziert und mit Weitsicht analysiert sowie Handlungsempfehlungen für die Gestaltung der Zukunft formuliert. Die Briefings basieren auf dem Konzept der gleichnamigen Buchreihe die W.I.R.E. über viele Jahre veröffentlicht hat und führen den Anspruch weiter Inspiration mit kritischem Denken zu verknüpfen.

Die Future Society Association (FSA) ist eine Initiative zur Neuverknüpfung von Wirtschaft und Gesellschaft mit dem Ziel, die Transformation einer digitalen und nachhaltigen Gesellschaft gemeinsam mit Wirtschaft, Wissenschaft und Politik zu gestalten. Die Basis legt die vorausschauende Identifizierung künftiger gesellschaftlicher Herausforderungen als Ergänzung zum bestehenden Technologiefokus vieler Innovationsprojekte, das Vermitteln von Wissen über neue gesellschaftliche Entwicklungen und das Verknüpfen von Akteuren zur Gestaltung der Gesellschaft von morgen. Ziel ist die Förderung von Geschäftsmodellen und Produkten, die gesellschaftliche Verantwortung integrieren und so langfristiges und stabiles wirtschaftliches Wachstum ermöglichen. Gleichzeitig fördert die FSA neue Perspektiven für eine wünschenswerte Zukunft, da ohne die Partizipation der Bevölkerung weder nachhaltige daten- noch kreislaufbasierte Geschäftsmodelle Akzeptanz finden.

Die FSA wurde 2020 durch den Think Tank W.I.R.E. initiiert und wird durch eine Gruppe von vorausschauenden Organisationen und Personen getragen.

GLOSSAR

Reinforcement learning

Bei der vielleicht einfachsten Form des Trainings, die es gibt, geht es beim verstärkenden Lernen darum, dem System jedes Mal Feedback zu geben, wenn es eine Aufgabe ausführt, so dass es daraus lernt, Dinge richtig zu tun. Dies kann ein langsamer und teurer Prozess sein, aber für Systeme, die mit der realen Welt interagieren, gibt es manchmal keinen besseren Weg.

Large-language models (LLMs)

Hierbei handelt es sich um eines der so genannten neuronalen Netze. Grosssprachige Modelle werden trainiert, indem sie mit Milliarden von Wörtern aus alltäglichen Texten aus Quellen wie Büchern, Tweets und allem, was dazwischen liegt, gefüllt werden. Die LLMs greifen auf all dieses Material zurück, um Wörter und Sätze in bestimmten Sequenzen vorherzusagen.

Generative adversarial networks (GANs)

Bei dieser Methode werden zwei neuronale Netze miteinander verbunden, um etwas Neues zu schaffen. Die Netze werden bei der kreativen Arbeit in der Musik, der bildenden Kunst oder beim Filmmachen eingesetzt. Einem Netzwerk wird die Rolle des Schöpfers zugewiesen, während ein zweites die Rolle des Markers übernimmt. Das erste lernt, Dinge zu schaffen, die das zweite gutheisst.

Symbolic AI

Es gibt sogar KI-Techniken, die sich von der Vergangenheit inspirieren lassen. Symbolische KI ist ein Ansatz, der die Idee verwirft, dass ein einfaches neuronales Netz die beste Option ist, und versucht, maschinelles Lernen mit sorgfältig strukturierten Fakten über die Welt zu kombinieren.

LITERATUR

M. Fouque, A. Calabrese, M. Diallo, N. Vey, J.-N. Monette, J.-P. Chevillet, & L. Besacier. (2021). HAL-03913837: The GPT-3 OpenAI Large Model: An Overview and Some Experiments. HAL. <https://hal.science/hal-03913837/>

S. Naushad, R. Kumar, & J. De Winter. (2022). ChatGPT for Next Generation Science Learning. ResearchGate. https://www.researchgate.net/publication/367281552_ChatGPT_for_Next_Gen...

J. De Winter & M. Farooq. (2022). Can ChatGPT pass high school exams on English Language Comprehension? ResearchGate. <https://www.researchgate.net/profile/Joost-De-Winter/publication/366659...>

J. De Winter. (2022). ChatGPT: Unlocking the Future of NLP in Finance. ResearchGate. https://www.researchgate.net/publication/367129318_ChatGPT_Unlocking_th...

H. Ismail, & M. Y. Kadhum. (2021). The study of the effect of using a language model in teaching writing skills. Iraqi Journal of Curriculum and Management Sciences, 8, 258-266. <https://journal.esj.edu.iq/index.php/IJCM/article/download/539/258>

M. Farooq & J. De Winter. (2021). The Future of Natural Language Processing: A Study of OpenAI GPT-3 Model. Social Science Research Network. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4312418

PARTNER



IMPRESSUM

© 2023 W.I.R.E.
www.thewire.ch
Konzept & Inhalte: W.I.R.E.
Gestaltung & Illustration: W.I.R.E.
Korrektur: Leube Translation and Language Services
Herausgabe: Juli 2023

TEILNEHMENDE DES LUNCH AND CREATE

Alexander Fleischer, Die Post
Gian Casutt, ETH Rat
Matthias Kiess, TBWA
Claudia Pletscher, Future Society Association
Roland Wittmann, W.I.R.E.
Stephan Sigrist, W.I.R.E.

