Automatisierungspotenziale von Büro- und Verwaltungsarbeiten anhand von Large Language Models am Beispiel von ChatGPT

Sina Stumpp

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
E-Mail: stumppsi@hs-pforzheim.de

Frank Morelli

Hochschule Pforzheim
Tiefenbronner Straße 65
75175 Pforzheim
E-Mail: frank.morelli@hs-pforzheim.de

ABSTRACT

Dieser Beitrag untersucht das Potenzial von Large Language Models (LLMs) am Beispiel von ChatGPT für die Automatisierung von Büroarbeiten in Beratungsunternehmen. Auf der Basis von Experteninterviews werden zwei Anwendungsfälle identifiziert und Kriterien für die Anwendung des Tools definiert. Es handelt sich dabei um "Erstellung von PowerPoint-Präsentationen" und "Deeskalation beim Verfassen von E-Mails". Die Studie unterstreicht die Notwendigkeit einer robusten KI-Strategie und intensiver Schulungen, um Risiken zu minimieren und Mitarbeiter zu schulen.

SCHLÜSSELWÖRTER –

Large Language Model, ChatGPT, Office Automation, Consulting, Transformer

EINLEITUNG

Natural Language Processing (NLP) als Teilgebiet der Künstlichen Intelligenz (KI) konzentriert sich auf sprachliche Disziplinen. Diese Disziplin hat zum Ziel, natürliche menschliche Sprache zu verstehen und zu generieren, um eine generische Konversation zwischen Mensch und Maschine zu ermöglichen und gleichzeitig sprachliche Herausforderungen wie grammatische Regeln, Ausnahmen, Zweideutigkeiten, Dialekte, Ironie und Sprachwandel zu bewältigen (Ghavami 2020; Moloi/Marwala 2021; Taulli 2022; Wang u. a. 2023).

Der Launch der Webanwendung ChatGPT gab im November 2022 den Startschuss für eine neue Welle disruptiver Sprachtechnologien (Chui u. a. 2016; Haleem u. a. 2022). Dieses Large Language Modell (LLM), mit der Fähigkeit natürliche Konversationen zu führen und permanent Wissen zur Verfügung zu stellen, hat seither Menschen weltweit begeistert. ChatGPT basiert auf dem NLP-Lernprozess: Während der Vorverarbeitung kommen dabei Tokenization, Part of Speech Tagging, Named Entity Recognition, Stemming und Stop Word Identification als NLP-Techniken innerhalb der „input embedding"-Schicht zum Einsatz (Onslow 2023; Ghavami 2020; Kuiler 2022).

Neben textintensiven Sektoren, wie der Bildungs- und Forschungsbranche, beschäftigen sich auch Beratungsunternehmen mit der Technologie (ZBW - Leibniz Information Centre for Economics 2023). Zum einen zeichnet es sich ab, dass die Implementierung von digitalen Technologien zu Produktivitätssteigerung in der internationalen Arbeitswelt führt (Dwivedi u. a. 2023). Zum anderen ist davon auszugehen, dass der Umgang mit KI-Technologien und -Tools wie ChatGPT für BeraterInnen in Zukunft wichtig sein wird. Die Suche nach potenziellen Automatisierungfällen innerhalb eines Büro- und Verwaltungsrahmens erweist sich im Kontext von LLMs als wesentlicher Bestandteil für eine zukunftsweisende Strategie. Sie hat das Potenzial, Arbeit effizienter zu gestalten. Die praktische Implementierung solcher Tools stellt Unternehmen jedoch vor Herausforderungen, da bisher die Einordnung und Konkretisierung von benutzerfreundlichen Einsatzszenarien fehlen: Zum gegenwärtigen Zeitpunkt besteht kein vollumfängliches Verständnis für die realistische Leistungsfähigkeit der Technologie. Die Integration von LLMs in Unternehmensumgebungen und Verwaltungsprozesse wirft zudem noch unbeantwortete Fragen der Datensicherheit, Ethik und Verantwortlichkeit auf. Diese sind aber entscheidend für eine zugehörige Implementierung in die Unternehmenslandschaft.

Ziel der Arbeit ist es, die Identifikation von geeigneten Anwendungsfällen der Technologie ChatGPT, basierend auf dem kostenlos zugänglichen Modell GPT 3.5, für den regelmäßigen Einsatz innerhalb von Beratungsunternehmen. Dabei sollen Potenziale zur Einsparung von Arbeitszeit und Ressourcen identifiziert sowie passende Anleitungen zur Automatisierung der Aufgaben mit ChatGPT erstellt werden. Als Rahmenbedingung ist die aktuelle Gesetzgebung (z.B. KI-Verordnung und DSGVO) zu berücksichtigen.

Die gewonnenen Erkenntnisse basieren zum einen auf einer qualitativen Datenerhebung mit Hilfe von fünf Experteninterviews. Es wurden sowohl Personen aus leitenden Positionen im Bereich KI-Consulting und klassischer Unternehmensberatung befragt, sowie Personen, die durch ihren Beruf als Programmierer mit ChatGPT konfrontiert waren. Zum anderen erfolgt die Analyse zweier Use Cases im Büro- und Verwaltungskontext, die sich im Rahmen der Experteninterviews als relevante Anwendungsfälle herauskristallisiert haben. Im einzelnen handelt es sich dabei um „PowerPoint Präsentation erstellen"

und „Erstellung einer E-Mail zur Deeskalation in der Kundenkommunikation“.

Large Language Models (LLMs)

Aufgrund ihrer Leistungsfähigkeit haben sich Large Language Models (LLMs), zu denen auch ChatGPT gehört, als neue Subkategorie von Sprachmodelle etabliert. Ein essenzieller Indikator für den Wissensumfang eines Sprachmodells ist die Parameteranzahl, ein Synonym für die Anzahl an Gewichten, die bestimmen wie das Netzwerk auf Eingaben reagiert und somit auch die Ausgaben angibt. Als „large“ gilt ein LLM, wenn es eine hohe Anzahl an Parametern besitzt und das Training mit Big Data, in Form von Textdaten, durchgeführt wurde. Studien belegen, dass eine wachsende Größe des LLMs die Leistung des Modells in erster Linie positiv beeinflusst, sodass derzeit der Trend dahin geht, immer größere LLMs zu entwickeln (Bender 2021; Goetze/Abramson 2021).
Grundsätzlich wird zum einen darauf abgezielt, korrekte natürliche Sprache zu erzeugen, die von Menschen nicht als maschinell generiert einzuordnen ist. Zum anderen geht es darum, dass Modelle selbst ein Sprachverständnis generieren. Darauf basierend können LLMs komplexe Textaufgaben wie Fragenbeantwortung, Klassifizierung etc. lösen (Zhao u. a. 2023). Mithilfe der Berechnung von Wahrscheinlichkeiten für Wörter und deren Reihenfolge werden dafür autoregressiv Wortsequenzen erzeugt (Bender u. a. 2021; Ghavami 2020).
Die Eingabe erfolgt über Prompts, d.h. schriftliche Befehle der Nutzer, die tokenisiert in die Encoderschicht übermittelt werden. LLMs sollen dabei angepasst an Nutzererwartungen handeln. Dafür werden meist entsprechende Beispiele in den Trainingsdatensatz integriert, damit das Modell anhand derer sprachliche Konzepte hinzulernen kann. In diesem Zusammenhang müssen Wahrheit und Unvoreingenommenheit der Ausgaben gesichert sein sowie Nutzungen mit falschen, illegalen oder gefährlichen Intentionen unterbunden werden (Ouyang u. a. 2022). Mittlerweile haben einige LLMs die Fähigkeit sich einen „moralischen Kompass“ anzueignen, indem sie schädliche Konzepte wie bspw. Diskriminierung oder Stereotypen als Teil ihres Trainings erkennen oder speziell von Entwicklern dazu instruiert werden. Anschließend ist es ihnen möglich, sich selbst und andere moralisch zu korrigieren oder Befehle abzuweisen (Ganguli u. a. 2023).
Neben der Möglichkeit zur Spezialisierung auf bestimmte Anwendungsbereiche, besitzen LLMs die Kompetenz zur Generalisierung. Dies ermöglicht die Bearbeitung von unbekannten Befehlen bzw. beschreibt die Fähigkeit, aus dem Training gezogenes,Wissen auf neue Situationen anzuwenden und kontextspezifische Antworten liefern können (Ferrara 2023).
Durch intensives Training mit Textdokumenten erlangen LLMs ein tiefgehendes Verständnis für menschliche Sprache, ermöglichen die Generierung von natürlichen Texten und können zudem multimodale Daten wie Bilder und Videos verarbeiten (Lund u. a. 2023; Yin u. a. 2023). Die Anwendungsbereiche sind vielseitig und umfassen Textgenerierung, Übersetzung, Klassifizierung, Informationsextraktion, Fragenbeantwortung und Sentiment-Analyse. LLMs bieten auch Formulierungshilfen, unterstützen kreative Inhalte, verstehen Emotionen in Texten, können auf wenigen Beispielen basierend generalisieren (Chohan 2023; Chui u. a. 2022; Zhang u. a. 2023; Zhao u. a. 2023). Ihr Verständnis für Kontext und Zusammenhänge ermöglicht situationsabhängige Reaktionen (Dwivedi u. a. 2023; Haleem u. a. 2022).

Sind LLMs auf Mensch-Computer-Dialoge ausgerichtet sind, ermöglicht dies natürliche Konversationen (Bozkurt u. a. 2023; Haleem u. a. 2022). Zudem agieren sie dabei als effektive Assistenten oder Co-Autoren in Zusammenarbeit mit Menschen, ohne diese zu ersetzen (Polonsky/Rotman 2023). Darüber hinaus erleichtern LLMs die Automatisierung von wiederholenden, zeitaufwändigen Textaufgaben. Auf Basis von maßgeschneiderten Trainingsdaten können LLMs flexibel an individuelle Szenarien angepasst werden, was eine vielseitige Anwendung ermöglicht (Ferrara, 2023). LLMs bewältigen mühelos große Mengen an unstrukturierten Daten, d.h. sie benötigen keine vorherige Datenstrukturierung und können Informationen aus weniger organisierten Quellen verarbeiten (Bender u. a. 2021; Haleem u. a. 2022).
Zudem lassen sie sich nahtlos in bestehende Systeme und Workflows integrieren, um zusätzliche Funktionalitäten und Verbesserungen bereitzustellen (OpenAI 2023a). Durch die Implementierung von LLMs in Geschäftsprozesse können weiterführend Routineaufgaben automatisiert werden und Mitarbeiter besitzen die Möglichkeit sich auf anspruchsvollere Aufgaben zu fokussieren (ZBW – Leibniz Information Centre for Economics 2023). Somit optimieren sie bei einer effektiven Nutzung Prozesse, sparen Zeit, steigern die Unternehmensproduktivität und tragen zur Kosteneffizienz bei, sodass Unternehmen sich dementsprechend einen Wettbewerbsvorteil sichern und das Wirtschaftswachstum fördern können (Dwivedi u. a. 2023; Goldstein u. a. 2023; Lund u. a. 2023).
In der Betrachtung der Risiken von LLMs offenbart sich eine Vielzahl von Herausforderungen, die sowohl ethische als auch praktische Implikationen mit sich bringen. Der undurchsichtige "Black Box"-Aufbau von LLMs führt zu einer mangelnden Nachvollziehbarkeit ihrer Ausgaben, da der komplexe interne Mechanismus nicht immer logisch ersichtlich ist (Aiyappa u. a. 2023). Diese Intransparenz birgt das Risiko von Falschinformationen, die durch Halluzinationen seitens der Modelle erzeugt werden (Wang u. a. 2023). Datenschutz wird ebenfalls gefährdet, da Eingabedaten in Trainigsdaten integriert werden, was nicht nur unerwünschte Datenschutzrisiken schafft, sondern auch LLMs zu potenziellen Zielen für Cyberangriffe macht (Kuiler 2022). Zusätzlich besteht die Gefahr, dass sich trotz Anonymisierung von Eingabedaten Personen oder Unternehmen durch Kontexthinweise identifizieren lassen (Patsakis/Lykousas 2023).
Die Fehlbarkeit des Trainings durch menschlichen Einfluss sowie die Unzureichendkeit von repräsentativen

Trainingsdaten verstärken Verzerrungen in den generierten Ausgaben und verursachen eine erhöhte Wahrscheinlichkeit für toxische und von Stereotypen durchzogene Inhalte (Bender u.a. 2021; Goldstein u.a. 2023; Ouyang u. a. 2022).
Diese Risiken erstrecken sich über verschiedene Bereiche, von der individuellen Nutzung bis hin zu potenziellen Auswirkungen auf Arbeitsmärkte und die Integrität wissenschaftlicher Forschung (Chohan 2023; Polonsky/Rotman 2023). Eine weitere Herausforderung besteht darin, dass die Anwendung von LLMs eine fundierte Kenntnis der korrekten Formulierung von Prompts erfordert, um effektiv genutzt werden zu können (Goldstein u. a. 2023). Insgesamt verdeutlichen die geannten Risiken die Notwendigkeit eines umfassenden Verständnisses und einer kritischen Reflexion im Umgang mit LLMs in diversen Anwendungsbereichen, um einem unberechtigt hohen Vertrauen in die KI vorzubeugen (Kasneci u. a. 2023).

ChatGPT

ChatGPT ist ein von OpenAI entwickeltes LLM, das seit seiner Veröffentlichung im November 2022 über die Grenzen der Technologiebranche hinaus große Wellen geschlagen und aufgrund seiner disruptiven Art einen neuen Standard in linguistischer Intelligenz gesetzt hat (Lund u. a. 2023; Wang u. a. 2023). Mit über 100 Millionen Nutzern weltweit innerhalb der ersten zwei Monate gilt das, in der Basisversion kostenlos zugängliche, LLM als erstes Sprachmodell mit Massenanwendung (Aiyappa u. a. 2023). Ein Grund dafür ist die Entwicklung des Modells als Chatbot, was die Zugänglichkeit für technisch-unaffine Menschen erleichtert (Dwivedi u. a. 2023).
Über ein Texteingabefeld können Nutzer Prompts formulieren und im Anschluss beantwortet ChatGPT die Befehle durch das Generieren von Text. In die Kategorie der Chatbots fällt ChatGPT durch die Fähigkeit, Bezug auf vorherige Prompts und Antworten zu nehmen, und somit einen menschlichen Dialog mit dem Nutzer zu führen (Lund u. a. 2023). ChatGPT kann Rückfragen stellen, unangemessene Anfragen verweigern, Fehler eingestehen sowie korrigieren, sodass es sehr intuitiv wirkt und den Anschein erwecken kann, es würde eine Art Gewissen besitzen (Ganguli u. a. 2023).
Zur Entwicklung von GPT-3 wurde das transformerbasierte neuronale Netzwerk zunächst mit einem sehr großen Datensatz, bestehend aus einer Vielzahl von ungekennzeichneten Textkorpora, trainiert, sodass dieses ein umfassendes Verständnis von Sprache entwickelt (Kasneci u.a. 2023; Narasimhan u.a. 2018). In diesem Schritt lernt ein Modell, wie Sprache gebildet wird, wie Kontext Aussagen beeinflusst und welche Bedeutung hinter Wörtern und Kombinationen dieser stecken, indem es Muster und Beziehungen in den Trainingsdaten enthaltenen Texten erkennt und seine Parameter entsprechend anpasst (Ferrara 2023; Zhao u.a. 2023). Beim Training von ChatGPT wurde dafür eine gefilterte Version von CommonCrawl, einer Bündelung von Websites und Dokumenten, die über die letzten acht Jahre durch Webcrawling gesammelt wurden, der Datensatz WebText, der Redditeinträge, inklusive angegebener Quellen, beinhaltet, sowie zwei Buchkorpora und alle englischen Wikipediaeinträge verwendet (Patsakis/Lykousas 2023). Auch wenn ChatGPT in über 50 Sprachen, wie bspw. Englisch, Spanisch, Französisch, Deutsch, Arabisch und Japanisch, nutzbar ist, bestanden mindestens 96 % der Trainingsdaten aus englischen Daten (Ouyang u. a. 2022). Um wie viele Dokumente und Wörter es sich genau handelt, wurde von OpenAI nicht veröffentlicht.

(1) Supervised Fine-Tuning:
Beim Vortraining wird ein Grundverständnis von Sprache vermittelt. Hierdurch besitzt das Sprachmodell zwar breitgefächerte Kompetenzen. Diese reichen jedoch i.d.R. nicht aus, um qualitative hochwertige Antworten in ausgewählten Anwendungsbereichen hervorzubringen. Entsprechend gilt es, diese mit mit Supervised Finetuning zu ermöglichen (Ouyang u. a. 2022). Hierbei wird das Sprachmodell mit einem kleineren, gelabelten und speziell auf das gewünschte Nutzungsverhalten ausgelegten Datensatz gefüttert (Wang u. a. 2023). Da ChatGPT in erster Linie einen einfachen Kommunikationsaustausch ermöglichen soll, wird das Modell in diesem Schritt für zwischenmenschliche Konversationen im Chatumfeld zugeschnitten. Eine ausgewählte Gruppe an Menschen, die als „Labeler“ (Kennzeichner) agieren, verfasst dafür eine Reihe an Prompts sowie die, nach ihrem Ermessen, gewünschten Ziel-Ausgaben auf diese, die das Modell, nach Verarbeitung der Daten, als Handlungsorientierung einsetzt (Ouyang u. a. 2022). Supervised Finetuning gilt als beliebter Ansatz, um den Aufwand, Daten manuell zu kennzeichnen, zu minimieren. Neben Domänenwissen lassen sich mit Hilfe von Finetuning auch bestimmte Richtlinien oder Einschränkungen implementieren, um klassische Risiken wie Bias oder toxische Ausgaben zu vermeiden und die Ausgaben an menschliche Wertevorstellungen anzupassen (Goldstein u. a. 2023; Ouyang u. a. 2022). Anschließend ist es dem Sprachmodell möglich sich selbst moralisch zu korrigieren.
(2) Reinforcement Learning from Human Feedback:
Das Sprachmodell ist zu diesem Zeitpunkt nicht dazu fähig, die Qualität von Antworten einzuschätzen, d.h. zu wissen, welche Antwort auf einen Prompt die beste ist, wenn es mehrere davon produziert. Deshalb findet die Anwendung eines Reward Models (Belohnungsmodell) Gebrauch (Gao u. a. 2023; Ouyang u. a. 2022). Zur Erstellung des Reward Models werden dem Sprachmodell verschiedene Prompts mehrmals gestellt, womit am Ende zu jedem Prompt eine Auswahl an Antwortmöglichkeiten existiert. Danach wird diese Sammlung Labelern vorgelegt, die die Ausgaben von „am besten“ zu „am schlechtesten“ ranken (Ziegler u. a. 2019). Wenn genug Rankingbeispiele gesammelt worden sind, erfolgt auf Basis der Feedbackdaten ein separates Reward Model Training, das vorhersagt, welche Ausgabe ein Mensch für eine gegebene Eingabe bevorzugt, indem es für jede Antwortmöglichkeit eine „Skalarbelohnung“ ausgibt (Ferrara 2023; Ouyang u. a. 2022). Anhand dessen wird eine Policy (Strategie) erlernt, die mit dem Proximal Policy

Optimization (PPO) Algorithmus, einer Gruppe von Policy Gradient Methoden von OpenAI, Skalarbelohnungen mittels Gradient Descent maximiert (Ouyang u. a. 2022; Schulman u. a. 2017). Das Sprachmodell gibt nach Implementierung der Policy bei Promptanfragen auf Nutzervorlieben zugeschnittene Texte aus und beendet den Trainingsvorgang, sodass ChatGPT von dort an einsatzbereit ist. Seit dem Launch von ChatGPT wurden mehrere geupdatete Versionen veröffentlicht, die laut OpenAI noch besser an die Nutzerwünsche angepasst wurden. Grund dafür ist vor allem der Einsatz von vergangenen Nutzeranfragen und Nutzerbewertungen zu den generierten Ausgaben als neue Trainingsdaten. ChatGPT wird somit aktiv weiterentwickelt und gewinnt mit mehr Nutzereingabedaten der auch an mehr Sprachverständnis (Aiyappa u. a. 2023).

Promptdesign

Die Formulierung der Prompts, innerhalb der Use Cases, ist an den gängigsten Promptdesign-Techniken sowie OpenAIs „GPT Best Practices“ orientiert (OpenAI 2023b). Grundlegend gibt es drei Ansätze für das Design von Prompts:

- *Zero-Shot:* Enthält ein Prompt eine Aufgabenstellung ohne passende Lösungsbeispiele, spricht man von einem zero-shot Prompt. Die Beantwortung des Befehls beruht stattdessen auf der Generalisierung von Trainingsdaten, indem mit allgemeinem Wissen und mit Hilfe von erkannten Mustern eine möglichst angemessene Lösung generiert wird.
- *One-Shot:* Liefert man dem Modell nur ein einziges Beispiel, in Form von einer Eingabe mit dazugehöriger Ausgabe, so kann ChatGPT, basierend darauf, eine eigene Ausgabe formulieren. Dafür lernt es aus diesem einzelnen Beispiel, um anschließend ähnliche Aufgaben ausführen können, ohne weitere Trainingsdaten zu benötigen.
- *Few-Shot:* Im Gegensatz zu einem one-shot Prompt, umfasst ein Few-Shot Prompt normalerweise mehrere Beispiele, mit denen das Modell die Aufgabe bewältigen soll.

Es kann hilfreich sein, ChatGPT mit one- oder few-shot Prompts in eine gewünschte Richtung zu lenken, wenn das Modell Schwierigkeiten hat einen Befehl zu verstehen (Brown u. a. 2020).

USE CASES

Die Auswahl der Use Cases ist auf Basis von durch Experteninterviews ermittelten Anwendungsbeispielen erfolgt, ersichtlich in Tabelle 1.

Insgesamt wurden sechs Kategorien mit 29 Anwendungsbeispielen identifiziert. Aufgrund ihrer täglichen Relevanz in der Consultingbranche wurden davon die Use Cases „PowerPoint Präsentation erstellen (17)“ und „Deeskalierende E-Mail-Kommunikation (22)“ für eine detaillierte Umsetzungsanalyse ausgewählt.

Tabelle 1: Durch Experteninterviews identifizierte Use Cases

Nebenkategorie	Nr	Use Case
Programmierung	1	Codegeneration
Wissens-management	2	Wissensrecherche
	3	Marktanalyse
	4	Extrahierung von Informationen, Zusammenfassungen/ Executive Summaries
	5	Strukturierung von Inhalten
	6	Extrahierung von Informationen & Hinzufügen von Informationen in Exceltabellen (über API)
Schreibassistenz	7	Formulierungshilfe
	8	Paraphrasieren
	9	E-Mail (-Kommunikation)
	10	Dokumentenerstellung (z.B. Proposals)
	11	Websitetextgeneration, Storeeintraggeneration für Apps
	12	Übersetzung
	13	Kreative Texte Schreiben & Inspiration
	14	Reden schreiben
	15	Redetext für Pitches & Präsentationen schreiben
	16	Corporate Language (vorheriges Finetuning notwendig)
Außenwirkung	17	Powerpoints erstellen
	18	Kommunikations- Coaching/ Training mit ChatGPT
Beratende & prüfende Tätigkeiten	19	Risikoerkennung
	20	Prüfung von Dokumenten
	21	Bewertungssystem
	22	Beschwerde Management/ Sentiment Analyse in (E-Mail) -kommunikation bspw. Zur Deeskalation
Freizeit-gestaltung	23	Urlaub planen
	24	Rezepte
	25	Ernährungsplan
	26	Workoutplan
	27	Anlageberater (über API)

„PowerPoint Präsentation erstellen”

In Consultingfirmen ist die Erstellung von Präsentation täglicher Bestandteil der Arbeit. Bis zur fertigen Präsentation müssen mehrere Schritte durchlaufen werden, die bei manueller Durchführung die folgenden Punkte umfassen:

1. Wissensrecherche: Sammlung von passenden Informationen über das Internet, Studien, Bücher o.ä.
2. Strukturierung der Informationen: Zusammenstellung und Sortierung der Informationen für ein Präsentationsformat. Dazu gehört die Bestimmung einer Präsentationsgliederung, die Formulierung von Titeln und Texten und einer Storyline mit passendem Fazit.
3. Präsentationserstellung: Umsetzung der Struktur in PowerPoint durch manuelles Hinzufügen von Folien, Titeln, Texten, Bildern, Grafiken o.ä.
4. Überarbeitung und Verbesserung: Anpassung von Inhalten und Design, aufgrund von Abänderungswünschen oder Unternehmensrichtlinien.

Da jeder Schritt einen Zeitaufwand von mehreren Stunden bedeuten kann, würde die Möglichkeit, diese Arbeit zu automatisieren, eine erhebliche Zeiteinsparung ermöglichen. Mit dem Use Case soll überprüft werden, ob ChatGPT als Automatisierungstechnologie in Frage kommt. Eine Visualisierung des Automatisierungsprozesses zeigt Abb. 1.

Anforderungsanalyse

Tabelle 2: Anforderungsanalyse Use Case 1

Ziel	Der Nutzer erhält eine qualitativ hochwertige Präsentation mit einer faktisch korrekten inhaltlichen Übersicht zu Generative AI in Deutschland, die für Menschen ohne Vorkenntnisse zum Thema verständlich und nachvollziehbar ist.
Umfang	- Neu erstellte PowerPoint - 5 Folien (1 Deckblatt & 4 Inhaltsfolien) - Titel & Folieninhalte
Zustand bei Erfolg	Es existiert eine neu erstellte PowerPoint Präsentation, die mit einer logischen Struktur eine Übersicht über das Thema Generative AI in Deutschland bietet. Die Präsentation hat leicht verständliche Textinhalte und ein ansprechendes Design.
Zustand bei Scheitern	- Stufe 1: es existiert keine Präsentation - Stufe 2: die PPT hat keine logische Struktur - Stufe 3: textliche Inhalte sind faktisch inkorrekt oder nicht verständlich - Stufe 4: die PPT hat kein ansprechendes Design
Untersuchtes System	ChatGPT
Primär-Akteure	Nutzer, ChatGPT
Sekundär-Akteure	PowerPoint
Erfolgsszenario	Ersichtlich in Abbildung 1
Erweiterungen	Drei Varianten
Potentielle direkte Probleme	- ChatGPT ist nicht verfügbar, da Server ausgelastet sind o.ä. - ChatGPT braucht mehrere Prompts, um einen einzelnen Schritt zufriedenstellend auszuführen - ChatGPT führt einzelne Schritte unabh.ngig von der Promptanzahl nicht richtig aus - Makro ist fehlerhaft oder lässt sich nicht ausführen

Stakeholder

Tabelle 3: Stakeholder Use Case 1

Primärstakeholder	Kunden, Beratungsunternehmen, Berater/ Angestellter, Teammitglieder
Sekundärstakeholder	Endnutzer, Konkurrenz

User Story

Laura, Mitarbeiterin bei "TechPro Consulting", nutzt ChatGPT-Makros, um kurzfristig eine umfangreiche PowerPoint-Präsentation über Generative AI für ein wichtiges Kundengespräch zu erstellen. Nach Abklärung der inhaltlichen Punkte erstellt ChatGPT automatisch eine vorläufige Präsentation als Makro. Laura führt das Makro im Visual Basic Editor aus, passt die Präsentation an die Markenrichtlinien an und ergänzt diese mit ihrer Expertise. Die Nutzung von ChatGPT-Makros steigert ihre Effizienz und reduziert i.d.R. Stress, sodass sie zuversichtlich für das Kundengespräch ist und weitere Projekte erfolgreich bewältigen kann.

Rahmenbedingungen

Um die Einhaltung der Unternehmens-Policy zu gewährleisten, darf der Unternehmensname nicht als Kontextorientierung in den Prompts genannt werden. Dementsprechend ist eine gebrandete Präsentation im Corporate Design mit ChatGPT 3.5 nicht umsetzbar, sondern muss manuell für diesen Zweck angepasst werden.

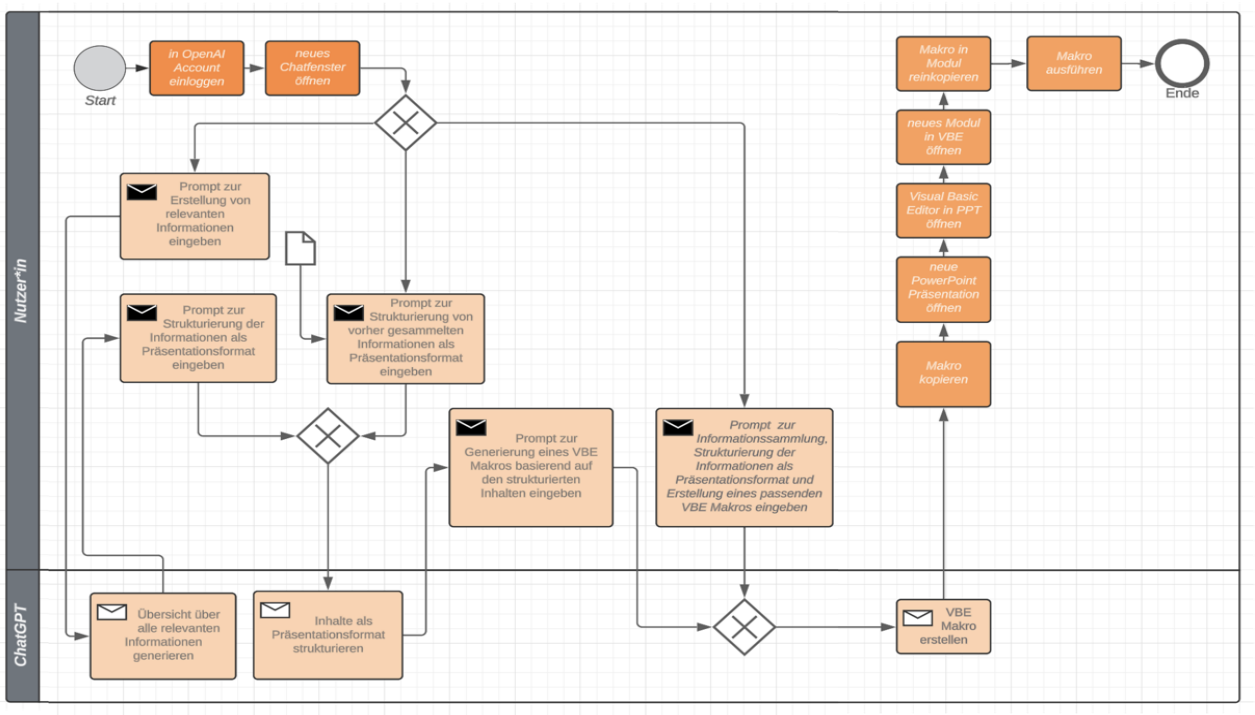

Abb. 1: Prozessablauf (Quelle: Eigene Darstellung)

Abb. 2: Teilergebnis Variante 1 zum Thema Generative KI (Quelle: Eigene Darstellung)

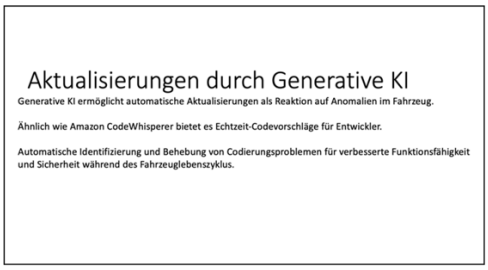

Abb. 3: Teilergebnis Variante 2 zum Thema Generative KI (Quelle: Eigene Darstellung)

Abb. 4:Teilergebnis Variante 3 zum Thema Generative KI in der Automobilbranche (Quelle: Eigene Darstellung)

Erkenntnisse

Haben Nutzer noch keine Erfahrung mit dem Einsatz von ChatGPT oder in der Erstellung von VBA-Makros, dauert die komplette Durchführung des Use Cases deutlich länger als bei erfahrenen Nutzern. Zudem kann Unwissen über die korrekte Formulierung von Prompts zu schlechteren Ergebnissen führen. Des Weiteren besteht die Möglichkeit, dass die Use Case Voraussetzungen noch nicht erfüllt sind und sich Nutzer beispielsweise über die Unternehmens-Policy informieren oder einen neuen Account erstellen müssen, was die Nutzung von ChatGPT aufwändiger als die manuelle Durchführung gestaltet. Grundlegende Kenntnisse sind daher erforderlich, damit der Use Case mit Hilfe von ChatGPT effizient ausgeführt wird. Zudem ist zu beachten, dass während der Durchführung der Use Cases mehrmals von ChatGPT generierte VBA-Makros bei Ausführung Fehler verursacht haben und so nicht funktioniert haben. Erst nach diversen Neugenerierungen konnten bestimmte Makros eingesetzt werden. Es hat sich jedoch als zeitsparender erwiesen, den Prompt neu zu stellen als nach einzelnen Fehlern zu fragen. Als technisch-unaffine Person können Codefehler nichtsdestotrotz verunsichernd wirken und zur Ablehnung von ChatGPT in diesem Einsatzszenario führen. Weiterhin ist zu erwähnen, dass ChatGPT 3.5 ein rein textbasiertes Tool ist und auch VBA-Makros nicht auf die Erstellung von visuellen Inhalten ausgelegt ist. Zwar besteht die Möglichkeit Bilder oder Grafiken per Makro einzufügen, dafür muss jedoch der passende Dateipfad zum Speicherungsort auf dem Desktop im Makro hinterlegt werden. Dies würde eine manuelle Speicherung von Bildern und Grafiken bedeuten und zusätzliche Zeit kosten. Die eingebaute Designerfunktion von PowerPoint, die sich als erweiterndes Tool zur visuellen Gestaltung nutzen lässt, kann ggf. im Unternehmenskontext aus Sicherheitsgründen gesperrt sein. Außerdem ist darauf zu achten, dass, aufgrund von dem begrenzten Kontextfelds ChatGPTs, sich keine großen Foliendecks kreieren lassen. Stattdessen bricht der VBA-Code frühzeitig ab und weist darauf hin die restlichen Folien in gleicher Weise manuell hinzuzufügen.

Neben den aus der einschlägigen Literatur abgeleiteten Risiken finden auch von Experten angesprochene Risiken Anwendung: Den Ausgaben von ChatGPT ist zu entnehmen, dass viele Formulierungen ähnlich klingen und persönliche Nuancen in Texten deutlich weniger werden. Hierdurch entsteht die Gefahr sich nicht von Konkurrenten, die ähnlich vorgehen, zu diffferenzieren. Dies kann zum einen Auswirkungen auf das Image des Beratungsunternehmens haben und zum anderen bei regelmäßiger Wiederholung zur negativen Beeinflussung von Soft Skills der Mitarbeiter führen, da sie verlernen Texte zu schreiben und strukturieren und potenziell dazu neigen, weniger achtsam in der Umsetzung von sprachlichen Aufgaben zu werden. Dies ist vor allem problematisch innerhalb der Wissensrecherche, da ChatGPT 3.5 kein Zugriff auf aktuelle Daten hat. Hinzukommend birgt die Wissensrecherche durch ChatGPT ein erhöhtes Risiko Falschinformationen zu verbreiten. Dies kann für Unternehmensberatungen projektspezifisch und / oder imagebezogene negative Auswirkungen zur Folge haben. Neben den Risiken durch ChatGPT existiert auch das Risiko mit Makros Malware-Infektionen oder unbefugten Zugriff auf sensible Daten zu ermöglichen, da diese Schadcode enthalten können, der beim Öffnen des Dokuments ausgeführt wird. Zurzeit ist dieses Risiko jedoch als gering einzustufen.

Trotz der Risiken sind die Use Case Anforderungen mehrheitlich erfüllt: So konnte eine PowerPoint mit logischer Struktur erstellt werden, die eine grundlegende Übersicht über generative KI mit Bezug zur Automobilbranche bietet (siehe hierzu Abb. 2 bis 4). Die Textinhalte sind überwiegend leicht verständlich und nachvollziehbar.

Als großes Defizit hat sich jedoch das Präsentationsdesign erwiesen. Es ist i.d.R. im Anschluss notwendig Folien manuell anzupassen, zu gestalten und visuelle Elemente hinzuzufügen. Folglich ist ChatGPT für die Vollautomatisierung des Prozesses ungeeignet, als unterstützendes Tool aber durchaus sinnvoll.

Unter Berücksichtigung der Grenzen und Risiken kann die Basis von Präsentation in kurzer Zeit erstellt und dafür mehr Aufwand in die Anpassung an Zielgruppen und ins Storytelling durch Design gesteckt werden. Auch

die von den Experten angegebene Maximalanzahl von sechs aufzuwendenden Prompts wird nicht überschritten. Für den Use Case existieren diverse Möglichkeiten, um an das gleiche Ziel zu kommen. Im vorliegenden Fall sind drei Varianten näher untersucht worden:

- Variante 1 (ein Prompt): In Fall von Zeitmangel liefert Variante 1 am schnellsten ein Ergebnis. Trotz kürzeren Inhalten ist die Wissensessenz vorhanden und kann als Grundlage dienen.
- Variante 2 (zwei Prompts): Möchte der Nutzer Kontrolle über die exakten Folieninhalte und Quellen erhalten, erscheint Variante 2 empfehlenswert. Da der Zeitaufwand für manuelle Recherche hoch ist, muss der Nutzer jedoch genug Zeit einplanen.
- Variante 3 (drei Prompts): Variante 3 ist für alle Nutzer interessant, die längere und genauere Texte erwarten, aber bei denen ChatGPT die Ausführung aller Schritte in Eigenleistung erbringen soll. Dazu ist es für Nutzer, die Einzelschritte nach ihren Vorstellungen personalisieren wollen, die beste Variante im direkten Vergleich.

In allen Fällen ist es essenziell, alle generierten Inhalte auf faktische Wahrheit zu überprüfen und toxische oder von Bias beeinflusste Texte zu entfernen.

„Deeskalierende E-Mail-Kommunikation"

Auch wenn Anwendungen wie Microsoft Teams oder Zoom sich zunehmend an Beliebtheit erfreuen, ist die E-Mail in der Consultingbranche nach wie vor das Hauptkommunikationstool mit Kunden und zwischen Mitarbeitern. Wenn es zu einem externen oder internen Konflikt kommt, erfolgt eine Beschwerde oftmals über dieses Medium. Dies erschwert häufig den Deeskalationsprozess im Vergleich zu persönlichen Gesprächen (Jong/ van Eekelen 1999). Eine manuelle Bearbeitung einer Beschwerdemail umfasst verschiedene Schritte:

1. Leseverständnis: Lesen und Verstehen der E-Mail
2. Stimmungseinschätzung: Inhaltliche Analyse und Abschätzung zur Stimmung des Absenders
3. Lösungsfindung: Entwicklung geeigneter Maßnahmen zur Konfliktlösung
4. Kommunikation: Formulierung einer deeskalierenden Antwort, inklusive Lösungsmöglichkeiten

Anforderungsanalyse

Tabelle 4: Anforderungsanalyse Use Case 2

Ziel	Der Nutzer erhält eine professionelle, empathische und deeskalierende E-Mail, die er als Antwort auf die Kunden-E-Mail einsetzen kann.
Umfang	- 1 Antwort-E-Mail - Sentiment Analyse der Kunden-E-Mail
Zustand bei Erfolg	Es existiert eine neu erstellte E-Mail, die einen logischen Aufbau hat und auf die Inhalte der Kunden-Email eingeht. Die generierte E-Mail muss die Bedenken vom Kunden konkret ansprechen und Lösungsansätze aufzeigen. Sie kann ohne große Abänderungen eingesetzt werden.
Zustand bei Scheitern	- Stufe 1: es existiert keine E-Mail - Stufe 2: die E-Mail hat keine logische Struktur - Stufe 3: die E-Mail kann nicht deeskalierend eingesetzt werden
Untersuchtes System	ChatGPT
Primär-Akteure	Nutzer, ChatGPT
Sekundär-Akteure	Kunde
Erfolgsszenario	Ersichtlich in Abbildung 5
Erweiterungen	-
Potentielle direkte Probleme	- ChatGPT ist nicht verfügbar, da Server ausgelastet sind o.ä. - ChatGPT braucht mehrere Prompts, um einen einzelnen Schritt zufriedenstellend auszuführen - ChatGPT führt einzelne Schritte unabhängig von der Promptanzahl nicht richtig aus

Die richtige Formulierung einer E-Mail in einer Konfliktsituation kann essenziell für die weitere Beziehung der Parteien und den Verlauf von Projekten sein. Eine Automatisierung des Prozesses anhand von ChatGPT (vgl. Abb. 2) vermag ggf. eine sachorientierte Perspektive auf die Situation fördern und eine optimierte Antwort, mit Berücksichtigung der Bedürfnisse der Involvierten, generieren.

Stakeholder

Tabelle 5: Stakeholder Use Case 2

Primärstakeholder	Kunden, Beratungsunternehmen, Berater/ Angestellter
Sekundärstakeholder	Andere Mitarbeiter, Andere Kunden, Konkurrenz

User Story

Als Consultant bei einer renommierten Beratungsfirma ist Lennart täglich mit der Kommunikation mit Kunden und Projektbeteiligten konfrontiert. Ein Kunde, Herr Müller, hat eine E-Mail gesendet, in der er emotional gereizt die Unzufriedenheit über das Nichteinhalten einiger Projektvorgaben ausdrückt. Lennart muss schnell und professionell reagieren, um die Situation zu deeskalieren und das Vertrauen des Kunden zu wahren. Er beschließt, das ChatGPT-System zu nutzen, um eine Sentiment-Analyse der E-Mail zu generieren und anschließend eine deeskalierende E-Mail als Antwort auf die Beschwerde von Herrn Müller zu erstellen.

Rahmenbegrenzungen

Um die Einhaltung der Unternehmens Policy zu gewährleisten, darf der Unternehmensname sowie der des betroffenen Kunde nicht als Kontextorientierung in den Prompts genannt werden. Weiterführend sollen keine projektinternen Informationen, sondern lediglich eine allgemeine Umschreibung der Gesamtsituation genutzt

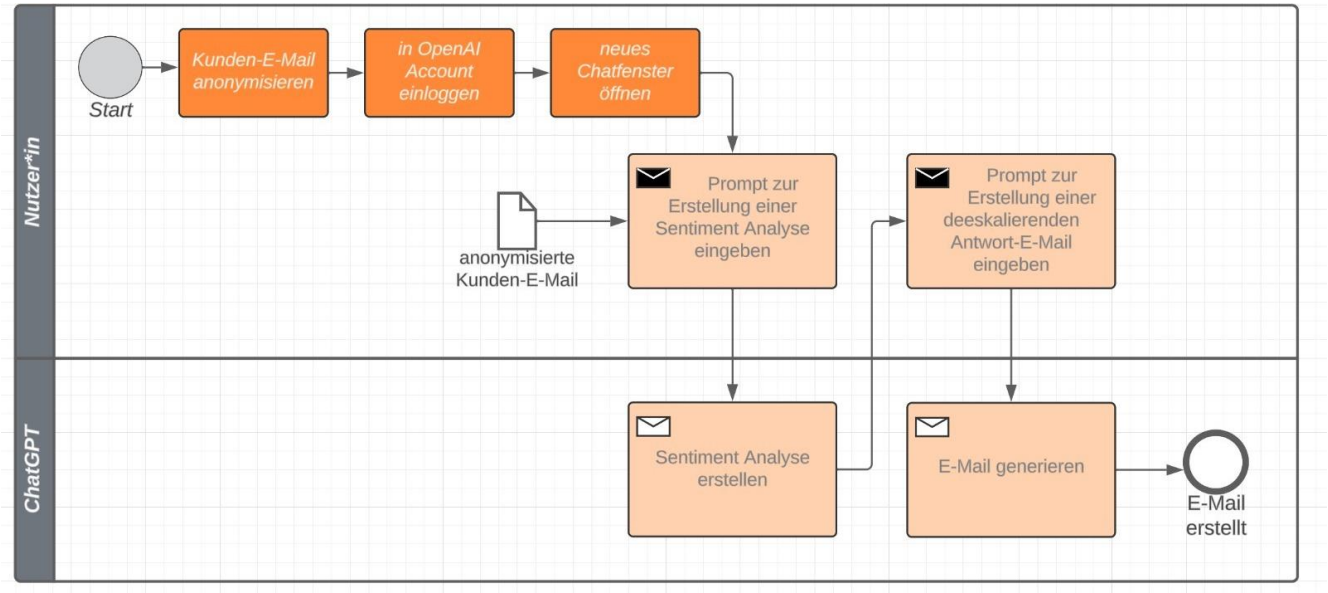

Abb. 5: Prozessablauf (Quelle: Eigene Darstellung)

"Betreff: Klärung und Stärkung unserer Geschäftsbeziehung

Sehr geehrte/r [Name des Absenders],

vielen Dank für Ihre offenen Worte und Ihr ehrliches Feedback in Ihrer E-Mail. Zunächst möchte ich betonen, dass es uns außerordentlich wichtig ist, eine vertrauensvolle und effiziente Zusammenarbeit mit Ihnen zu pflegen, und es tut uns aufrichtig leid, dass Sie mit einigen Aspekten unserer aktuellen Zusammenarbeit unzufrieden sind.

Ihr Anliegen nehmen wir äußerst ernst, und wir möchten alles daran setzen, die Situation zu klären und gemeinsam eine Lösung zu finden. Es liegt uns am Herzen, Ihre Erwartungen zu erfüllen und unsere Geschäftsbeziehung wieder auf den richtigen Weg zu bringen.

Wir verstehen vollkommen Ihre Frustration und Enttäuschung angesichts der nicht eingehaltenen Vorgaben. Bitte seien Sie versichert, dass wir die Gründe dafür genau untersuchen werden, um eine Wiederholung künftig zu vermeiden. Ihr Vertrauen in unsere Dienstleistungen ist uns äußerst wichtig, und wir werden alles daran setzen, dieses Vertrauen wieder aufzubauen.

Unser Team wird die Angelegenheit sorgfältig prüfen und die notwendigen Schritte ergreifen, um sicherzustellen, dass die Vereinbarungen eingehalten werden und unsere Zusammenarbeit reibungslos verläuft. Wir sind dankbar für Ihre Bereitschaft zur Zusammenarbeit, um gemeinsam eine Lösung zu finden.

Wir möchten Sie darüber informieren, dass wir uns zeitnah mit Ihnen in Verbindung setzen werden, um Ihnen unsere konkreten Pläne zur Verbesserung unserer Zusammenarbeit darzulegen. Ihre Zufriedenheit ist unser oberstes Ziel, und wir möchten sicherstellen, dass Sie sich auf unseren professionellen Service verlassen können.

Nochmals möchten wir uns aufrichtig bei Ihnen für die aufgetretenen Unannehmlichkeiten entschuldigen und versichern Ihnen, dass wir uns mit der gebotenen Dringlichkeit um eine positive Entwicklung unserer Kundenbeziehung kümmern werden.

Sollten Sie weitere Fragen oder Anliegen haben, zögern Sie bitte nicht, uns direkt zu kontaktieren. Wir schätzen unsere Zusammenarbeit mit Ihnen und sind zuversichtlich, dass wir gemeinsam zu einer befriedigenden Lösung gelangen werden.

Mit freundlichen Grüßen,

[Name des Unternehmens]"

Abb. 6: Ergebnis deeskalierende E-Mail

Erkenntnisse

Zusätzlich zu den Einschränkungen aus dem vorherigen Use Case besteht hierbei die Möglichkeit, dass Nutzer den Vorgang einer Sentiment Analyse nicht kennen und so möglicherweise Schwierigkeiten haben, ihr Anliegen in Worte zu fassen. Des Weiteren ist zu erwähnen, dass Konflikte meist nicht plötzlich auftreten, sondern i.d.R. eine Entstehungsgeschichte haben. Es ist folglich fragwürdig, ob eine einzige E-Mail als Orientierung den Konflikt vollumfänglich abbilden kann, vor allem wenn Teile der Kommunikation auch in persönlichem, mündlichem Kontakt stattgefunden haben. Ferner ist es möglich, dass ChatGPT aufgrund von Datenschutzrichtlinien keine umfangreiche Analyse durchführen kann. Da bestimmte Konfliktbereiche eine eigene Interpretation erfordern, ist dies für ChatGPT folglich nicht möglich. Zusätzlich ist die Sentiment Analyse auf das Spektrum negativ bis positiv begrenzt. Trotz einer umfangreichen Anzahl von nuancierten Emotionen sind diese nicht angegeben. Die emotionalen Hintergründe des Absenders sind entsprechend vereinfacht dargestellt und bilden eventuell nicht die Wirklichkeit ab.

Neben den aus wissenschaftlichen Erkenntnissen abgeleiteten Risiken sehen die befragten Experten weitere Risiken: Die Fähigkeit von ChatGPT, Gefühle aus Sätzen zu erkennen, basiert auf mathematischen Algorithmen und nicht auf einem wirklichen Verständnis von Emotionen. Entsprechend ist davon auszugehen, dass die Sentiment Analyse einer Maschine eine andere ist als die eines Menschen. Die Frage, ob ChatGPT die relevanten Emotionsnuancen verstehen kann, ist daher unklar. Des Weiteren ist vor allem in einem interkulturellen Kontext die Beratung von ChatGPT kritisch zu sehen, da der Datensatz und die Labeler, die ChatGPT trainieren, hauptsächlich aus der westlichen Welt stammt und Stimmungen in verschiedenen Kulturen sprachlich unterschiedlich ausgedrückt werden (Dwivedi u. a. 2023). Dementsprechend kann es zu Missinterpretationen kommen. Ferner entstehen durch die Weiterentwicklung von Sprache neue Wörter und Wortbedeutungen, die, aufgrund von ChatGPTs veralteten Datenstand, sich nicht mehr angemessen interpretieren lassen. Entsprechend kann der Einsatz von ChatGPT im Unternehmenskontext einen negativen Einfluss auf das Image des Unternehmens und dessen Arbeit haben.

Die Erprobung des Use Cases zeigt dennoch ein positives Ergebnis, da die für das Erfolgsszenario erforderlichen Kriterien erfüllt werden konnten (vgl. Abb. 6). So ist eine

professionell formulierte E-Mail generiert worden, die auf die Bedenken des Kunden eingeht und eine Lösungsfindung anbietet, ohne die maximale Anzahl an Prompts von sechs zu überschreiten. Vor dem Versenden müsste in diesem Einsatzszenario die E-Mail nur noch geringfügug individualisiert werden. Daraus folgt, dass die Nutzung von ChatGPT zur Deeskalation in der E-Mail-Kommunikation für Unternehmenberatungen ein sinnvolles Hilfstool sein kann, wenn die geltenden Vorschriften und Rahmenbedingugnen dabei berücksichtigt werden. Nichtsdestotrotz ist zwischen verschiedenen Konfliktsituationen zu differenzieren. Besteht die Möglichkeit, die Situation in einem persönlichen Gespräch aufzuklären, sollte man diese Maßnahme wählen, um Missverständnisse leichter aus der Welt zu schaffen und besser auf individuelle Bedürfnisse der Konfliktparteien eingehen zu können. Ist dies nicht umsetzbar, kann ChatGPT, vor allem bei schriftlichen Konversationen eine Mediatorrolle einnehmen. Gerade aufgrund der Tatsache, dass Menschen oftmals mit negativen Emotionen auf Infragestellungen der eigenen Leistung reagieren, können Consultingfirmen von ChatGPT profitieren und dieses Risiko einschränken. Dennoch erweist es sich als empfehlenswert, dieses Konzept hauptsächlich bei einfachen Konflikten einzusetzen, die keine längere Konfliktvergangenheit haben. Abschließend ist es unerlässlich, die Ausgaben nochmals zu überprüfen und für die individuellen Konversationspartner anzupassen.

DISKUSSION

Generelle Anwendungsempfehlung

Eine große Stärke von ChatGPT liegt in Analyse, Strukturierung und Schreiben von Texten in menschlicher Sprache. Folglich ist vor allem die Eingabe von eigenen Referenztexten, auf dessen Basis ChatGPT Befehlen nachgeht, eine geeignete Lösung, um das Risiko, Falschinformationen zu erhalten, zu reduzieren. Eine weitere Kompetenz des LLMs ist die Erstellung von Inhalten nach Vorgaben und Referenzeingaben des Nutzers, die als erster Entwurf anschließend weiterverarbeitet und angepasst werden kann. Dabei vermag es sich um Newsletter, Blogs, E-Mails, Reden, Präsentationsinhalte oder ähnliches halten, die mit Hilfe von ChatGPT weiterführend an eine vom Nutzer definierte Zielgruppe angepasst oder in eine oder mehrere, vom Tool beherrschte Sprache übersetzt werden. Weiterhin besteht die Möglichkeit, Referenztexte mit bestimmten Kriterien abzugleichen, die im Vorfeld definiert wurden, um den Text zu kategorisieren. Eine weitere Stärke von ChatGPT besteht in der Möglichkeit zur Einübung von verschiedenen rollen- bzw. Persona-basierten Dialogen. Im direkten Kundenkontakt ist jedoch tendenziell vom Einsatz abzuraten, da im Kundensupport i.d.R. nicht genügend Kontrollmechanismen vorhanden sind. Entsprechend sollte ChatGPT als unternehmensinternes Tool angesehen werden. Unabhängig vom jeweiligen Anwendungsfall sind die Ausgaben jedoch stets noch von Menschen zu überprüfen, bevor man sie final verwendet.

Limitationen

In der vorliegenden Arbeit setzen sich die Limitationen aus drei beeinflussenden Faktoren zusammen: Erstens beschränkt die Wahl von Experteninterviews als Teilmethodik den Umfang der Daten, sodass die Ergebnisse nicht auf die Allgemeinheit übertragbar sind. Zudem ist die geringe Zahl der Befragten nicht repräsentativ für die Gesamtheit der Beratungsindustrie, zumal die Experten auch angegeben haben, sich privat mit neuen Technologien zu beschäftigen. Dementsprechend ist zu vermuten, dass bei ihnen das Interesse, LLMs in ihren Arbeitsalltag zu integrieren, höher als im Durchschnitt ist. Die Untersuchung der Use Cases wird des Weiteren durch das Nutzungswissen der Forschenden eingeschränkt, sodass folglich nur so viel Wissen und Fähigkeiten von ChatGPT nutzbar waren, wie die Forschenden durch die eigenen Promptingkompetenzen extrahieren können. Weiterhin gibt es, aufgrund der Aktualität des Themas, permanente Weiterentwicklungen und Neuveröffentlichungen. So lagen zum Zeitpunkt der Untersuchung noch keine Informationen über den Einsatz von GPT-4 vor. Es ist davon auszugehen, dass die GPT 3.5 Version für Unternehmen zunehmend unattraktiv wird.

SCHLUSSFOLGERUNG UND AUSBLICK

Diese Arbeit untersucht das Potenzial von LLMs am Beispiel von ChatGPT 3.5 für die Automatisierung von Büro- und Verwaltungsarbeiten innerhalb von Consultingunternehmen. Anhand von Experteninterviews sind verschiedene Anwendungsfälle identifiziert worden, in denen man ChatGPT sinnvoll einsetzen kann, um Arbeitszeit und Ressourcen einzusparen. Als Voraussetzung sind passende Anforderungskriterien an das Tool zu definieren.

Die Ergebnisse zeigen, dass die Integration von ChatGPT in die Unternehmensumgebung vielversprechende Möglichkeiten bietet, bestimmte Aufgaben (teilweise) zu automatisieren und Prozesse zu optimieren. Die praxisorientierten Use Cases "Powerpoint Präsentation erstellen" und „Erstellung einer E-Mail zur Deeskalation in der Kundenkommunikation“ verdeutlichen, wie man ChatGPT in der Realität anwenden kann, um die Effizienz und Genauigkeit von Bürotätigkeiten zu verbessern. Aus den Interviews lässt sich eine allgemeine Bereitschaft zur Nutzung von ChatGPT im Arbeitskontext feststellen. Allerdings weisen die Ergebnisse der Befragung auch auf Risikopotenziale hin, insbesondere in Bezug auf Datensicherheit, Ethik und Verantwortlichkeit.

Die Integration von LLMs muss zudem von einer passenden KI-Strategie des Unternehmens begleitet und umfängliche Aufklärungsarbeit geleistet werden, um die Wahrscheinlichkeit des Eintretens von Risiken zu minimieren und die Mitarbeiterkompetenzen an den zeitgemäßen Standard anzupassen. Die Arbeit hat gezeigt, dass der aktuelle Forschungsstand im Bereich der Automatisierung und LLMs bereits zahlreiche Einsatzmöglichkeiten in verschiedenen Branchen aufzeigt, aber speziell für den Einsatz von LLMs in Büro- und Verwaltungsarbeiten noch weiterführende Forschung benötigt wird.

In den kommenden Jahren ist mit einem deutlichen Anstieg der Größe und Leistungsfähigkeit von LLMs sowie mit der Entwicklung anwendungsspezifischer LLMs zu rechnen. Das bedeutet, dass es auf der einen Seite die Basis-LLMs geben wird, die eine generelle Sprachkompetenz besitzen und eine Vielzahl an Textaufgaben befriedigend erledigen können und auf der anderen Seite werden LLMs zukünftig auf konkrete Umgebungen und Bedürfnisse, wie Unternehmen oder Codingplattformen, angepasst werden.

LITERATUR

Aiyappa, Rachith u. a. (2023): Can we trust the evaluation on ChatGPT?, arXiv preprint arXiv:2303.12767v1.

Bender, Emily M. u. a. (2021): On the Dangers of Stochastic Parrots, in: Association for Computing Machinery (Hrsg.): FAccT'21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, New York, S. 610–623.

Bozkurt, Aras u. a. (2023): Speculative Futures on ChatGPT and Generative Artificial Intelligence (AI). A Collective Reflection from the Educational Landscape, in: *Asian Journal of Distance Education*, 18. Jg., S. 53–130.

Brown, Tom B. u. a. (2020): Language Models are Few-Shot Learners, in: Neural Information Processing Systems Foundation, Inc. (NeurIPS) (Hrsg.): Advances in Neural Information Processing Systems 33, Red Hook, NY, S. 1877–1901.

Chohan, Usman W. (2023): Generative AI, ChatGPT & The Future of Jobs, Notes on the 21 st Century, Critical Blockchain Initiative.

Chui, Michael/Manyika, James/Miremadi, Mehdi (2016): Where machines could replace humans - and where they can't (yet). The technical potential for automation differs dramatically across sectors and activities, in: McKinsey Quarterly, S. 1–12.

Chui, Michael/Roberts, Roger/Yee, Lareina (2022): Generative AI is here. How tools like ChatGPT could change your business, Quantum Black AI by McKinsey.

Ferrara, Emilio (2023): Should ChatGPT be Biased? Challenges and Risks of Bias in Large Language Models, arXiv preprint arXiv:2304.03738.

Ganguli, Deep u. a. (2023): The Capacity for Moral Self-Correction in Large Language Models, arXiv preprint arXiv:2302.07459v2.

Gao, Leo/Schulman, John/Hilton, Jacob (2023): Scaling Laws for Reward Model Overoptimization, in: Krause, A. u. a. (Hrsg.): Proceedings of the 40th International Conference on Machine Learning, PMLR 202, S. 265–279. 131

Ghavami, Peter (2020): Big Data Analytics Methods. Analytics techniques in data mining, deep learning and natural language processing, De Gruyter, Boston/ Berlin.

Goetze, Trystan S./Abramson, Darren (2021): Bigger Isn't Better. The Ethical and Scientific Vices of Extra-Large Datasets in Language Models, in: Association for Computing Machinery (Hrsg.): WebSci '21. Proceedings of the 13th ACM Web Science Conference 2021, S. 69–75.

Goldstein, Josh A. u. a. (2023): Generative Language Models and Automated Influence Operations. Emerging Threats and Potential Mitigations, arXiv preprint arXiv:2301.04246.

Dwivedi, Yogesh K. u. a. (2023): So what if ChatGPT wrote it?. Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy, in: *International Journal of Information Management*, 71. Jg., S. 102642.

Haleem, Abid/Javaid, Mohd/Singh, Ravi Pratap (2022): An era of ChatGPT as a significant futuristic support tool. A study on features, abilities, and challenges, in: *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, 2. Jg., Nr. 4, S. 1–8.

Jong, Jan A. de/van Eekelen, Ilse M. (1999): Management consultants: what do they do?, in: *Leadership & Organization Development Journal*, 20. Jg., Nr. 4, S. 181–188.

Kasneci, Enkelejda u. a. (2023): ChatGPT for good? On opportunities and challenges of large language models for education, in: *Learning and Individual Differences*, 103. Jg., Nr. 102274, S. 1–9.

Kuiler, Erik W. (2022): Ethics, in: Schintler, Laurie A./ McNeely, Connie L. (Hrsg.): Encyclopedia of Big Data, Springer International Publishing, Cham, S. 452–454.

Lund, Brady D. u. a. (2023): ChatGPT and a new academic reality: Artificial Intelligence-written research papers and the ethics of the large language models in scholarly publishing, in: *Journal of the Association for Information Science and Technology*, 74. Jg., Nr. 5, S. 570–581.

Moloi, Tankiso/Marwala, Tshilidzi (2021): Artificial Intelligence and the Changing Nature of Corporations. How Technologies Shape Strategy and Operations, Springer International Publishing, Cham.

Narasimhan, Karthik u. a. (2018): Improving Language Understanding by Generative Pre-Training, OpenAI.134 Anwendungsfelder und Herausforderungen der Künstlichen Intelligenz. Wie KI die Gesellschaft und Marketing beeinflusst, Springer Fachmedien, Wiesbaden.

Onslow, Mike (2023): Unraveling the Enigma of Tokens in ChatGPT. Your Pathway to Proficiency, URL: https://medium.com/@mike_onslow/unraveling-the-enigma-of-tokens-in-chatgpt-your-pathway-to-proficiency-4ee200e18cc3, Stand: 23. August 2023.

OpenAI (2023a): ChatGPT plugins, URL: https://openai.com/blog/chatgptplugins, Stand: 19. August 2023.

OpenAI (2023b): GPT Best Practices, URL: https://platform.openai.com/docs/guides/gpt-best-practices, Stand: 20. August 2023.

Ouyang, Long u. a. (2022): Training language models to follow instructions with human feedback, arXiv preprint arXiv:2203.02155.

Patsakis, Constantinos/Lykousas, Nikolaos (2023): Man vs the machine. The Struggle for Effective Text Anonymisation in the Age of Large Language Models, arXiv preprint arXiv:2303.12429v1.

Polonsky, Michael Jay/Rotman, Jeffrey D. (2023): Should Artificial Intelligent Agents be Your Co-author? Arguments in Favour, Informed by ChatGPT, in: *Australasian Marketing Journal*, 31. Jg., Nr. 2, S. 91–96.

Schulman, John u. a. (2017): Proximal Policy Optimization Algorithms, arXiv preprint arXiv:1707.06347, S. 1–12.

Taulli, Tom (2022): Grundlagen der Künstlichen Intelligenz. Eine nichttechnische Einführung, Springer, Berlin/Heidelberg.

Wang, Fei-Yue u. a. (2023): What Does ChatGPT Say. The DAO from Algorithmic Intelligence to Linguistic Intelligence, in: *IEEE/CAA Journal of Automatica Sinica*, 10. Jg., Nr. 3, S. 575–579.

Yin, Shukang u. a. (2023): A Survey on Multimodal Large Language Models, arXiv preprint arXiv:2306.13549v1.

ZBW - Leibniz Information Centre for Economics (2023): Economics of ChatGPT. A Labor Market View on the Occupational Impact of Artificial Intelligence, Kiel.

Zhang, Wenxuan u. a. (2023): Sentiment Analysis in the Era of Large Language Models. A Reality Check, arXiv preprint arXiv:2305.15005v1.

Zhao, Wayne Xin u. a. (2023): A Survey of Large Language Models, arXiv preprint arXiv:2303.18223.

Ziegler, Daniel M. u. a. (2019): Fine-Tuning Language Models from Human Preferences, arXiv preprint arXiv:1909.08593v2.