Studienprojekt I

Nachhaltigkeitsrisiken und -potenziale von KI: Eine vergleichende Literatur- und Modellanalyse

Studierende:	Fynnian Kolbe, Jamie Jentsch, Finn Luca Ritschel
Matrikelnummer:	77201605870, 77208395999, 77201740614
Ausbildungsbetrieb:	Flughafen Berlin Brandenburg GmbH
Studienjahrgang:	2023
Fachbereich:	Fachbereich 2 – Duales Studium
Studiengang:	Informatik
Modul:	IT3161 – Studienprojekt I
Betreuer Hochschule:	Prof. Dr. Dagmar Monett Díaz
Anzahl der Wörter:	8693 Wörter
Datum der Fertigstellung:	19.08.2025

Kolbe, Fynnian — Digital unterschrieben von Kolbe, Fynnian Datum: 2025.08.19 14:55:13 +02'00'

Jentsch, Jamie — Digital unterschrieben von Jentsch, Jamie Datum: 2025.08.19 14:52:51 +02'00'

Ritschel, Finn — Digital unterschrieben von Ritschel, Finn Datum: 2025.08.19 14:57:22 +02'00'

(Datum, Unterschrift Studierende)

Kurzfassung

Diese Studienarbeit befasst sich mit der Beziehung zwischen Künstlicher Intelligenz (KI) und Nachhaltigkeit. Dieses Thema gewinnt in letzter Zeit zunehmend an Bedeutung, einerseits durch das größere Umweltbewusstsein, andererseits durch den KI-Boom der letzten Jahre in Form von ChatGPT oder Gemini, um nur wenige Beispiele zu nennen. Hinter jeder Nutzung von KI steckt ein häufig unterschätzter Ressourcenaufwand, da diese mit enormen Mengen an Daten arbeiten. Diese Studienarbeit setzte sich zum Ziel, durch den systematischen Vergleich vieler Studien, die sich mit diesem Thema beschäftigen, einen umfassenden Gesamtüberblick zu schaffen. Um die Anzahl der Studien einzugrenzen, wurden mithilfe der PRISMA-Methode die relevantesten Studien herausgefiltert. Bei der Vergleichsstudie wurden die vier Kernpunkte „Stromverbrauch von KI“, „Emissionen durch KI“, „Reduzierung von Emissionen durch KI“ sowie „Ziele von Unternehmen und Staaten“ untersucht. Es ergab sich das eindeutige Bild, dass KI negative Auswirkungen auf die Nachhaltigkeit hat, was vor allem mit dem großen Stromverbrauch der Rechenzentren und der Herstellung der Hardware zusammenhängt. Gezeigt werden konnte aber auch, dass KI durchaus helfen kann, Emissionen zu reduzieren. Die Beziehung zwischen KI und Nachhaltigkeit ist also nicht einseitig. Im Rahmen einer praktischen Implementierung wurden drei KI-Modelle mit jeweils unterschiedlichen Parameteranzahlen der Gemma 3 Reihe von Google auf einem Server betrieben und dabei der Stromverbrauch gemessen, während alle Modelle dieselben Testfragen beantworteten. Daraus konnte geschlussfolgert werden, dass die Antworten umso präziser waren, je höher der Zeitaufwand und damit der Stromverbrauch bei der Beantwortung der Fragen war. So konnte sowohl mit der Vergleichsstudie als auch mit der praktischen Umsetzung gezeigt werden, dass zwischen KI und Nachhaltigkeit eine enge Beziehung besteht und KI sich meist negativ auf die Nachhaltigkeit auswirkt.

Inhaltsverzeichnis

Abkürzungsverzeichnis

CO_2	Kohlenstoffdioxid
CO_2e	Kohlenstoffdioxid-Äquivalente
EE	Embodied Emissions
EU	Europäische Union
FBB	Flughafen Berlin Brandenburg GmbH
IKT-Sektor	Informations- und Kommunikationstechnik-Sektor
KI	Künstliche Intelligenz
KNN	Künstliche Neuronale Netze
LCF	Load Capacity Factor
LLM	Large Language Model
ML	Machine Learning
NLP	Natural Language Processing
OE	Operational Emissions
PFM	Pretrained Foundation Model
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
TPU	Tensor Processing Unit

1 Einleitung

1.1 Motivation

Der Begriff Künstliche Intelligenz (KI) ist aus dem Alltag kaum noch wegzudenken. Fast in jedem Bereich des Lebens kommt KI mittlerweile zum Einsatz. Einmal stärker ausgeprägt, ein anderes Mal schwächer. Wurde vor einigen Jahren nach KI gefragt hat, dann gingen die ersten Gedanken vielleicht in Richtung Smart Assistants wie Amazon Alexa oder Apple Siri (Holder et al., 2018). Mittlerweile ist ChatGPT der Inbegriff für KI geworden (Ertel, 2025). Dabei ist zu beachten, dass diese Branche in den letzten drei Jahren einen enormen Boom verzeichnet hat (Ertel, 2025). So hat OpenAI ChatGPT erst 2022 veröffentlicht (Ertel, 2025), weitere Chatbots wie Copilot, Gemini oder DeepSeek kamen sogar noch später (DeepSeek, 2025; Google, 2024; Microsoft, 2023). Ein wesentlicher Grund für diesen Boom ist der niederschwellige Zugang zu diesen Technologien (OECD, 2025). So kann mit sogenannten Large Language Model (LLM)-basierten Chatbots wie mit einem Menschen über einen Chat kommuniziert werden, während es kaum bemerkbar ist, dass die Antworten maschinengeneriert sind (Harwardt et al., 2023; Zhou et al., 2023). Aber auch der tatsächliche Nutzen spielt dabei keine unerhebliche Rolle. Denn solche Chatbots können dabei helfen, die eigene Produktivität zu steigern (Peng et al., 2023), indem unter anderem die Recherche erleichtert wird und sich dadurch Wissen effizienter und gezielter angeeignet werden kann (OECD, 2025). Aber Unterstützung bei organisatorischen sowie kreativen Tätigkeiten oder alltäglichen Problemen kann ebenso erfahren werden. Auch in der Wirtschaft wird KI immer häufiger genutzt, unter anderem, um Produktivität und Effizienz zu steigern und Zeitersparnisse zu erzielen (Li et al., 2024).

Doch die Technik und der Aufwand, die hinter alldem stecken, sollte nicht unterschätzt werden. IBM (2023) erklärt, dass hinter jeder künstlichen Intelligenz ein sogenanntes KI-Modell steckt, welches Inputs entgegennimmt, diese auf der Basis seiner Algorithmen und Struktur verarbeitet und einen Output zurückgibt. Weiterhin wird erklärt, dass die Modelle eine immense Menge an Daten benötigen, um die Inputs korrekt verarbeiten zu können und dementsprechende Outputs zu liefern. Mit diesen Daten werden jene trainiert, einerseits mit menschlicher Unterstützung, andererseits selbstüberwacht. So lernen die Modelle, einen akkuraten Output zu liefern (IBM, 2023). Jedoch ist dieses Verfahren enorm rechenaufwändig und verbraucht viele Ressourcen von etwaigen Rechenzentren (Hacker, 2023; Luccioni et al., 2024b). Außerdem ist ein regelmäßiges Aktualisieren der Trainingsdaten erforderlich (Tomlinson et al., 2024). Das Training der Modelle ist jedoch nur ein Teil des Ressourcenverbrauchs von KI. Auf der anderen Seite kommen noch sämtliche Anfragen der Nutzer hinzu, welche ebenfalls verarbeitet werden müssen, um ein Ergebnis zurückzuliefern (Tomlinson et al., 2024). Da das verarbeitende Modell hochtrainiert ist, muss dieses entsprechend viele „Entscheidungen“ treffen, welche einen wesentlichen Teil des Verbrauchs darstellen (IBM, 2023). Mit Ressourcen ist dann die gesamte Infrastruktur von Rechenzentren gemeint, wozu nicht nur die Computer selbst gehören, sondern auch das Kühlungssystem bzw. die Klimaanlage (Guidi et al., 2024; OECD, 2022). Diese beiden Komponenten stellen den Hauptverbrauch dar (Guidi et al., 2024). So verbrauchten alle Rechenzentren weltweit im Jahr 2022 geschätzt zwischen 240 und 340 TWh Strom. Dies entspricht 1

bis 1,3 % des globalen Strombedarfs. Weiterhin führen Guidi et al. aus, dass geschätzt wird, dass sich dieser Stromverbrauch bis 2026 verdoppelt auf ca. 480 bis 680 TWh, was dem Jahresstrombedarf von Kanada entspricht. Diese Zahlen verdeutlichen, dass das Thema Nachhaltigkeit bzgl. künstlicher Intelligenz definitiv Aufmerksamkeit verdient. Um dem ein Stück weit gerecht zu werden, sollen einige wesentliche Aspekte davon in dieser Studienarbeit beleuchtet werden. Dazu werden im nächsten Kapitel zunächst die Projektziele formuliert.

1.2 Projektziele

Die Projektziele leiten sich aus einer konkreten Aufgabenstellung ab. So soll im Kern eine Vergleichsstudie verfasst werden, die mehrere Studien und Papers, welche sich mit dem Thema künstliche Intelligenz und Nachhaltigkeit befassen, systematisch miteinander vergleicht. Dabei soll der Fokus hauptsächlich auf aktuellen Ansätzen und Entwicklungen in dieser Thematik liegen. Zusätzlich dazu ist auch noch die Umsetzung eines geeigneten praktischen Teils erforderlich.

Um diese Aufgabenstellung erfüllen zu können, werden im Folgenden einige Ziele definiert. Zur Umsetzung der Vergleichsstudie soll die PRISMA-Methode gewählt werden, welche es ermöglicht, aus einer Vielzahl von Studien und Quellen die relevantesten herauszufiltern, um diese dann systematisch miteinander zu vergleichen (Page et al., 2021). Somit sollen einige Statistiken erstellt werden, um ermitteln zu können, wie viele Quellen sich mit einem bestimmten Aspekt von KI und Nachhaltigkeit befassen. Anhand dieser herausgearbeiteten Aspekte kann dann der letztendliche Vergleich stattfinden, um zu erfahren, was die verschiedenen Quellen zu diesen jeweils beinhalten. Damit ist es am Ende möglich, jeweils eine übergreifende Gesamterkenntnis in Form von einer Generalisierung zu erlangen, welche dann ein Gesamtüberblick über das Thema bieten. Während des Vergleichs sollen insbesondere drei Kernpunkte betrachtet werden. Einerseits der Stromverbrauch sowie die verursachten Emissionen durch künstliche Intelligenz, andererseits aber auch, wie KI zur Reduzierung von Emissionen beitragen kann. Für den praktischen Teil sollen zwei bis drei vortrainierte, verschiedene Large-Language-Modelle mit jeweils kleinen, mittleren und größeren Modellgrößen verglichen werden, indem der Stromverbrauch von diesen bei ca. 500 eindeutig zu beantwortenden Fragen gemessen wird. Damit lassen sich Aussagen über die Unterschiede bei Präzision und Stromverbrauch bzgl. der Modellgrößen ableiten.

Um das Gesamtthema und den Hintergrund besser verstehen zu können, werden in den folgenden Abschnitten einige theoretischen Grundlagen erläutert.

2 Theoretische Grundlagen

2.1 Einführung in die Künstliche Intelligenz

2.1.1 Definition Künstliche Intelligenz

Die wichtigste Grundlage ist das Verständnis, was künstliche Intelligenz bedeutet und wie sie in etwa funktioniert. Dazu soll zunächst versucht werden, KI zu definieren. Auch wenn dieses Thema bereits seit mehr als 50 Jahren untersucht wird, gibt es bis heute keine einheitliche Definition (Ertel, 2025). Das Grundproblem liegt bereits darin, dass selbst das Wort Intelligenz keine hat (Harwardt et al., 2023). Meist ist Intelligenz definiert als „die Fähigkeit eines Menschen […], seine kognitiven Fähigkeiten zur Lösung von Problemen einzusetzen“ (Harwardt et al., 2023, S.21). Da kognitiv so viel wie „das Wahrnehmen, Denken, Erkennen betreffend“ (Cornelsen Verlag GmbH, o.J.) bedeutet, lässt sich festhalten, dass mittels Intelligenz komplexe Informationen verstanden und verarbeitet werden, um anhand dessen entsprechende Schlussfolgerungen zu gewinnen (Bruhn, 2016). Außerdem kann laut Bruhn die Intelligenz durch Bildung, Erfahrungen und Training beeinflusst werden.

In der Literatur gibt es viele verschiedene Definitionen von KI, von denen im Folgenden einige erwähnt werden. Die erste stammt aus dem Jahr 1955 von John McCarthy, welcher sagte: „Ziel der KI ist es, Maschinen zu entwickeln, die sich verhalten, als verfügten sie über Intelligenz.“ (Ertel, 2025, S.1). Diese Definition hat jedoch die Schwäche, dass sich Maschinen mit einfachen Mitteln so konfigurieren lassen, dass sie intelligentes Verhalten lediglich vortäuschen (Ertel, 2025). Im Jahr 1991 wurde dann jene Definition von Rich und Knight aufgestellt: „Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.” (Harwardt et al., 2023, S.22) Diese Definition bessert die Schwäche aus und findet bis heute Anwendung (Ertel, 2025). „Künstliche Intelligenz beschreibt Informatik-Anwendungen, deren Ziel es ist, intelligentes Verhalten zu zeigen. Dazu sind in unterschiedlichen Anteilen bestimmte Kernfähigkeiten notwendig: Wahrnehmen, Verstehen, Handeln und Lernen.“ (Harwardt et al., 2023, S.22) definierten hingegen 2017 das Deutsche Forschungszentrum für Künstliche Intelligenz gemeinsam mit Bitkom. Aus der Wirtschaft, konkret von Amazon, stammt das Folgende: „Künstliche Intelligenz (AI) ist der Bereich der Informatik, der sich mit dem Erwerb kognitiver Fähigkeiten beschäftigt, die in der Regel menschlicher Intelligenz zugeordnet werden. Hierzu zählen Lernen, Problemlösung und Mustererkennung.“ (Harwardt et al., 2023, S.22f).

Bei Betrachtung dieser verschiedenen Definitionen fällt auf, dass diese zwar unterschiedliche Herangehensweisen haben, aber im Wesentlichen auf denselben Kern abzielen. Dieser ist die künstliche bzw. maschinelle Nachbildung von menschlicher Intelligenz. Bei Übertragung der vorherigen Überlegungen bzgl. Intelligenz auf die technische Seite, dann trifft ein KI-System anhand einer Eingabe eine Entscheidung bzw. Schlussfolgerung auf der Basis von komplexen Algorithmen, welche durch Unmengen an Daten antrainiert wurden. Durch dieses Training wurde dem System eine Art Intelligenz verliehen. Wie genau diese Entscheidung aber zustande kommt, ist nicht nachvollziehbar, ähnlich wie der Denkprozess im Gehirn. Daher stammt auch die Bezeichnung der KI als „Black Box“, da der Nutzer nur sieht, was Ein- und Ausgabe sind, aber nichts, was dazwischen liegt (Luber, 2024).

2.1.2 Teilgebiete der Künstlichen Intelligenz

Wie im vorherigen Abschnitt bereits deutlich wurde, ist das Thema künstliche Intelligenz sehr komplex. Aus diesem Grund wird zwischen Teilgebieten unterschieden, von denen die wichtigsten im Folgenden basierend auf den Ausführungen von Harwardt et al. (2023) kurz dargestellt werden sollen.

Das wohl wichtigste und bekannteste ist Machine Learning (ML) bzw. Maschinelles Lernen. Hierbei geht es darum, aus Daten zu lernen und das Gelernte auf Neues anzuwenden. Dazu wird ein KI-Modell mit Daten trainiert, in denen es Muster und Gesetzmäßigkeiten erkennt und somit die Trainingsdaten verallgemeinert. Dieses trainierte Modell kann dann auf neue Datensätze angewandt werden, bei denen es anhand der gelernten Verallgemeinerungen und Muster eine Entscheidung treffen kann. Es ist bei ML folglich keine komplette Programmierung des Algorithmus notwendig, sondern lediglich ein Training mit signifikanten Daten. Dazu gibt es verschiedene Lernparadigmen. Das erste Paradigma ist das „Supervised Learning“ bzw. „Überwachtes Lernen“, bei dem das Modell mit Daten trainiert wird, zu denen die richtige Ausgabe klar angegeben ist, d.h. die Daten gelabelt sind. Damit soll das Modell so trainiert werden, dass es zu noch unbekannten Eingaben die richtigen Ausgaben liefert sowie Eingaben korrekt klassifizieren kann. Im Gegensatz dazu gibt es das „Unsupervised Learning“ bzw. „Unüberwachte Lernen“, wobei lediglich mit Daten ohne richtige Ausgabe – also ungelabelten Daten – trainiert wird. Ziel ist es, dass das Modell selbstständig Muster findet und die Daten entsprechend clustert. Beim „Reinforcement Learning“ soll das Modell hingegen selbstständig die optimale Lösung für ein Problem finden, indem mit dem Prinzip Trial-and-Error gearbeitet wird. Dabei wird das Modell belohnt, wenn die gerade getroffene Entscheidung gut war und bestraft, wenn sie schlecht war. Somit lernt es sukzessive, immer bessere Entscheidungen zu treffen (Harwardt et al., 2023).

Als besondere Ausprägung von ML wird das sogenannte „Deep Learning“ von Harwardt et al. ausgemacht, bei dem zusätzlich zu den gerade aufgeführten Lernparadigmen noch künstliche neuronale Netze (KNN) zum Einsatz kommen. Ein KNN ist eine Imitation des menschlichen Gehirns mithilfe künstlicher Neuronen, wodurch die Modelle noch deutlich komplexere Zusammenhänge erkennen können. ML-Modelle kommen beispielsweise bei personalisierter Werbung oder in sozialen Medien zum Einsatz.

Als weiteres Teilgebiet führen Harwardt et al. das Natural Language Processing (NLP) aus, das sich mit der Verarbeitung natürlicher Sprache beschäftigt. Dazu gehört einerseits das Verständnis der natürlichen Sprache und das Erkennen der Bedeutung, wofür das „Natural Language Understanding“ verantwortlich ist und andererseits das Erzeugen natürlicher Sprache, wofür es das „Natural Language Generation“ gibt. Dazu muss ein Modell auch mit großen Datenmengen trainiert werden und dann Muster erkennen, um letztendlich die Bedeutung des Textes zu ermitteln. Das Ziel dieses Teilgebietes ist es, die Kommunikation zwischen Mensch und Computer so wie zwischenmenschliche Kommunikation zu ermöglichen. Damit wird auch die automatisierte Verarbeitung von in natürlicher Sprache vorliegenden Informationen ermöglicht. Beispiele für die Anwendung dieses Teilgebietes sind Chatbots, Übersetzungssysteme oder Sprachassistenten wie Siri und Alexa.

2.1.3 Large Language Models (LLM)

Eine wichtige Grundlage zum Verständnis der Relevanz des praktischen Teils dieses Studienprojekts ist das Verständnis, was ein LLM ist und wie es – vereinfacht betrachtet – funktioniert. Ein LLM ist ein KI-Modell, das Eingaben in natürlicher Sprache – sogenannte Prompts – erfassen und verarbeiten kann und anhand dieser Eingabe dann eine generierte Ausgabe ebenfalls in natürlicher Sprache zurückgibt (Naveed et al., 2025; Zhou et al., 2023). In den letzten Jahren gab es in diesem Fachgebiet erhebliche Fortschritte, welche es erlauben, dass ein LLM mittlerweile viele Aufgaben auf Menschenniveau ausführen kann (Naveed et al., 2025). Typische Anwendungsfälle sind laut Naveed et al. Übersetzungen, Zusammenfassungen und mittlerweile verstärkt Wissensabfragen, was bei vielen Einzug in den Alltag erlangt hat, vor allem in Form von Chatbots.

Ein Large Language Model gehört zur Gruppe der Pretrained Foundation Models (PFM), welche mit immensen Datenmengen vortrainierte KI-Modelle sind, um eine Basis für unterschiedlichste Anwendungsfälle bereitzustellen, für diese sie dann später feinjustiert werden (Zhou et al., 2023). So können PFMs einerseits auf NLP spezialisiert sein, wo schließlich LLMs einzuordnen sind, andererseits aber auch auf graphische Verarbeitung. Weiterhin führen Zhou et al. aus, dass die Hauptkomponente eines PFM und damit eines LLM der Transformer ist, der eine spezielle Deep-Learning-Architektur ist. Im Falle eines LLM ist dieser zuständig, die Texteingaben zu verarbeiten, indem er den einzelnen Wörtern Gewichtungen zuordnet sowie statistische Beziehungen zwischen den Satzbestandteilen erkennt, um somit den „Sinn“ und Kontext der Eingabe zu erfassen. Nach der Verarbeitung der Eingabe wird eine entsprechende Ausgabe generiert, indem – vereinfacht – jeweils das wahrscheinlichste nächste Wort an die Ausgabe angefügt wird. Damit jenes korrekt funktioniert, besitzt ein LLM mehrere Milliarden Parameter. Es lässt sich also festhalten, dass ein LLM im Kern auf den beiden Teilgebieten Machine Learning und Natural Language Processing basiert (Zhou et al., 2023).

Doch aufgrund dieser Komplexität verursacht das Training und die Nutzung von LLMs einen enormen Ressourcenverbrauch, was in entsprechenden umwelttechnischen und finanziellen Kosten mündet, worauf in einem späteren Kapitel nochmal präziser eingegangen wird (Naveed et al., 2025). Aufgrund dieses Umstands und der zunehmenden Bedeutung von LLMs im Alltag, bieten diese einen interessanten Ansatzpunkt für den praktischen Teil.

2.2 Einführung in die Nachhaltigkeit

2.2.1 Dimensionen der Nachhaltigkeit

Das Wort Nachhaltigkeit bedeutet, dass die Bedürfnisse der aktuellen Generation so befriedigt werden, dass die Bedürfnisse zukünftiger Generation nicht kleiner sein müssen. (1987: Brundtland Report, o. J.)

Nachhaltigkeit umfasst dabei ökologische, soziale und ökonomische Dimensionen. Ökologisch bedeutet, dass natürliche Ressourcen geschont und geschützt werden. Soziale Nachhaltigkeit sorgt für gerechte Lebensverhältnisse weltweit, etwa durch faire Löhne oder auch den Zugang zu Bildung. Ökonomisch nachhaltiges Handeln sieht eher langfristigen Wohlstand. Dabei geht es darum, dass die

Wirtschaft nicht auf Gewinnmaximierung fokussiert sein sollte, sondern eher darauf, Menschen und Umwelt zu schonen. Ein Beispiel für die Beziehung der drei Dimensionen sieht wie folgt aus: Der faire Bio-Kaffeeanbau schützt die Umwelt durch Verzicht auf Pestizide (ökologisch), sichert den Bauern faire Löhne (sozial) und ermöglicht ihnen langfristig wirtschaftliche Stabilität (ökonomisch). Dabei ist wichtig zu verstehen, dass alle drei Dimensionen auf einer Ebene sind und keine der anderen übergeordnet ist. (1987: Brundtland Report, o. J.)

Nachhaltigkeit ist entscheidend, um unseren Planeten lebenswert zu erhalten. Jeder Mensch kann durch kleine Veränderungen im Alltag einen Unterschied machen. Nur durch gemeinsames Handeln kann eine gerechte und lebenswerte Zukunft für alle Menschen gesichert werden.

2.2.2 CO_2-Fußabdruck

Der CO_2-Fußabdruck beschreibt die Menge an Kohlenstoffdioxid (CO_2), das durch den Menschen freigesetzt wird. Dieser dient als Maßstab zur Bewertung der Klimawirkung von Konsum- und Produktionsprozessen und wird meist in Kilogramm oder Tonnen Kohlenstoffdioxid-Äquivalente (CO_2e) pro Jahr angeben. Dabei ist es wichtig zu verstehen, dass Emissionen in unterschiedlichen Einheiten angegeben werden. So wird unterschieden zwischen reinen CO_2-Emissionen und CO_2e-Emissionen. CO_2e fasst die Emissionen verschiedener Treibhausgase zusammen, indem jene auf Grundlage ihres Global Warming Potentials in eine gleichwertige Menge an CO_2 umgerechnet werden (*CO_2 Äquivalente | ClimatePartner*, o. J.). Zu den Hauptverursachern zählen typischerweise der Energieverbrauch, der Verkehr sowie die Lebensmittelproduktion.

Auch im Bereich der digitalen Infrastruktur gewinnt der CO_2-Fußabdruck zunehmend an Bedeutung. Besonders rechenintensive Prozesse wie das Training großer KI-Modelle führen zu erheblichen CO_2-Emissionen, da sie enorm viel Energie verbrauchen. Diese Emissionen hängen wiederum stark davon ab, in welchen Ländern und unter welchen Strommixbedingungen die Rechenzentren betrieben werden.

Während auf individueller Ebene Maßnahmen wie der Umstieg auf öffentliche Verkehrsmittel, der Verzicht auf Inlandsflüge oder die Nutzung erneuerbarer Energien empfohlen werden, stehen im Technologiebereich vor allem energieeffiziente Hardware, nachhaltige Trainingsmethoden und der Einsatz grüner Rechenzentren im Fokus der Diskussion.

Ein bewusster Umgang mit dem CO_2-Fußabdruck ist somit nicht nur auf persönlicher Ebene relevant, sondern betrifft auch die verantwortungsvolle Entwicklung und Nutzung von KI-Technologien. Im Folgenden sollen verschiedene Studien verglichen werden, die den CO_2-Ausstoß durch KI-Systeme in Rechenzentren untersucht haben. (Hollstein, 2025)

Neben den Emissionen, die während der Nutzung eines Produktes entstehen, gibt es auch noch oft Emissionen, welche weniger sichtbar sind und trotzdem einen großen Einfluss haben. Dazu zählen zum einen die Embodied Emissions (EE), die bereits während der Herstellung und Errichtung anfallen. Bzgl. KI entstehen diese beispielsweise bei der Produktion von Hardware oder dem Bau der Rechenzentren.

Zum anderen gibt es auch noch die Operational Emissions (OE), die während des Betriebes freigesetzt werden. Bei KI wären das zum Beispiel die Emissionen durch das Kühlen der Rechenzentren oder die Versorgung der Infrastruktur. (Engel, 2021)

3 Methodik der systematischen Analyse

3.1 Methodik PRISMA

Die PRISMA-Methode (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) von Prof. Joanne McKenzie und Dr. Matthew Page ist eine international anerkannte Reporting-Guideline zur Verbesserung der Qualität und Transparenz von systematischen Übersichtsarbeiten und Meta-Analysen. Sie wurde erstmals 2009 veröffentlicht und 2020 umfassend aktualisiert. Die Methode wird verwendet, um den Prozess der Literaturrecherche zu standardisieren. Dies ist wichtig, da ein wichtiges Merkmal einer Vergleichsstudie die Reproduzierbarkeit ist. Forschende müssen die Möglichkeit haben, Ergebnisse zu einem späteren Zeitpunkt weiterhin nachvollziehen zu können. Um dies zu gewährleisten, existieren PRISMA-Richtlinien. Diese bieten ein Framework, welches die Forschenden unterstützt, alles korrekt umzusetzen. Im Wesentlichen gibt es vier Schritte, um die Analyse nach PRISMA durchzuführen. (Page et al., 2021)

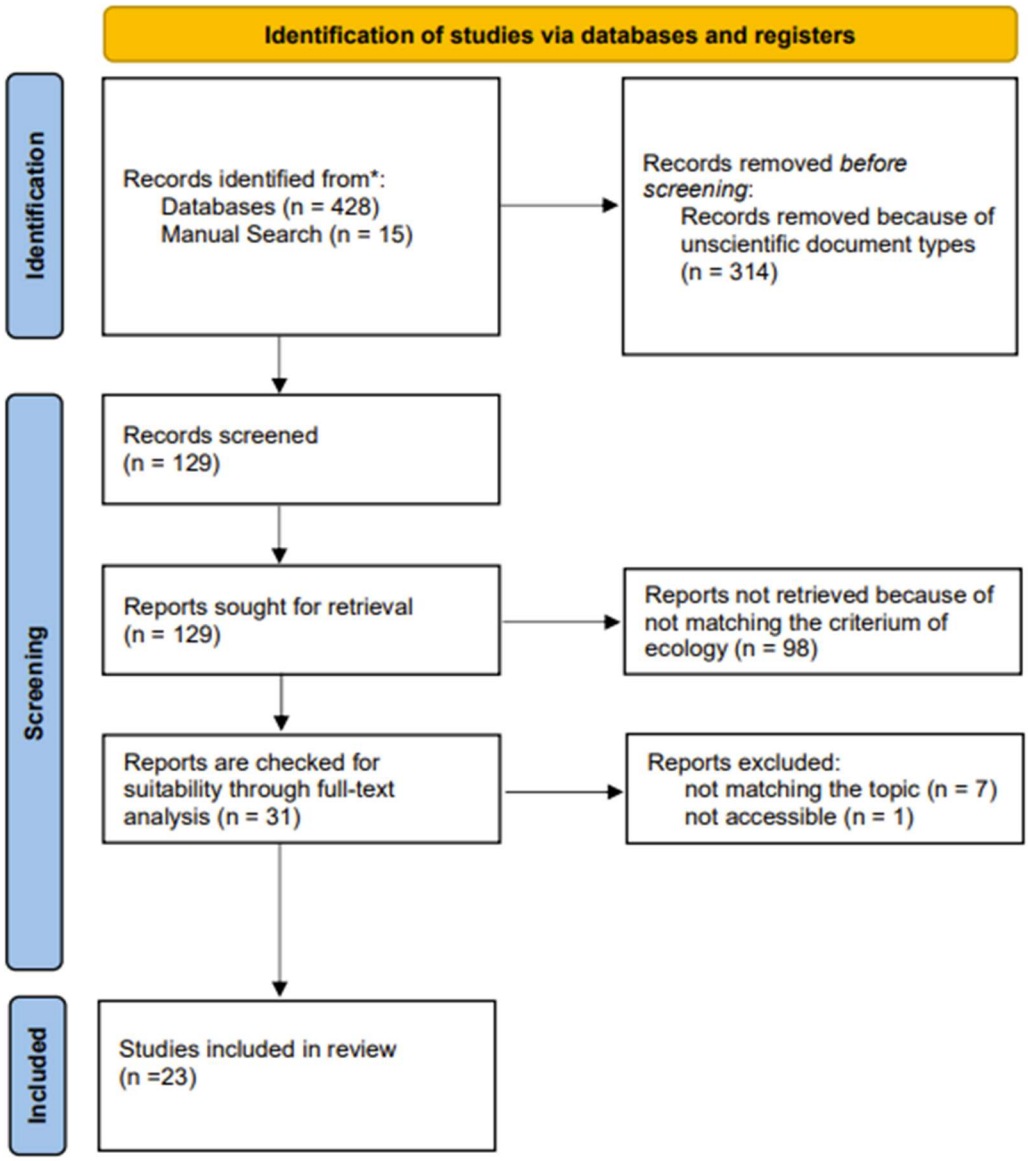
1. Identifikation
   In diesem ersten Schritt werden potenziell relevante Studien durch systematische Recherchen in Datenbanken sowie durch manuelle Suchen erfasst. Hierbei wird die Anzahl der gefundenen Studien aus den verschiedenen Quellen erfasst und notiert. Des Weiteren werden doppelte Datensätze gezählt, notiert und entfernt. Das Ziel dieser Phase ist es, alle potenziell relevanten Studien zur Fragestellung zu identifizieren.
2. Screening
   In dieser Phase werden die Titel und Abstracts der verbleibenden Studien gesichtet, um offensichtliche, nicht relevante Studien auszusortieren. Dafür werden vorab Einschluss- und Ausschlusskriterien definiert. Ziel ist es, eine erste grobe Filterung anhand oberflächlicher Informationen vorzunehmen
3. Eignung
   Die Volltexte der im Screening als relevant eingeschätzten Studien werden nun vollständig gelesen und detailliert beurteilt. Dies kann z.B. mit einer Entscheidungsmatrix geschehen. Es wird überprüft, ob sie die methodischen und inhaltlichen Kriterien für den Einschluss erfüllen. Studien, die methodische Mängel aufweisen oder nicht zur Forschungsfrage passen, werden ausgeschlossen.
4. Einschluss
   Dies ist die letzte Stufe, hier werden die final in den Review einbezogenen Studien angegeben. Nach Abschluss dieser Phase ist eine Studienbasis geschaffen, mit welcher die Studien zuverlässig analysiert werden können.

3.2 Durchführung PRISMA

Zur Durchführung der PRISMA-Methode wird strikt nach den oben beschriebenen Punkten vorgegangen. Das von PRISMA veröffentliche Flussdiagramm unterstützt hierbei, jederzeit den Überblick zu behalten. Dieses Flussdiagramm ist in Abbildung 1 dargestellt.

Abbildung 1:

PRISMA Flussdiagramm



Quelle: Eigene Darstellung

In der Identifikationsphase werden 428 Quellen aus der „Green Intelligence Ressource Hub“ Datenbank gewonnen. Diese sind in verschiedene Kategorien unterteilt, wie Podcasts, wissenschaftliche Artikel, Richtliniendokumente, Seminare, Konferenzen oder ähnliche Formate. Durch manuelle Recherche werden zusätzlich 15 weitere Quellen identifiziert. Insgesamt werden somit 443 Quellen identifiziert. Nach dem 1. Februar 2025 werden keine Quellen mehr berücksichtigt, zu diesem Zeitpunkt gilt die Identifikationsphase als abgeschlossen.

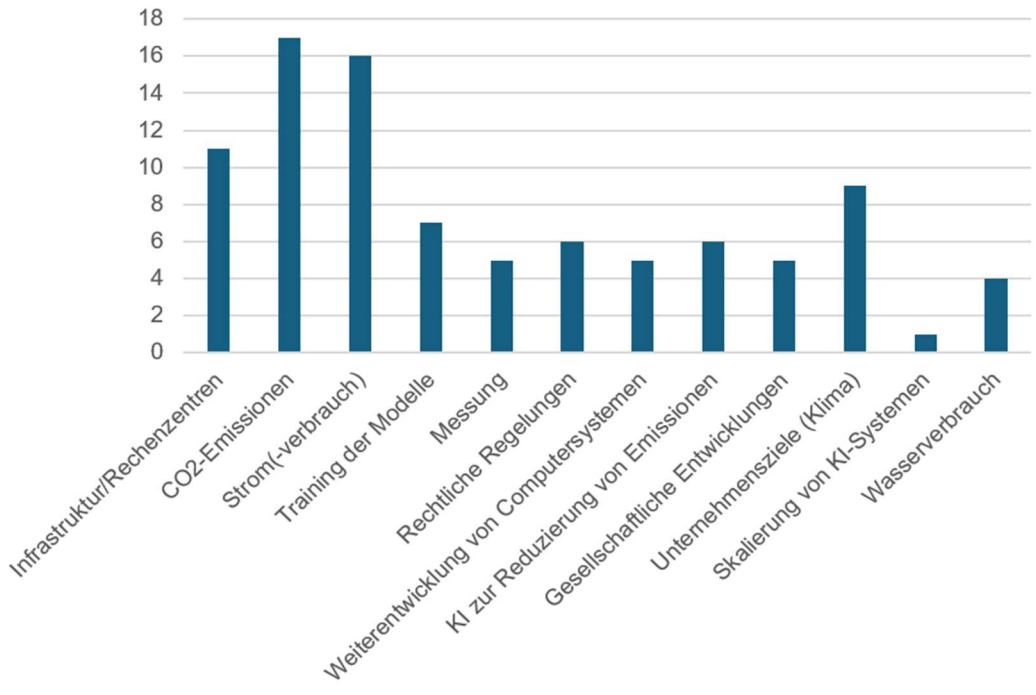
Vor der zweiten Phase, dem Screening, werden jedoch 314 Quellen entfernt. Dies hängt mit der Struktur der Datenbank „Green Intelligence Ressource Hub“ zusammen, welche eine Vielzahl an nicht-wissenschaftlichen Formaten enthält. Aus diesem Grund werden ausschließlich die 114 wissenschaftlichen Paper berücksichtigt.

Zu diesem Zeitpunkt befinden sich 129 Quellen in der Screening-Phase. Um in dieser Phase nicht relevante Studien herauszufiltern, werden spezifische Einschlusskriterien festgelegt. Die zentralen Kriterien umfassen „Carbon Footprint“, Environmental Impact“, „Sustainable AI“ und „Water Consumption“. Nach Sichtung der Abstracts werden 89 Studien ausgeschlossen, da sie nicht den definierten Kriterien entsprechen und somit nicht zur Nachhaltigkeitsstudie beitragen.

In der Phase der Eignung werden die 31 als relevant eingeschätzten Studien im Volltext analysiert und beurteilt. In dieser Phase wurden acht Studien ausgeschlossen, sieben erfüllten die inhaltlichen Kriterien nicht, da sie keine ökologischen Themen behandelten. Eine weitere Studie konnte aufgrund fehlenden Zugriffs nicht berücksichtigt werden. Die Beurteilung erfolgt unter Verwendung einer Entscheidungsmatrix, welche den Vorteil bietet, eine strukturierte und effiziente Übersicht über die inhaltlichen Schwerpunkte der einzelnen Studien zu ermöglichen. Die Kriterien der Entscheidungsmatrix werden dynamisch ergänzt, um einen gezielten Einblick in die jeweiligen Themenfelder zu erhalten.

Abbildung 2:

Thematische Verteilung der geeigneten Studien



Quelle: Eigene Darstellung

Die thematische Verteilung der inhaltlich geeigneten Studien ist in Abbildung 2 dargestellt. Auf der x-Achse sind die thematischen Schwerpunkte aufgeführt, die im Rahmen der Entscheidungsmatrix identifiziert wurden, während die y-Achse die Anzahl der jeweiligen Studien angibt. Besonders häufig werden Themen wie CO_2-Emissionen oder der Stromverbrauch behandelt. Aspekte wie Wasserverbrauch und die Skalierung von KI-Systemen werden hingegen deutlich seltener thematisiert. Die Verteilung verdeutlicht somit eine klare thematische Gewichtung innerhalb der untersuchten Literatur und zeigt, welche Nachhaltigkeitsaspekte derzeit im Fokus der Forschung stehen.

Studien, die trotz vorheriger Selektion in der Screening-Phase als nicht relevant identifiziert werden, werden in diesem Schritt ausgeschlossen. Mithilfe der Entscheidungsmatrix lassen sich insbesondere thematische Gemeinsamkeiten herausarbeiten, die als Grundlage für spätere Vergleiche dienen.

In der letzten Phase der PRISMA-Methode werden die final einbezogenen Studien benannt. Diese sind Anhang 1 zu entnehmen.

4 Ergebnisse der systematischen Analyse

4.1 Stromverbrauch von KI

Der Ressourcenverbrauch von KI-Modellen stellt nach wie vor eine zentrale Herausforderung im Kontext der Nachhaltigkeit dar. Zahlreiche Studien setzen sich mit dem Energiebedarf auseinander, insbesondere mit dem Stromverbrauch von Rechenzentren, die als infrastrukturelle Grundlage für KI-Anwendungen dienen. Dabei ist zu beachten, dass nicht nur die Recheneinheiten selbst, sondern auch die Kühlung der Server einen erheblichen Teil des Energieaufwands ausmachen.

Laut der Studie von Guidi et al. (2024) lag der weltweite Stromverbrauch von Rechenzentren im Jahr 2022 bei etwa 240–340 TWh, was rund 1–1,3 % des globalen Strombedarfs entspricht. Dieses Ergebnis deckt sich mit den Daten aus, die ebenfalls einen globalen Anteil von 1–1,3 % bestätigen. Zusätzlich wird in Luccioni et al (2024a) ein Anstieg des Stromverbrauchs in den letzten Jahren um 20–40 % aufgezeichnet. Besonders kritisch ist dabei, dass Rechenzentren nicht Strom verbrauchen, sondern auch etwa 1 % der weltweiten Treibhausgasemissionen verursachen.

Eine stärker regional fokussierte Betrachtung bietet eine Studie (Luccioni et al., 2024b), die sich auf die USA konzentriert. Hier liegt der Anteil von Rechenzentren am gesamten Stromverbrauch bereits bei 2–3 %, mit einer prognostizierten Verdreifachung in den kommenden Jahren. Besonders relevant im Kontext künstlicher Intelligenz ist der dort aufgeführte Anteil von 10–20 %, den KI-Anwendungen derzeit am Stromverbrauch von Rechenzentren ausmachen. Dieser Wert könnte laut Prognosen auf bis zu 70 % steigen, was auf eine deutliche Zunahme von Nutzung Künstlicher Intelligenz, welche rechenintensiver sind, schließen lässt.

Insgesamt zeigen die Studien deutliche Übereinstimmung hinsichtlich des aktuellen Energiebedarfs, unterscheiden sich jedoch in ihrem Fokus und ihrer Prognoseintensität. Während die Studien (Guidi et al., 2024), (Luccioni et al., 2024) eher die globale Lage analysieren, hebt die Studie von Luccioni et al. (2024b) insbesondere die Dynamik des US-amerikanischen Markts sowie den wachsenden Einfluss von KI hervor. Der Vergleich legt nahe, dass insbesondere der steigende Anteil von KI am Gesamtverbrauch ein entscheidender Treiber zukünftiger Energiebedarfe sein wird.

4.2 Emissionen durch KI

4.2.1 Verursachung der Emissionen

Wie bereits im voranstehenden Abschnitt erwähnt, verbrauchen KI-Systeme eine enorme Menge an Strom. Dieser Abschnitt soll die Studien nun etwas näher aufschlüsseln, wo genau die durch KI verursachten Emissionen entstehen. Mit diesem Thema beschäftigen sich 17, also knapp Dreiviertel der Studien. Zu beachten ist dabei der Unterschied zwischen Embodied Emissions und Operational Emissions.

Auf Seiten der EE stehen vor allem die Emissionen, die sowohl durch die Herstellung der Hardware für die Server (Lee et al., 2024) als auch durch das Training der Modelle (Luccioni et al., 2024b) verursacht werden. Hierbei weichen die Studien nicht voneinander ab (Hacker, 2023; Lee et al., 2024; Luccioni et al., 2024b; Taşdelen, 2024). Sowohl Produktion als auch Training sind mit einem großen Stromverbrauch verbunden (Lee et al., 2024; Luccioni et al., 2024b). Einige Studien führen noch auf, dass auch die Entwicklung der Modelle sowie der Bau von Rechenzentren Emissionen verursachen (Lee et al., 2024; OECD, 2022). Beim Gewinnen der Ressourcen zur Hardwareherstellung werden ebenso Emissionen verursacht, bei der Halbleiterherstellung stammen sogar 25 % der Emissionen von der Verwendung von diversen Gasen (Lee et al., 2024). Eine Studie erwähnt jedoch, dass laut NVIDIA ca. 80 bis 90 % des von KI verbrauchten Stroms auf den Betrieb zurückzuführen ist, was heißt, dass die EE deutlich geringer als die OE sind (Patterson et al., 2021). Sie beschränkt sich aber auf Deep Neural Network-Modelle und bezieht sich nicht auf KI allgemein (Patterson et al., 2021).

Die OE entstehen dementsprechend durch den Stromverbrauch der Server, die zum Hosten der Modelle und Verarbeitung der Anfragen in den Rechenzentren betrieben werden (Olawade et al., 2024). Entscheidend ist beim Stromverbrauch, dass – hier gibt es zwischen den Studien keine eindeutige Datengrundlage – nur 32 % (OECD, 2022) bzw. 40 % des Stroms (Luccioni et al., 2024b) aus erneuerbaren Energiequellen kommt. Wo sich die Studien wiederum überschneiden, ist, dass der Wahl des Standorts der Rechenzentren nicht unerheblich dazu beiträgt, wie groß die Emissionen sind (Guidi et al., 2024; Luccioni et al., 2022, Patterson et al., 2021; Taşdelen, 2024). Denn je nach Region stehen unterschiedlich viele erneuerbare Energien zur Verfügung, was somit auch Auswirkungen auf den Anteil der fossilen Energien und damit aus technischer Sicht auf die CO_2-Emissionen je kWh hat (Acun et al., 2023; Guidi et al., 2024). So können Regionen mit viel Sonne, aber wenig Wind, nur tagsüber sauberen Strom bereitstellen (Acun et al., 2023). In den USA sind beispielsweise 95 % aller Rechenzentren in Gebieten mit überdurchschnittlicher CO_2-Intensität (Guidi et al., 2024). Einigkeit besteht auch dabei, dass die Wahl der Hardware in den Rechenzentren selbst sowie die der Architektur der KI-Modelle und der Trainingsstrategien einen Unterschied machen kann (Lee et al., 2024; Olawade et al., 2024; Patterson et al., 2021; Wu et al., 2022). Aber auch der Strom zum Betrieb der Kühlung sollte bei den Operational Emissions nicht vergessen werden (Hacker, 2023).

Es lässt sich also festhalten, dass vor allem bei den großen Verursachern wie Training, Hardwareproduktion und Betrieb der Rechenzentren eine große Übereinstimmung zwischen den Studien zu erkennen ist.

4.2.2 Auswirkungen der Emissionen

Diese großen Emissions-Verursacher haben folglich Auswirkungen ökologischer, aber auch finanzieller Art (Fraisl et al., 2025). So ist der Information- und Kommunikationstechnik-Sektor (IKT-Sektor) mittlerweile für 2 % der globalen CO_2-Emissionen verantwortlich (Luccioni et al., 2022). Eine andere Studie geht sogar von 3,9 % der globalen Treibhausgasemissionen aus (Hacker, 2023). Während die 150 größten Tech-Unternehmen 1,6 % des globalen Stromverbrauchs verursachen (OECD, 2022), sind bei Google die Treibhausgasemissionen seit 2019 um 48 % und bei Microsoft seit 2020 um 29,1 %

gestiegen (Luccioni et al., 2024b). Dies fällt zeitlich zusammen mit dem Boom von KI (Ertel, 2025). Der Load Capacity Factor (LCF) ist eine Variable, die den Zustand der Ökosysteme darstellt und sich darauf bezieht, wie sehr ein Land die Kapazität an Ressourcen besitzt, seine Bevölkerung mit ihrem Lebensstil zu versorgen bzw. ob der Ressourcenverbrauch langfristig getragen werden kann (Shewly et al., 2024). Der LCF stützt sich dabei auf die Parameter Bruttoinlandsprodukt, finanzielle Zugänglichkeit, Globalisierung, Urbanisierung, aber mittlerweile auch künstliche Intelligenz (Shewly et al., 2024).

All dies zeigt, dass KI eine zunehmend größere Rolle hinsichtlich Emissionen und Nachhaltigkeit spielt, auch wenn die Studien unterschiedliche Zahlen vorlegen.

4.2.3 Konkrete Emissionswerte

In diesem Abschnitt sollen ein paar konkrete Emissionswerte vorgestellt werden, die jene Auswirkungen verdeutlichen. Dabei ist zu beachten, dass diese in unterschiedlichen Einheiten angegeben werden, wie bereits in Kapitel zwei erklärt wurde.

Eine signifikante Entwicklung ist bei den von US-Rechenzentren verursachten Emissionen zu beobachten. Haben jene im Jahr 2018 noch 31,5 Mio. Tonnen CO_2e verursacht, waren es im Zeitraum von September 2023 bis August 2024 schon 105 Mio. Tonnen CO_2e (Guidi et al., 2024). Weiterhin berichten drei Studien übereinstimmend, dass das Training von BLOOM – einem LLM mit 176 Milliarden Parametern – insgesamt ca. 50 Tonnen CO_2e verursacht hat, d.h. inklusive Hardwareherstellung, Stromverbrauch ebendieser sowie Kühlung (Bhardwaj et al., 2025; Luccioni et al., 2022; Tomlinson et al., 2024). Das entspricht dem CO_2-Ausstoß einer Person, wenn sie 74-mal von Berlin nach Palma de Mallorca und zurück mit einem typischen Flugzeug auf dieser Strecke in der Economy-Class fliegt (Stiftung myclimate, 2025). Das Training von GPT-3 hat laut zwei Studien dagegen mehr als das Zehnfache, genauer 552 Tonnen CO_2e, verursacht, was in etwa 800 derartiger Flugreisen entspricht (Patterson et al., 2021; Tomlinson et al., 2024). Während sich diese Studien auf das Training beschränkt, gehen Tomlinson et al. (2024) noch auf die Emissionen im Gesamtbetrieb ein. Rechnet man sowohl EE als auch OE zusammen, dann verursacht eine durchschnittliche Anfrage an BLOOM 1,6 g CO_2e und an GPT-3 2,2 g CO_2e (Tomlinson et al., 2024). Weitere interessante Vergleiche sind unter anderem, dass die Bildgenerierung so viel Strom wie eine Smartphoneladung verbraucht (Hacker, 2023) und dass durch Rechenzentren in den nächsten Jahren so viele Treibhausgase wie 16 Millionen Autos ausstoßen könnten (Luccioni et al., 2024).

Auch wenn die Studien unterschiedlichste Werte bereitstellen, so zeigt sich dennoch eindeutig, dass KI mittlerweile für große Mengen an Emissionen verantwortlich ist.

4.2.4 Herausforderungen

Im selben Zug erwähnen einige Studien aber auch, dass es einige Herausforderungen gibt bzgl. Emissionen von KI. So ist bereits die korrekte Messung der verursachten Emissionen ein nicht zu unterschätzendes Problem (Lee et al., 2024; Patterson et al., 2021). Denn während die verbrauchte

Energie einfach gemessen werden kann, sind die Emissionen durch den Abbau der Ressourcen und die Hardwareherstellung selbst kaum erfassbar (Tomlinson et al., 2024). Neben diesem Problem sehen weitere Studien eine Herausforderung darin, effizientere Hardware und Algorithmen zu entwickeln, ohne dabei in den Rebound-Effekt zu laufen, d.h. die Einsparungen durch intensivere Verwendung zunichtezumachen (Lee et al., 2024; Olawade et al., 2024). Acun et al. (2023) sehen aber auch generell ein Problem darin, Rechenzentren durchgängig emissionsfrei zu betreiben, da erneuerbare Energien nicht konstant verfügbar sind.

Dies lässt die Schlussfolgerung zu, dass die Studien zwar übereinstimmend Probleme hinsichtlich der Erfassung und Reduzierung von Emissionen sehen, diese aber aus unterschiedlichen Perspektiven wie methodisch bis hin zu technisch betrachten.

4.3 Ziele von Unternehmen und Staaten

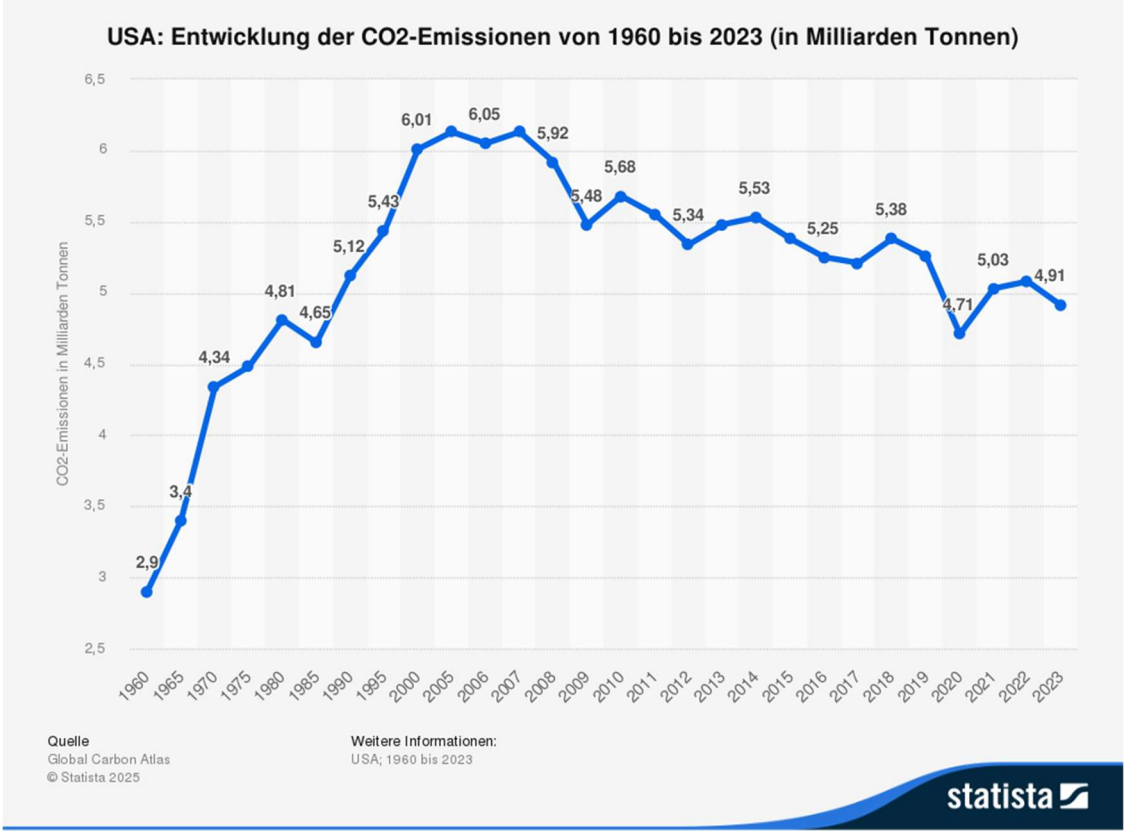
Zahlreiche Unternehmen, darunter auch Meta, Microsoft und Google, verfehlen infolge der rasant fortschreitenden Entwicklung von KI ihre selbst gesetzten Nachhaltigkeitsziele (Guidi et al., 2024). Grundsätzlich streben diese Unternehmen Klimaneutralität an, jedoch wird dieses Ziel aufgrund des steigenden Energiebedarfs von KI-Anwendungen verschoben oder vollständig aufgegeben (Bhardwaj, 2025). Hinzu kommt, dass 64 % der Unternehmen den tatsächlichen Energieverbrauch durch KI als zu komplex für eine präzise Messung einschätzen (Desroches et al., 2025).

Google verfolgt trotz des KI-Booms weiterhin das Ziel, innerhalb der nächsten fünf bis fünfzehn Jahre die Klimaneutralität zu erreichen (Luccioni et al., 2025). Allerdings stiegen die CO_2-Emissionen des Unternehmens im Zeitraum von 2019 bis 2023 um 50 % an (Bhardwaj, 2025). Allein im Jahr 2023 betrug der Anstieg im Vergleich zum Vorjahr 13 % (Guidi et al., 2024). Parallel dazu erhöhte sich der Stromverbrauch. Zwischen 2015 und 2021 wurde ein jährlicher Zuwachs von 25 % verzeichnet, wobei diese Entwicklung nicht nur Google, sondern ebenso Meta und Microsoft betrifft (Lee et al., 2024). Die Klimaziele für 2024 kann Google, aufgrund der Entwicklung von KI nicht einhalten (Desroches et al., 2025). Zwar sind Investitionen in erneuerbare Energien geplant, doch wird kurzfristig weiterhin auf fossile Energiequellen zurückgegriffen (Bhardwaj, 2025).

Auf globaler Ebene sehen sich Staaten wie die Vereinigten Staaten und China sowie supranationale Akteure wie die Europäische Union (EU) mit vergleichbaren Herausforderungen konfrontiert. Die Vereinigten Staaten haben sich zum Ziel gesetzt, ihre Treibhausgasemissionen bis zum Jahr 2030 um 50 bis 52 % gegenüber dem Niveau von 2005 zu reduzieren. Dies hängt damit zusammen, dass die Vereinigten Staaten im Jahr 2005 einen historischen Hochstand an Treibhausgasemissionen verzeichnet haben, wie in Abbildung 3 zu erkennen ist. Die Volksrepublik China verfolgt ein vergleichbares Ziel, allerdings sind keine konkreten Werte bekannt. Im Februar 2020 haben 50 Länder angekündigt, Strategien zu entwickeln, um den KI-Fortschritt nachhaltig zu gestalten. Die EU strebt an, die Emissionen bis 2030 auf das Niveau des Jahres 1990 zurückzuführen (Ding et al., 2024).

Abbildung 3:

Entwicklung der CO_2-Emissionen der Vereinigten Staaten vom Amerika



Quelle: Global Carbon Atlas, 2024

4.4 Reduzierung von Emissionen

4.4.1 Weiterentwicklung von Computersystemen

Um durch KI verursachte Emissionen zu reduzieren, muss der Ressourcenverbrauch gesenkt werden, wofür es im Wesentlichen zwei Ansätze gibt: Einerseits kann natürlich die Nutzung von KI verringert werden, andererseits wäre ein effizienterer und gezielterer Umgang mit Ressourcen denkbar. Da der erste Ansatz angesichts des Potenzials von KI nicht praktikabel ist, bleibt der zweite. Fünf der im Rahmen dieser Vergleichsarbeit analysierten Studien beschäftigen sich unter anderem mit der Weiterentwicklung von Hard- und Softwaresystemen in Verbindung mit künstlicher Intelligenz. All diese haben gemeinsam, dass sie ebenso den zweiten Ansatz wählen.

Drei von jenen setzen direkt auf „Sustainability By Design“ (Hacker, 2023; Lee et al., 2024; Patterson et al., 2021). Darunter versteht man, dass Produkte und Dienstleistungen bereits so entwickelt werden, dass sie Nachhaltigkeitsstandards genügen, was sowohl auf technischem als auch organisatorischem Wege erfolgen kann (Hacker, 2023). Dazu zählen z.B. Langlebigkeit, einfache Reparierbarkeit, Wieder- bzw. Weiterverwendbarkeit oder auch effiziente Energieverwendung (Rüdiger, 2023). Während eine dieser Studien lediglich dieses Konzept selbst fordert, liefern die beiden anderen konkrete Umsetzungsvorschläge. Die Studie von Lee et al. (2024) fokussiert sich vor allem auf die Embodied

Emissions und konkret auf die drei „R“ der Kreislaufwirtschaft. Das erste betrifft die Reduzierung, d.h. die gleiche Rechenleistung soll mit weniger Hardwareressourcen erreicht werden. Dies kann durch spezialisierte Chips und die dynamischere Allokation von Ressourcen in Rechenzentren geschehen. Das zweite „R“ bezieht sich bei Lee et al. auf die Wiederverwendung (Reusing), d.h. dass durch komponentenbasierte Systeme einzelne Bestandteile statt der gesamten Einheit gewechselt werden können bei Defekten oder Upgrades. Als Letztes gibt es das Recycling, das vor allem zum Ziel hat, Hardware ein zweites Leben zu ermöglichen (Lee et al., 2024). Die zweite Studie von Patterson et al. (2021) setzt zum Teil ebenfalls auf für die spezifische Anwendung spezialisierte Hardware. So wird beispielsweise der Einsatz von sogenannten Tensor Processing Units (TPU) vorgeschlagen. Eine TPU ist ein von Google entwickelter Chip, der vor allem auf maschinelles Lernen optimiert ist (Litzel & Luber, 2019). Außerdem werden Algorithmen und Technologien genannt, die die Energieeffizienz verbessern sollen, z.B. die Nutzung vortrainierter Modelle oder das Transferieren des Wissens größerer Modelle in kleinere (Patterson et al., 2021).

Die zwei anderen Studien setzen weniger direkt auf Sustainability By Design, verfolgen aber einen ähnlichen Grundgedanken (Acun et al., 2023; Taşdelen, 2024). Die Studie von Acun et al. (2023) hat ein Framework entwickelt, um die Zeiten, in denen Rechenzentren emissionsfrei betrieben werden, zu maximieren. Dieses Framework nennt sich „Carbon Explorer“ und betrachtet sowohl EE als auch OE. Dazu wurden drei wesentliche Strategien von Acun et al. ausgemacht, die am besten zum Ziel beitragen. Die offensichtlichste ist die Verwendung erneuerbarer Energien, wobei diese allein sehr wetterabhängig sind. Zum Ausgleich dessen werden Batteriespeicher eingesetzt. Zuletzt sollen auch Scheduler verwendet werden, die einige geplante Aufgaben verschieben in Zeiten, wo mehr erneuerbare Energien verfügbar sind. Die Studie geht davon aus, dass ca. 40 % der Tasks eines Rechenzentrums verschoben werden können (Acun et al., 2023). Eine ähnliche Strategie beinhaltet auch die Studie von Lee et al. (2024). So sollen Operational Emission reduziert werden, indem ein sogenannter Demand Response steuert, wann und wo die Aufgaben verarbeitet werden in Abhängigkeit von der Sauberkeit des Stroms und Tageszeit (Lee et al., 2024). Carbon Explorer als Framework hat dann die konkrete Aufgabe, eine ideale Kombination aus allen drei Strategien zu erreichen, um die emissionsfreien Zeiten zu maximieren (Acun et al, 2023). So kann laut Acun et al. abhängig von der Region der CO_2-Fußabdruck von Rechenzentren durch diese Strategien um 15 bis 65 % reduziert werden. Einen komplett anderen Ansatz hat dabei die zweite Studie von Taşdelen (2024) entwickelt, um Embodied Emissions zu reduzieren. Dazu hat sie einen sogenannten „Early Stopping“ Mechanismus entwickelt, der das Modell-Training vorzeitig abbricht, um das Training energieeffizienter zu gestalten. Dafür werden parallel zum Training Leistungsmetriken gemessen, wobei das Training dann abgebrochen wird, wenn keine Verbesserungen mehr festgestellt werden oder die Genauigkeit sich sogar verschlechtert durch „Over-Fitting“. Bei einer Fallstude von Taşdelen mit einem „Schere, Stein, Papier“-Datensatz hat sich ergeben, dass die Trainingszeit erheblich reduziert werden konnte, während bei der Genauigkeit keine relevanten Unterschiede festgestellt werden konnten. D.h. es lässt sich noch nicht verallgemeinern, ob „Early Stopping“ ein zukunftsweisendes Konzept ist, aber es bietet definitiv Potenzial (Taşdelen, 2024).

Zusammenfassend lässt sich sagen, dass bei der Weiterentwicklung von Computersystemen die Nachhaltigkeit eine große Rolle spielt und es die verschiedensten Ansätze gibt, um Emissionen von KI

zu reduzieren, indem Rechenressourcen effizienter werden und deren Nutzung gezielter geplant wird. Drei dieser Ansätze basieren auf dem Ansatz „Sustainability By Design“, während die anderen zwei mehr den Fokus auf die effiziente und bedachtere Verwendung der Ressourcen legen, wobei es zwischen diesen Ansätzen durchaus auch Überschneidungen gibt. Jedoch sollte darauf geachtet werden, dass nicht der Rebound-Effekt eintritt und durch diese Einsparungen der Anreiz geschaffen wird, mehr KI einzusetzen und damit die Einsparungen zu egalisieren (Lee et al., 2024).

4.4.2 Künstliche Intelligenz gegen Emissionen

Ein Viertel der untersuchten Studien beschäftigen sich aber auch mit dem Thema, wie künstliche Intelligenz so eingesetzt werden kann, dass Emissionen reduziert oder gar vermieden werden. So gibt es zahlreiche Anwendungsfälle, in denen KI einen größeren Effekt erzielen kann, als dass sie mehr Emissionen verursacht. Laut einer Studie kann der LCF sogar um 0,029 % auf kurze Sicht bzw. 0,142 % auf lange Sicht gesteigert werden, wenn der Parameter künstliche Intelligenz um 1 % steigt (Shewly et al., 2024). D.h. durch KI verbessert sich der Zustand der Ökosysteme tatsächlich, wenn sie richtig eingesetzt wird. Von diesen Fällen sollen im Folgenden einige vorgestellt werden.

Der am häufigsten genannte Ansatz ist die Integration von KI in den Strommarkt und die Stromproduktion (Hacker, 2023; Olawade et al., 2024; Rolnick et al., 2022). So kann KI anhand von Wettervorhersagen die Stromproduktion vor allem mit erneuerbaren Energien vollautomatisch anpassen (Olawade et al., 2024; Rolnick et al., 2022). Bei für Erneuerbare gutem Wetter kann diese deren Produktion erhöhen und die mit fossilen Energien reduzieren (Olawade et al., 2024; Rolnick et al., 2022). Bei vorauszusehendem schlechteren Wetter kann diese entscheiden, einen Teil des produzierten, sauberen Stroms in Batterien zu speichern (Rolnick et al., 2022). Außerdem kann die KI auf der Grundlage von historischen Daten den Strombedarf korrekter vorhersagen und eine Überproduktion vor allem mit fossilen Energien vermeiden (Hacker, 2023; Olawade et al., 2024).

Ein weiterer bedeutender Ansatz ist das Bauen von oder die Weiterentwicklung zu Smart Buildings in Kombination mit Internet Of Things zur Steigerung der Energieeffizienz (Ding et al., 2024; Olawade et al., 2024; Shewly et al., 2024). So können Sensoren und Managementsysteme dazu beitragen, Heizungen, Lüftungen, Klimaanlagen oder Beleuchtungen gezielter und effizienter zu steuern (Ding et al., 2024, Olawade et al., 2024). Wenn diese Systeme mit KI kombiniert werden, kann diese beispielsweise Nutzungsmuster oder -zeiten erkennen und damit die Anlagen auf ein Basisniveau herunterfahren, wenn mit keiner Nutzung zu rechnen ist und so letztendlich Energie sparen (Ding et al., 2024, Olawade et al., 2024). Dies kann sowohl in Privatgebäuden als auch in öffentlichen Gebäuden angewandt werden (Olawade et al., 2024).

Geht es um das Thema Stadtentwicklung, dann spielt auch der Ansatz der Entwicklung intelligenter Verkehrsnetze und Mobilität eine Rolle, welcher von Olawade et al. (2024) etwas näher beleuchtet wird. Dabei könnten Kameras und Sensoren die aktuelle Verkehrslage erfassen und KI könnte dann gemeinsam mit historischen Daten Maßnahmen treffen, den Verkehrsfluss zu optimieren, indem z.B.

Ampelschaltungen oder Geschwindigkeitsbegrenzungen angepasst werden. So entstehen weniger Staus und Anfahrmanöver, was ebenso Emissionen und Ressourcen einspart.

Das Problem an diesen Lösungen ist, dass diese noch teuer und komplex in der Anschaffung sind und somit auch nicht so weit verbreitet sind (Ding et al., 2024). Jedoch sagen Ding et al. auch, dass KI unterstützen kann, indem sie Planungen automatisiert, die Konfiguration durch selbstständiges Lernen erleichtert und bei Entscheidungen unterstützt. Dadurch reduzieren sich die Kosten sowie die Komplexität und es kommt zu einem weiteren Einsatz bzw. Skalierung dieser Systeme, was wiederum die Emissionen verringert.

KI kann aber auch genutzt werden, um Emissionen oder andere Umweltprobleme besser zu tracken und monitoren (Fraisl et al., 2025; Olawade et al., 2024). Dadurch könnten beispielsweise systematisch Schwach- bzw. Problemstellen gefunden und auf dieser Datengrundlage behoben werden durch fundierte Entscheidungen. Die Studie von Fraisl et al. (2025) schlägt sogar vor, diese Möglichkeit mit Citizen Science zu kombinieren, um gewisse Synergien zu entfachen. Denn Citizen Science Projekte können die Verfügbarkeit und Qualität von lokalen Daten deutlich verbessern, was sich positiv auswirkt.

Eine völlig andere Richtung geht eine weitere Studie, die untersucht hat, ob KI bei manchen Aufgaben weniger CO_2 verursacht als Menschen (Tomlinson et al., 2024). Dazu hat sie die vier KI-Modelle ChatGPT, BLOOM, DALL-E2 sowie Midjourney genommen und diese sowohl Texte als auch Bilder generieren lassen. Bei den CO_2e-Emissionen verursacht durch die Modelle wurden einerseits diejenigen durch Training und andererseits diejenigen durch die Abfragen berücksichtigt. Bei den menschlichen Emissionen wurden Durchschnittswerte verwendet, wie viel CO_2e pro Stunde und Bürger verursacht werden. Die Studie von Tomlinson et al. kommt zu dem Ergebnis, dass die Modelle zwischen 130- und 1500-mal weniger CO_2e zur Textgenerierung und zwischen 310- und 2900-mal weniger zur Bildgenerierung verursachen. Die großen Intervalle kommen dadurch zustande, dass die Modelle unterschiedlich viel Emissionen verursachen und dass z.B. ein US-Bürger im Durchschnitt einen größeren CO_2-Fußabdruck als ein indischer Bürger hat. KI reduziert folglich auch Emissionen, wenn es bestimmte Aufgaben von Menschen übernimmt. Jedoch lässt sich diese Erkenntnis nicht generalisieren, da der Anwendungsfall sehr eingeschränkt ist (Tomlinson et al., 2024).

Es lässt sich also festhalten, dass KI selbst durchaus das Potenzial hat, Emissionen zu reduzieren, wenn sie an den richtigen Stellen zum Einsatz kommt. Vor allem beim Thema Integration von KI in die Stromproduktion mit erneuerbaren Energien sind sich diese Studien sehr ähnlich (Hacker, 2023; Olawade et al., 2024; Rolnick et al., 2022). Aber auch die Erhöhung von Energieeffizienz durch intelligente Gebäudesteuerungsfunktionen findet häufig Erwähnung (Ding et al., 2024; Olawade et al., 2024; Shewly et al., 2024). Weitere Lösungen wie intelligente Verkehrssysteme, exakteres Monitoring von Emissionen oder die Übernahme von Aufgaben von Menschen spielen eine nicht zu unterschätzende Rolle (Fraisl et al., 2025; Olawade et al., 2024; Tomlinson, 2024). Während jedoch Shewly et al. (2024) von langfristig positiven Auswirkungen der KI auf die Umwelt ausgehen, warnen Tomlinson et al. (2022) erneut vor dem Rebound-Effekt.

4.5 Zwischenfazit

Nachdem die unterschiedlichen Vergleichsaspekte umfassend beschrieben wurden, soll nun ein Gesamtüberblick über das Verhältnis von KI und Nachhaltigkeit gegeben werden, welcher die verschiedenen Studien zusammenfasst. Insgesamt wurde dieses Verhältnis aus vier Perspektiven betrachtet.

Zum Stromverbrauch lässt sich sagen, dass dieser vor allem durch den Verbrauch der Rechenzentren, in denen die KI betrieben wird, zustande kommt. Den größten Teil macht dabei der Stromverbrauch der Server-Hardware aus. Nicht vergessen werden sollte jedoch, dass auch der Betrieb der Kühlungs- bzw. Klimaanlagen einen erheblichen Teil dazu beiträgt. Derzeit haben Rechenzentren einen Anteil von 1,0 bis 1,3 % am globalen Stromverbrauch, wobei KI selbst – zumindest in US-Rechenzentren – 10 bis 20 % des Stroms der Rechenzentren benötigt. Bei beiden Anteilen ist in den nächsten Jahren mit einem deutlichen Anstieg zu rechnen.

Die zweite Perspektive umfasst die Emissionen, die durch KI verursacht werden. Dabei unterscheiden die Studien zwischen Embodied Emissions einerseits und Operational Emissions andererseits. Zu den Embodied Emissions gehören im Wesentlichen die Emissionen, die durch Förderung der Ressourcen für die Hardware sowie die Herstellung ebendieser verursacht werden, aber auch der Stromverbrauch beim Training der KI. Die Operational Emissions setzen sich dagegen aus den Emissionen vom Strom zum Betrieb der KI-Server zusammen. Bei der Nutzung von Strom entstehen Emissionen, da nach wie vor der Großteil des Stroms aus fossilen Quellen kommt. So ist je nach Quelle der IKT-Sektor für mittlerweile 2,0 bis 3,9 % der globalen Emissionen verantwortlich. Diese Abweichung kommt unter anderem dadurch zustande, dass die konkrete Erfassung und Messung der Emissionen eine große Herausforderung darstellen.

Bzgl. der Ziele von Unternehmen und Staaten kann festgehalten werden, dass sich die größten Unternehmen wie Google, Microsoft und Meta zwar vorgenommen haben, ihre Emissionen zu reduzieren, indem sie sich zum Teil sogar konkrete Ziele gesetzt hatten, durch den Boom von KI in den letzten Jahren diese aber deutlich verfehlt und auch weiterhin verfehlen werden. Dennoch streben sie weiterhin die Erreichung jener an. Auch verschiedene Industrienationen wie die USA, China oder die Mitgliedsstaaten der EU haben Ziele und Strategien entwickelt, um ihre allgemeinen Emissionen sowie die Emissionen durch KI zu reduzieren. Das Bewusstsein für die hohen Emissionen durch KI ist folglich vorhanden, jedoch mangelt es noch an der Umsetzung.

Der letzte Aspekt bezieht sich darauf, inwiefern die Emissionen reduziert werden können und wie KI sogar dabei helfen kann. Zur Reduzierung gibt es vor allem Konzepte im Rahmen von „Sustainability By Design“, also dass Produkte bereits unter Gesichtspunkten der Nachhaltigkeit entwickelt werden. So sollen KI-Modelle mit effizienteren Strategien trainiert werden und auf KI spezialisierte Hardware eingesetzt werden. Aber auch Scheduler in Abhängigkeit von der Sauberkeit des Stroms sollen zum Einsatz kommen. KI kann Emissionen aber auch aktiv reduzieren, indem sie zum Beispiel bei der Produktion von erneuerbarem Strom oder in Gebäudesteuerungssystemen eingesetzt werden.

5 Systematische Verbrauchsmessung vortrainierter LLMs

5.1 Einführung

KI ist im Alltag mittlerweile ein fast nicht mehr wegzudenkender Begleiter. Wie im bisherigen Teil dieser Studienarbeit erklärt, verbrauchen KI-Modelle viele Ressourcen. Zu Beginn dieser Studienarbeit standen drei Konzepte zur Auswahl. Diese wurden mit Experten im Bereich KI der Flughafen Berlin Brandenburg GmbH (FBB) evaluiert. Das Ergebnis war hierbei ein Vergleich zwischen einem Model mit verschiedenen Parametern. Im Folgenden werden die Schritte dargestellt, welche zur Durchführung erforderlich sind, sowie eine Auswertung des Vergleichs.

5.2 Vorbereitung

5.2.1 Herausforderungen

Im Rahmen der praktischen Umsetzung ergeben sich zwangsläufig verschiedene Herausforderungen. Diese gliedern sich in zwei Bereiche auf, zum einen in Herausforderungen mit der Infrastruktur und Regelungen der FBB und zum anderen in allgemeine Herausforderungen unabhängig vom Praxispartner.

Die größte Herausforderung innerhalb der FBB war die geringe zur Verfügung stehende Rechenleistung. Zum aktuellen Zeitpunkt steht dem Unternehmen nur ein Server für KI zur Verfügung. Auf diesem Server befindet sich normalerweise der Flughafenassistent „Freddy“, ein Chatbot welcher für Fragen rund um den Flughafen zur Verfügung steht. Dieser wird mit einem LLM der Firma Meta mit 70 Milliarden Parametern betrieben. Daraus resultiert die nächste Herausforderung. In dem Zeitraum der Nutzung für die Studienarbeit können andere Mitarbeiter nicht auf diesen Chatbot zugreifen, dies bedeutet, dass die Zeit zum Installieren, Testen und Auswerten begrenzt ist.

Die Auswahl geeigneter LLMs gestaltet sich aufgrund mehrerer Faktoren als herausfordernd. Einerseits beschränkt die verfügbare Rechenleistung die Auswahl auf Modelle mit etwa 70 Milliarden Parametern. Andererseits ist es für die Vergleichbarkeit der Modelle vorteilhaft, wenn sie vom selben Hersteller stammen, um Unterschiede in Trainingsdaten und Modellarchitekturen zu minimieren. Modelle wie Qwen von Alibaba Cloud oder DeepSeek eignen sich ebenfalls nicht für diese Studienarbeit, da diese LLMs nachweislich Antworten verzerren (Noels et al., 2025) und dementsprechend keine eindeutige Vergleichsgrundlage schaffen. Somit wird die Auswahl der Modelle stark eingegrenzt. Um die Modelle zu implementieren, müssen diese von „Hugging Face“ lokal heruntergeladen werden, dies ist bei größeren Modellen mit 70 Milliarden Parametern zeitintensiv. Des Weiteren ist der Zugriff auf verschiedene Modelle nicht jeder Person gestattet. Modelle müssen erst angefragt werden, bevor diese heruntergeladen werden können.

5.2.2 Auswahl der Modelle

Die Auswahl geeigneter LLMs gestaltet sich aufgrund der Vielzahl verfügbarer Anbieter und Modellgenerationen als herausfordernd. Insbesondere auf der Plattform „Hugging Face“ ist die Modellvielfalt sehr groß, allerdings existieren auch zahlreiche Modelle, die nicht primär für die Textgeneration konzipiert wurden. Zudem bieten nur wenige Unternehmen Modelle mit unterschiedlichen Parametergrößen an, weshalb der Vergleich hinsichtlich der Skalierbarkeit erheblich erschwert wird.

Wie bereits erwähnt, scheiden Modelle wie Qwen und DeepSeek aus verschiedenen Gründen aus. Im Fokus stehen deshalb weit verbreitere Modelle, welche bereits eine positive Resonanz erhalten haben, wie Meta-LLaMA, Google Gemma sowie verschiedene Modelle der Firma Mistral. Schlussendlich ist die Wahl auf die Gemma-Modelle von Google gefallen, die mit der kürzlich veröffentlichten „Gemma 3“ Reihe sowohl hinsichtlich Aktualität als auch hinsichtlich der verfügbaren Parametergrößen von 1, 4 und 12 Milliarden eine geeignete Grundlage für eine systematische Analyse bieten.

5.2.3 Erstellung der Fragen

Um KI-Modelle vergleichbar zu machen, wurden von verschiedenen Unternehmen und Forschungseinrichtungen standardisierte Benchmarks entwickelt. Dazu zählen unter anderem das Massive Multitask Language Understanding (MMLU)-Benchmark, das auf akademischem Weltwissen aus über 50 Disziplinen basiert (Hendrycks et al., 2021), der Grade School Math 8K (GSM8K)-Datensatz, der mathematische Problemlösen im Grundschulniveau testet, sowie „TruthfulQA“, das darauf abzielt, die Fakten- und Wahrheitsgenauigkeit von Modellen zu überprüfen.

Diese Benchmarks orientieren sich an etablierten Bewertungskriterien wie Genauigkeit, Robustheit, Fähigkeiten, Effizienz, Fairness sowie Nachvollziehbarkeit der Modellantworten.

Für diese Studienarbeit werden die Fragen in Anlehnung an diese Benchmarks erstellt. Mit dem Ziel, eine möglichst breite Abdeckung kognitiver Fähigkeiten, wie logisches Denken, Weltwissen, Sprachverständnis und mathematische Problemlösungen zu erzielen. Die Fragen werden in deutscher Sprache formuliert, um eine objektive und vergleichbare Bewertung der Modellantworten zu ermöglichen.

5.3 Durchführung

Der Vergleich der oben genannten KI-Modelle wird in JupyterLab durchgeführt. Dafür wird ein Python Skript entwickelt, welches die 150 Fragen aus einer Textdatei ausliest und nacheinander an das jeweilige Modell übergibt. Zum Festhalten der Messwerte gibt es zwei Dateien. Eine Datei dient dazu, sekündlich die aktuellen Werte des Video Random Access Memorys (VRAM) und Stromverbrauchs festzuhalten. Die zweite Datei beinhaltet durchschnittlichen Stromverbrauch, Antwort auf die Frage und den tatsächlichen Verbrauch (angegeben in Watt * Sekunde). Die Methoden zum Loggen der genannten

Werte wurden eigens implementiert. Mit diesen zwei Dateien wird eine differenzierte Vergleichbarkeit zwischen den Modellen ermöglicht. Um die Ergebnisse statistisch besser abzusichern, wird jedes Modell in drei unabhängigen Durchläufen getestet.
Die Tests wurden auf einem Server mit einer NVIDIA L40S Graphics Processing Unit (GPU) durchgeführt, die über 46.068 MiB Grafikspeicher verfügt. Zur Sicherstellung der Testbedingungen wurden allen Modellen die gleichen Werte übergeben:

- Die **Maximale Tokenanzahl** ist auf 2048 begrenzt, um überlange Antworten zu vermeiden.
- Das **Top-p Sampling** (top_p = 0.7) erhöht die inhaltliche Vielfalt der Antworten, ohne sie unkontrollierbar zu machen.
- Die **Temperatur** (temperature = 0.5) sorgt für fokussierte, weniger kreative, aber präzisere Ausgaben.
- Die **Repetitionsstrafe** (repition_penalty = 1.2) verhindert, dass wiederholte Ausgaben identischer Satzstrukturen erscheinen.

Diese Einstellungen werden gewählt, um die Modelle möglichst effizient und kontrolliert zu betreiben.

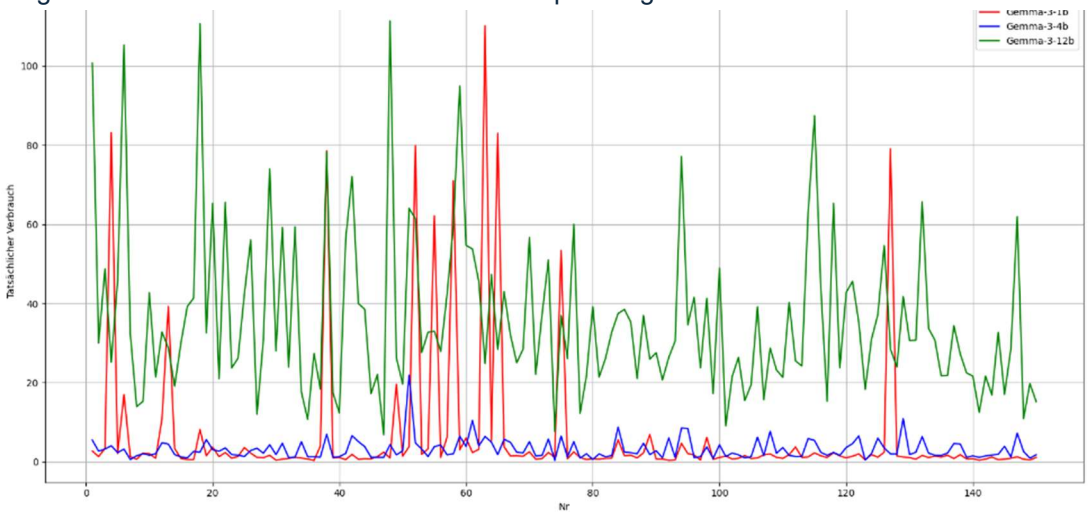
Am Ende jedes Durchlaufes werden die Messwerte in einzelnen comma-separated values-Dateien (CSV) gespeichert. So besteht die Möglichkeit einer strukturierten Auswertung und statistischen Analyse.

5.4 Auswertung

Ziel des praktischen Teils war es, verschiedene KI-Modelle unter dem Aspekt der Nachhaltigkeit miteinander zu vergleichen. Dabei wurden Stromverbrauch, Antwortgenauigkeit und die daraus resultierende Energieeffizienz untersucht. Der Stromverbrauch wurde dabei pro Anfrage über die Rechenzeit gemessen. Aus dieser Messung ergibt sich der tatsächliche Verbrauch, welcher ein guter Maßstab für den Stromverbrauch ist.

Abbildung 4:

Vergleich der Modelle: tatsächlicher Verbrauch pro Frage



Quelle: Eigene Darstellung

Wie in Abbildung 4 zu sehen ist, ist der Verlauf vom Modell mit 1 Milliarde (1B) und 4 Milliarden (4B) Parametern im tatsächlichen Stromverbrauch pro Frage ähnlich. Beim Modell mit 1B Parametern traten jedoch vereinzelte Leistungsspitzen – sogenannte Peaks – auf. Diese Leistungsspitzen sind auf längere Berechnungszeiten und damit einen höheren Stromverbrauch über die Zeit, zurückzuführen. Das Modell mit 12 Milliarden Parametern (12B) zeigte hingegen über nahezu den gesamten Testzeitraum hinweg einen erhöhten Energieverbrauch. Dafür gibt es zwei Hauptursachen: Erstens war die verwendete Grafikkarte maximal ausgelastet, was zu einer längeren Rechenzeit geführt hat. Zweitens benötigt ein größeres Modell, wie beispielsweise das 12B-Modell, mehr Ressourcen als ein 1B- oder 4B-Modell.

Ein weiterer Aspekt der Auswertung betrifft die Qualität der generierten Antworten. Hier schnitt das Gemma-3-1B-Modell mit lediglich 66,5 % korrekt beantworteter Fragen über drei Testdurchläufe hinweg deutlich schlechter ab. Das Gemma-3-4B-Modell erreichte 91,33 %, während das Gemma-3-12B Modell mit 96,67 % die höchste Genauigkeit zeigt.

Werden nun die Antwortgenauigkeit in Relation zum Energieverbrauch gesetzt, ergibt sich folgendes Bild:

- Das 1B-Modell belegt den dritten Platz. Es konnte in Genauigkeit nicht überzeugen und zeigte zudem einzelne ineffiziente Verbrauchsspitzen.
- Das 12B-Modell erreicht den zweiten Platz. Zwar lieferte dieses Modell die besten Ergebnisse bei der Antwortgenauigkeit, jedoch auf Kosten eines höheren Stromverbrauchs.
- Das 4B-Modell stellt den effizientesten Kompromiss dar: Es kombiniert einen niedrigen Energieverbrauch mit einer hohen Antwortgenauigkeit und belegt somit den ersten Platz im Vergleich.

Zusammenfassend lässt sich festhalten, dass bei der Gewichtung von Nachhaltigkeit und Leistungsfähigkeit nicht zwingend das Modell mit der geringsten Anzahl an Parametern die optimale Wahl darstellt.

6 Fazit

Zusammenfassend lässt sich sagen, dass die im Rahmen dieses Projekts verfasste Vergleichsstudie erfolgreich verwirklicht werden konnte. So konnten mit PRISMA gezielt die Studien ausgewählt werden, die für das Thema „KI und Nachhaltigkeit“ relevant waren. Die Vergleichsstudie ergab das sehr eindeutige Gesamtbild, dass KI klar negative Auswirkungen auf die Nachhaltigkeit hat. So ist vor allem der Stromverbrauch zum Betrieb der Rechenzentren ein Haupttreiber der Emissionen, aber auch diejenigen Emissionen, die beim Abbau der Ressourcen zur Herstellung der Hardware entstehen, sind nicht zu unterschätzen. Die Studien zeigten aber auch den gegensätzlich Aspekt, dass KI dazu beitragen kann, Emissionen in bestimmten Bereichen zu reduzieren, weshalb die Beziehung zwischen KI und Nachhaltigkeit nicht einseitig betrachtet werden sollte. Alle drei Hauptaspekte, die in den Projektzielen formuliert wurden, wurden im Rahmen dieser Vergleichsstudie umfassend beleuchtet und konnten sogar um einen neuen erweitert werden.

Hinsichtlich des praktischen Teils konnten die Projektziele ebenso erreicht sein, obwohl bei der Umsetzung einige Probleme auftraten. So musste die ursprüngliche Herangehensweise geändert werden, da die Modellauswahl sowie die begrenzten Server-Ressourcen Schwierigkeiten bereiteten. Dennoch konnte am Ende eindeutig die Aussage getroffen werden, dass die Antworten umso präziser waren, je mehr Zeit und damit Strom die Beantwortung dieser Fragen in Anspruch nahm.

7 Ausblick

Auch wenn alle Projektziele erreicht werden konnten, gibt es dennoch zahlreiche interessante Aspekte, die in Zukunft zu diesem Thema noch betrachtet werden und an diese Studienarbeit vertiefend anknüpfen könnten.

Einerseits könnte die Vergleichsstudie um mehrere Themen erweitert werden. So ergab die Analyse weitere Vergleichspunkte wie „Rechtliche Regelungen“, „Gesellschaftliche Entwicklungen“, „Wasserverbrauch“ oder „Messung“, die in dieser Studienarbeit jedoch nicht mehr betrachtet werden konnten.

Andererseits könnte der Vergleich der KI-Modelle intensiviert werden. Denn neben der Gegenüberstellung von Modellen mit unterschiedlicher Parameteranzahl könnten auch allgemeine und auf bestimmte Aufgaben spezialisierte Modelle gegenübergestellt werden. So ließen sich Rückschlüsse darauf ziehen, wie effizient spezialisierte Modelle tatsächlich sind und welches Potenzial sie hinsichtlich Nachhaltigkeit haben. Ebenso könnten Modelle mit gleicher Parameteranzahl, aber von verschiedenen Unternehmen wie NVIDIA, Google und Meta, verglichen werden, um zu ermitteln, ob Architektur und Trainingsmethoden Einfluss auf den Ressourcenverbrauch haben. Parallel dazu könnte ein Tool entwickelt werden, dass die CO_2-Emissionen oder sogar den reinen Stromverbrauch in Größen aus dem Alltag wie Flugzeug- oder PKW-Kilometer umrechnet. So könnten die Auswirkungen von KI auf die Nachhaltigkeit veranschaulicht werden.

Literaturverzeichnis

Acun, B., Lee, B., Kazhamiaka, F., Maeng, K., Chakkaravarthy, M., Gupta, U., Brooks, D. & Wu, C.-J. (2023). *Carbon Explorer: A Holistic Approach for Designing Carbon Aware Datacenters*. 118–132. https://doi.org/10.1145/3575693.3575754

Bhardwaj, E., Alexander, R. & Becker, C. (2025). *Limits to AI Growth: The Ecological and Social Consequences of Scaling* (arXiv:2501.17980; Version 1). arXiv. https://doi.org/10.48550/arXiv.2501.17980

Bruhn, L. (2016). *Intelligenz: Was ist das genau?* https://www.neuronation.com/science/de/definition-der-intelligenz-was-ist-das-eigentlich/

Cornelsen Verlag GmbH. (o. J.). *Kognitiv ▶ Rechtschreibung, Bedeutung, Definition, Herkunft ▶ Duden*. Duden. Abgerufen 1. August 2025, von https://www.duden.de/rechtschreibung/kognitiv

CO2 Äquivalente | ClimatePartner. (o. J.). Abgerufen 4. August 2025, von https://www.climatepartner.com/de/wissen/glossar/co2-aequivalente-co2e

De Vries, A. (2023). The growing energy footprint of artificial intelligence. *Joule*, *7*(10), 2191–2194. https://doi.org/10.1016/j.joule.2023.09.004

DeepSeek. (2025). *DeepSeek-R1 Release*. DeepSeek API Docs. Abgerufen 1. August 2025, von https://api-docs.deepseek.com/news/news250120

Desroches, C., Chauvin, M., Ladan, L., Vateau, C., Gosset, S. & Cordier, P. (2025). *Exploring the sustainable scaling of AI dilemma: A projective study of corporations' AI environmental impacts* (arXiv:2501.14334). arXiv. https://doi.org/10.48550/arXiv.2501.14334

Ding, C., Ke, J., Levine, M. & Zhou, N. (2024). Potential of artificial intelligence in reducing energy and carbon emissions of commercial buildings at scale. *Nature Communications*, *15*(1), 5916. https://doi.org/10.1038/s41467-024-50088-4

Engel, J. (2021). Embodied carbon vs operational carbon: What's the difference, and why does it matter? *Factor This™*. https://www.renewableenergyworld.com/energy-business/policy-and-regulation/embodied-carbon-vs-operational-carbon-whats-the-difference-and-why-does-it-matter/

Ertel, W. (2025). *Grundkurs Künstliche Intelligenz: Eine praxisorientierte Einführung*. Springer Fachmedien. https://doi.org/10.1007/978-3-658-44955-1

Fraisl, D., See, L., Fritz, S., Haklay, M. & McCallum, I. (2025). Leveraging the collaborative power of AI and citizen science for sustainable development. *Nature Sustainability*, *8*(2), 125–132. https://doi.org/10.1038/s41893-024-01489-2

Google. (2024). *Gemini-Apps: Release-Updates und Verbesserungen*. Gemini. Abgerufen 1. August 2025, von https://gemini.google.com/updates

Guidi, G., Dominici, F., Gilmour, J., Butler, K., Bell, E., Delaney, S. & Bargagli-Stoffi, F. J. (2024). *Environmental Burden of United States Data Centers in the Artificial Intelligence Era* (arXiv:2411.09786; Version 1). arXiv. https://doi.org/10.48550/arXiv.2411.09786

Hacker, P. (2023). *Sustainable AI Regulation* (SSRN Scholarly Paper No. 4467684). Social Science Research Network. https://doi.org/10.2139/ssrn.4467684

Harwardt, M. & Köhler, M. (2023). Künstliche Intelligenz. In M. Harwardt & M. Köhler (Hrsg.), *Künstliche Intelligenz entlang der Customer Journey: Einsatzpotenziale von KI im E-Commerce* (S. 21–29). Springer Fachmedien. https://doi.org/10.1007/978-3-658-39109-6_3

Hendrycks, D., Burns, C., Basart, S., Zou, A., Mazeika, M., Song, D. & Steinhardt, J. (2021). *Measuring Massive Multitask Language Understanding* (arXiv:2009.03300). arXiv. https://doi.org/10.48550/arXiv.2009.03300

Holder, C., Khurana, V. & Watts, M. (2018). *Artificial Intelligence: Public perception, attitude and trust. Bristows*. https://www.bristows.com/app/uploads/2019/06/Artificial-Intelligence-Public-Perception-Attitude-and-Trust.pdf

IBM. (2023). *Was ist ein KI-Modell?*. Abgerufen 1. August 2025, von https://www.ibm.com/de-de/think/topics/ai-model

Lee, B. C., Brooks, D., Benthem, A. van, Gupta, U., Hills, G., Liu, V., Pierce, B., Stewart, C., Strubell, E., Wei, G.-Y., Wierman, A., Yao, Y. & Yu, M. (2024). *Carbon Connect: An Ecosystem for Sustainable Computing* (arXiv:2405.13858). arXiv. https://doi.org/10.48550/arXiv.2405.13858

Li, P., Yang, J., Islam, M. A. & Ren, S. (2025). *Making AI Less „Thirsty“: Uncovering and Addressing the Secret Water Footprint of AI Models* (arXiv:2304.03271). arXiv. https://doi.org/10.48550/arXiv.2304.03271

Li, Z. S., Arony, N. N., Awon, A. M., Damian, D. & Xu, B. (2024). *AI Tool Use and Adoption in Software Development by Individuals and Organizations: A Grounded Theory Study* (No. arXiv:2406.17325). arXiv. https://doi.org/10.48550/arXiv.2406.17325

Litzel, N. & Luber, S. (2019). *Was ist eine Tensor Processing Unit (TPU)?* BigData-Insider. https://www.bigdata-insider.de/was-ist-eine-tensor-processing-unit-tpu-a-750292/

Luber, S. (2024). *Was ist eine Black Box?* BigData-Insider. https://www.bigdata-insider.de/was-ist-eine-black-box-a-9d64b128e437b6a727849cf82695dbe9/

Luccioni, A. S., Jernite, Y. & Strubell, E. (2024a). Power Hungry Processing: Watts Driving the Cost of AI Deployment? *The 2024 ACM Conference on Fairness, Accountability, and Transparency*, 85–99. https://doi.org/10.1145/3630106.3658542

Luccioni, A. S., Trevelin, B. & Mitchel, M. (2024b). *The Environmental Impacts of AI – Primer*. Abgerufen 4. August 2025, von https://huggingface.co/blog/sasha/ai-environment-primer

Luccioni, A. S., Strubell, E. & Crawford, K. (2025). *From Efficiency Gains to Rebound Effects: The Problem of Jevons' Paradox in AI's Polarized Environmental Debate* (arXiv:2501.16548; Version 1). arXiv. https://doi.org/10.48550/arXiv.2501.16548

Luccioni, A. S., Viguier, S. & Ligozat, A.-L. (2022). *Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model* (arXiv:2211.02001). arXiv. https://doi.org/10.48550/arXiv.2211.02001

Microsoft. (2023). *Release Notes for Microsoft 365 Copilot*. Microsoft Learn. Abgerufen 1. August 2025, von https://learn.microsoft.com/en-us/copilot/microsoft-365/release-notes

1987: Brundtland Report. (o. J.). Abgerufen 4. August 2025, von https://www.are.admin.ch/are/en/home/media/publications/sustainable-development/brundtland-report.html

Naveed, H., Khan, A. U., Qiu, S., Saqib, M., Anwar, S., Usman, M., Akhtar, N., Barnes, N., & Mian, A. (2025). A Comprehensive Overview of Large Language Models. *ACM Trans. Intell. Syst. Technol.* https://doi.org/10.1145/3744746

Noels, S., Bied, G., Buyl, M., Rogiers, A., Fettach, Y., Lijffijt, J. & Bie, T. D. (2025). *What Large Language Models Do Not Talk About: An Empirical Study of Moderation and Censorship Practices* (arXiv:2504.03803). arXiv. https://doi.org/10.48550/arXiv.2504.03803

OECD. (2022). *Measuring the environmental impacts of artificial intelligence compute and applications*. https://www.oecd.org/en/publications/measuring-the-environmental-impacts-of-artificial-intelligence-compute-and-applications_7babf571-en.html

OECD. (2025). *Unlocking productivity with generative AI: Evidence from experimental studies*. https://www.oecd.org/en/blogs/2025/07/unlocking-productivity-with-generative-ai-evidence-from-experimental-studies.html

Olawade, D. B., Wada, O. Z., David-Olawade, A. C., Fapohunda, O., Ige, A. O. & Ling, J. (2024). Artificial intelligence potential for net zero sustainability: Current evidence and prospects. *Next Sustainability*, *4*, 100041. https://doi.org/10.1016/j.nxsust.2024.100041

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., … Moher, D. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, n71. https://doi.org/10.1136/bmj.n71

Patterson, D., Gonzalez, J., Le, Q., Liang, C., Munguia, L.-M., Rothchild, D., So, D., Texier, M. & Dean, J. (2021). *Carbon Emissions and Large Neural Network Training* (arXiv:2104.10350). arXiv. https://doi.org/10.48550/arXiv.2104.10350

Peng, S., Kalliamvakou, E., Cihon, P. & Demirer, M. (2023). *The Impact of AI on Developer Productivity: Evidence from GitHub Copilot* (No. arXiv:2302.06590). arXiv. https://doi.org/10.48550/arXiv.2302.06590

Rolnick, D., Donti, P. L., Kaack, L. H., Kochanski, K., Lacoste, A., Sankaran, K., Ross, A. S., Milojevic-Dupont, N., Jaques, N., Waldman-Brown, A., Luccioni, A. S., Maharaj, T., Sherwin, E. D., Mukkavilli, S. K., Kording, K. P., Gomes, C. P., Ng, A. Y., Hassabis, D., Platt, J. C., … Bengio, Y. (2022). Tackling Climate Change with Machine Learning. *ACM Computing Surveys*, *55*(2), 1–96. https://doi.org/10.1145/3485128

Rüdiger, A. (2023). *Was ist Sustainability by Design?* DataCenter-Insider. https://www.datacenter-insider.de/was-ist-sustainability-by-design-a-aec206214ddb688e58563097df903345/

Shewly Bala, Abdulla Al Shiam, S., Shamsul Arefeen, S. M., Abir, S. I., Hemel Hossain, Hossain, M. S., Shoha, S., Akhter, A., Mohammad Ridwan & Sumaira. (2024). Measuring How AI Innovations and Financial Accessibility Influence Environmental Sustainability in the G-7: The Role of Globalization with Panel ARDL and Quantile Regression Analysis. *Global Sustainability Research*, 1–29. https://doi.org/10.56556/gssr.v3i4.974

Statista. (2024). *USA - CO2-Emissionen bis 2023*. Abgerufen 7. Juli 2025, von https://de.statista.com/statistik/daten/studie/1382237/umfrage/entwicklung-der-co2-emissionen-in-den-usa/

Stiftung myclimate. (2025). *CO_2 Rechner: CO_2 Ausstoss berechnen*. Myclimate. Abgerufen 1. August 2025, von https://co2.myclimate.org/de/portfolios?calculation_id=8076071

Taşdelen, A. (2024). *Enhancing green computing through energy-aware training: An early stopping perspective*. Karabuk University - Computer Engineering and Software Engineering Departments. https://doi.org/10.71074/ctc.1594291

Tomlinson, B., Black, R. W., Patterson, D. J. & Torrance, A. W. (2024). The carbon emissions of writing and illustrating are lower for AI than for humans. *Scientific Reports*, *14*(1), 3732. https://doi.org/10.1038/s41598-024-54271-x

Varoquaux, G., Luccioni, A. S. & Whittaker, M. (2025). *Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI* (arXiv:2409.14160). arXiv. https://doi.org/10.48550/arXiv.2409.14160

Wayback Machine. (2011, Juni 19). https://web.archive.org/web/20110619071609/http://www.bdi.eu/download_content/PCF-Leitfaden_100810_Online.pdf

Wu, C.-J., Raghavendra, R., Gupta, U., Acun, B., Ardalani, N., Maeng, K., Chang, G., Behram, F. A., Huang, J., Bai, C., Gschwind, M., Gupta, A., Ott, M., Melnikov, A., Candido, S., Brooks, D., Chauhan, G., Lee, B., Lee, H.-H. S., … Hazelwood, K. (2022). *Sustainable AI: Environmental Implications, Challenges and Opportunities* (arXiv:2111.00364). arXiv. https://doi.org/10.48550/arXiv.2111.00364

Zhou, C., Li, Q., Li, C., Yu, J., Liu, Y., Wang, G., Zhang, K., Ji, C., Yan, Q., He, L., Peng, H., Li, J., Wu, J., Liu, Z., Xie, P., Xiong, C., Pei, J., Yu, P. S. & Sun, L. (2023). *A Comprehensive Survey on Pretrained Foundation Models: A History from BERT to ChatGPT* (No. arXiv:2302.09419). arXiv. https://doi.org/10.48550/arXiv.2302.09419

Abbildungsverzeichnis

Anhangsverzeichnis

Anhang 1: Inkludierte Studien

Siehe Datei „PRISMA_inkludierte_Studien.xlsx“

Anhang 2: Python Skripte

Siehe Datei „python_skripte_gemma_3.zip“

Ehrenwörtliche Erklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit in allen Teilen selbstständig angefertigt und keine anderen als die in der Arbeit angegebenen Quellen und Hilfsmittel benutzt habe. Sämtliche wörtlichen oder sinngemäßen Übernahmen und Zitate, sowie alle Abschnitte, die mithilfe von KI-basierten Tools entworfen, verfasst und/oder bearbeitet wurden, sind kenntlich gemacht und nachgewiesen. Im Anhang meiner Arbeit habe ich sämtliche KI-basierte Hilfsmittel angegeben. Diese sind mit Produktnamen und formulierten Eingaben (Prompts) in einem KI-Verzeichnis ausgewiesen.

Ich bin mir bewusst, dass die Verwendung von Texten oder anderen Inhalten und Produkten, die durch KI-basierte Tools generiert wurden, keine Garantie für deren Qualität darstellt. Ich verantworte die Übernahme jeglicher von mir verwendeter maschinell generierter Passagen vollumfänglich selbst und trage die Verantwortung für eventuell durch die KI generierte fehlerhafte oder verzerrte Inhalte, fehlerhafte Referenzen, Verstöße gegen das Datenschutz- und Urheberrecht oder Plagiate.

Kolbe, Fynnian — Digital unterschrieben von Kolbe, Fynnian Datum: 2025.08.19 14:55:59 +02'00'

Jentsch, Jamie — Digital unterschrieben von Jentsch, Jamie Datum: 2025.08.19 14:53:40 +02'00'

Ritschel, Finn — Digital unterschrieben von Ritschel, Finn Datum: 2025.08.19 14:59:02 +02'00'

(Datum, Unterschrift Studierende)