

**KI-gestützte E-Mail-Automation im Garten-
und Landschaftsbau: Entwicklung und
Evaluation eines Prototyps**

Bachelorarbeit

Prototypische Entwicklung und formative Evaluation

Bachelorstudium

“Management, Communication & IT”

MCI | DIE UNTERNEHMERISCHE HOCHSCHULE®

Betreuer:

Prof. Dr. Reinhard Bernsteiner

Verfasser:

Julius Lindenlaub

WS2023

52315201

07.06.2026

Abstract

Deutsch

Betriebliche Büros im Garten- und Landschaftsbau (GaLaBau) empfangen täglich ein inhaltlich vielfältiges E-Mail-Aufkommen, das vor jeder inhaltlichen Bearbeitung manuell gesichtet, eingeordnet und weitergeleitet werden muss. Der damit verbundene Koordinationsaufwand betrifft zwar grundsätzlich viele Betriebe, ist in Organisationen ohne dedizierte IT-Abteilung oder spezialisiertes Büromanagement jedoch besonders schwer strukturell abzufangen. Die vorliegende Arbeit untersucht, wie sich eine KI-gestützte E-Mail-Automation in einem mittelständischen GaLaBau-Betrieb prototypisch implementieren lässt und welche technischen, prozessualen und organisatorischen Voraussetzungen ihre Praxistauglichkeit bestimmen.

Als Forschungsrahmen dient eine gestaltungsorientierte Einzelfallstudie bei einem mittelständischen GaLaBau-Betrieb. In iterativer Entwicklung (P1 bis P4) entstand ein Prototyp, der eingehende E-Mails über einen n8n-Workflow abrufen, mit dem lokal betriebenen Open-Source-Sprachmodell llama3.1:8b klassifiziert und mit Outlook-Kategorie-Labels versieht. Die datenschutzorientierte Architektur verzichtet vollständig auf externe KI-Dienste. Zur Evaluation wurden eine explorative Vorerhebung (n=6 GaLaBau-Betriebe), ein Mail-Tagebuch über vier Arbeitstage, eine Labelauswertung von 57 E-Mails sowie eine qualitative Nachbesprechung mit dem Partnerunternehmen eingesetzt.

Die Ergebnisse belegen die grundsätzliche technische Machbarkeit. 25 von 57 E-Mails (43,9 Prozent) wurden korrekt klassifiziert. Eine produktive Vollautomation wurde nicht aktiviert, da die Evaluation mehrere Implementierungsgrenzen sichtbar machte. Dazu zählten fehlende Prozessklarheit, ein unzureichend abgegrenztes Kategorienschema sowie Lücken in der Stammdatenpflege des angeschlossenen Fachsystems. Der reguläre manuelle Bearbeitungsaufwand von etwa 15 Minuten täglich ließ einen belastbaren wirtschaftlichen Nutznachweis nicht zu.

Aus der Evaluation wurden acht Implementierungskriterien für praxistaugliche KI-Automation im Handwerk abgeleitet. Die zentrale Schlussfolgerung lautet: KI-Projekte sollten nicht mit der Modellauswahl, sondern mit der Klärung der zu automatisierenden Entscheidung beginnen. Prozessmapping, Kategorienvvalidierung und Stammdatenpflege sind grundlegende Voraussetzungen jeder erfolgreichen technischen Implementierung.

English

Business offices in the landscape and horticulture sector receive heterogeneous incoming emails that must be manually reviewed, categorized, and forwarded before substantive processing can begin. This coordination effort affects many organizations, but it is especially difficult to absorb structurally in companies without dedicated IT departments or specialized administrative staff. This study investigates how an AI-based email automation system can be prototypically implemented in a medium-sized landscaping company and which technical, process-related, and organizational prerequisites determine its practical viability.

The research follows a design-oriented single-case study with an anonymized medium-sized GaLaBau company. Through iterative development (P1 to P4), a prototype was built that retrieves emails via an n8n workflow, classifies them with the locally deployed open-source language model llama3.1:8b, and assigns Outlook category labels. The privacy-oriented architecture avoids external AI services. Evaluation instruments included an exploratory pre-survey (n=6), a four-day email diary, a label review of 57 emails, and a qualitative follow-up discussion.

The results confirm basic technical feasibility. 25 out of 57 emails (43.9 percent) were correctly classified. Full automation was not activated because the evaluation revealed several implementation limits, including insufficient process clarity, an inadequately distinctive category schema, and gaps in master data maintenance. A regular manual processing effort of approximately 15 minutes per day did not allow for a substantiated economic benefit assessment.

Eight implementation criteria for practically viable AI automation in the skilled trades are derived from the evaluation. The central conclusion is that AI projects should begin not with model selection, but with clarifying the decision to be automated. Process mapping, category validation, and master data maintenance are fundamental prerequisites for successful technical implementation

Inhaltsverzeichnis

Abbildungsverzeichnis.....	III
Tabellenverzeichnis	IV
Abkürzungsverzeichnis.....	V
1 Einleitung.....	1
1.1 Ausgangssituation und Problemdarstellung.....	1
1.2 Zielsetzung und Forschungsfrage	2
1.3 Aufbau der Arbeit	3
2 Theoretischer Rahmen	5
2.1 Digitalisierung im Handwerk.....	5
2.1.1 Begriff und Transformationsebenen	5
2.1.2 Digitalisierungsstand im Handwerk und GaLaBau	6
2.1.3 Hemmnisse, Kompetenzen und Designimplikationen.....	7
2.2 Künstliche Intelligenz	10
2.2.1 Von maschinellem Lernen zu Large Language Models	11
2.2.2 Betriebliche Anwendung und Grenzen.....	13
2.3 Prozessautomatisierung im E-Mail-Management.....	16
2.3.1 E-Mail als Geschäftsprozess	16
2.3.2 Robotic Process Automation und KI-gestützte Klassifizierung	18
2.4 Implementierungsbedingungen im KMU-Kontext.....	20
2.4.1 Prozess- und Datentauglichkeit	20
2.4.2 Pilotierung, Kontrolle und Wirtschaftlichkeit	22
3 Methodik.....	24
3.1 Forschungsdesign und methodischer Rahmen.....	24
3.2 Erhebungsinstrumente	26
3.2.1 Vorerhebung: Fragebogen	26

3.2.2 Prozessmapping beim Partnerunternehmen	26
3.2.3 Mail-Tagebuch	28
3.2.4 Labelauswertung und Nachbesprechung	28
3.3 Evaluationskonzept und Limitationen	29
4 Prototypentwicklung	32
4.1 Entwicklungskontext	32
4.1.1 Ausgangssituation und Vorerhebung	32
4.1.2 Iterative Prototypentwicklung und Prozessmapping	33
4.2 Finaler Prototyp (P4)	34
4.2.1 Technische Architektur und Datenfluss	35
4.2.2 Kategorisierungslogik und Einführungsstrategie	37
4.2.3 Geocapture-Anbindung und Routing	38
4.2.4 Datenschutz und Dokumentation	39
5 Evaluation	40
5.1 Datengrundlage der Evaluation	40
5.2 Ergebnisse der Labelphase	41
5.3 Prozessuale und datenbezogene Grenzen	43
5.4 Wirtschaftliche Praxistauglichkeit	44
5.5 Kriterienrahmen	45
6 Fazit	47
6.1 Beantwortung der Forschungsfrage	47
6.2 Limitationen	49
6.3 Ausblick und weiterer Forschungsbedarf	50
Literaturverzeichnis	51
Eidesstattliche Erklärung	59
Transparenzerklärung	60
Anhang	i

Abbildungsverzeichnis

Abbildung 1: Design der Forschungsmethode.....	24
Abbildung 2: Veränderung der Prozessannahme durch das Prozessmapping.....	27
Abbildung 3: Vereinfachte Darstellung des Prototyps	36
Abbildung 4: Ergebnis der Labelauswertung	42

Tabellenverzeichnis

Tabelle 1: Bewertungsdimensionen und Aussagegrenzen der Evaluation	30
Tabelle 2: Prototypentwicklung	34
Tabelle 3: Technische Komponenten des Prototyps	35
Tabelle 4: Mail-Tagebuch Auswertung	41
Tabelle 5: Implementierungskriterien für KI-gestützte E-Mail-Automation	45

Abkürzungsverzeichnis

Abkürzung	Bedeutung
API	Application Programming Interface (Engl.)
ART	Action Regulation Theory (Engl.)
DSGVO	Datenschutz-Grundverordnung (Dt.)
GaLaBau	Garten- und Landschaftsbau (Dt.)
HwK	Handwerkskammer (Dt.)
IPO	Input-Processing-Output (Engl.)
JSON	JavaScript Object Notation (Engl.)
LLM	Large Language Model (Engl.)
ML	Maschinelles Lernen (Dt.)
NLP	Natural Language Processing (Engl.)
OECD	Organisation for Economic Co-operation and Development (Engl.)
RPA	Robotic Process Automation (Engl.)
VPS	Virtueller privater Server (Dt.)
XAI	Explainable Artificial Intelligence (Engl.)

1 Einleitung

Die Einleitung führt in die Problemstellung der Arbeit ein, leitet daraus die Forschungsfrage ab und beschreibt den Aufbau der Untersuchung.

1.1 Ausgangssituation und Problemdarstellung

Betriebliche E-Mail-Bearbeitung bildet im GaLaBau-Kontext einen vorgelagerten Koordinationsprozess, bevor fachliche Arbeit überhaupt beginnen kann. Angebotsanfragen, Lieferbestätigungen, Korrespondenz mit Subunternehmen, Reklamationen und Bewerbungen treffen häufig im gleichen Postfach ein und müssen manuell gesichtet, eingeordnet und an zuständige Personen weitergeleitet werden. Im Austausch mit dem Digitalisierungsausschuss des Fachverbands Garten-, Landschafts- und Sportplatzbau Hessen-Thüringen zeigte sich, dass digitale Lösungen für solche administrativen Abläufe grundsätzlich auf Interesse stoßen, der konkrete Einstieg für viele Betriebe jedoch unklar bleibt. Die Problemstellung dieser Arbeit setzt daher nicht bei einer abstrakten Automatisierungsidee an, sondern bei einem wiederkehrenden Sortier- und Weiterleitungsprozess, der im untersuchten Branchenkontext praktisch beobachtbar ist.

E-Mail-Bearbeitung ist in wissensintensiven Tätigkeiten ein gut belegter Belastungsfaktor. Studien zu Wissensarbeit, also zu Tätigkeiten mit hohem Anteil an Informationsverarbeitung, Kommunikation und Entscheidungsvorbereitung, zeigen einen erheblichen täglichen Zeitaufwand für das Lesen, Sortieren und Beantworten eingehender Nachrichten (Mark et al., 2016; Martin et al., 2025). Mark et al. (2016) verbinden intensive E-Mail-Nutzung zudem mit geringer wahrgenommener Produktivität und höherem arbeitsbezogenem Stress. Lancot und Duxbury (2025) zeigen ergänzend, dass der Erwartungsdruck zeitnahe Reaktion zunehmend auch außerhalb regulärer Arbeitszeiten wirkt. Für GaLaBau-Betriebe sind diese Ergebnisse nur eingeschränkt übertragbar, da die Untersuchungen nicht im Handwerkskontext durchgeführt wurden. Die Befunde liefern jedoch Hinweise darauf, dass auch hier ein ungefiltertes zentrales Postfach zu zusätzlicher Belastung und ineffizienten Arbeitsabläufen beitragen kann.

Im Handwerk verschärft sich die E-Mail-Problematik durch begrenzte organisatorische und digitale Ressourcen. Administrative Automatisierung wird zwar als relevantes Zukunftsfeld beschrieben, ist in der betrieblichen Realität vieler Handwerksbetriebe jedoch noch nicht umfassend angekommen (Beuchel et al., 2024). Gerade kleine und

mittlere Handwerksbetriebe verfügen häufig nicht über spezialisierte IT- oder Prozessmanagementstrukturen, sodass Sortier-, Rückfrage- und Weiterleitungsaufgaben auf wenige Personen im Büroalltag konzentriert sind. Damit entsteht ein Spannungsfeld aus digitalem Potenzial, knappen Kapazitäten und unklarer Umsetzbarkeit, das für den GaLaBau-Kontext besonders relevant ist.

KI-gestützte E-Mail-Klassifikation bietet einen technischen Ansatzpunkt, löst die Implementierungsfrage im Handwerk jedoch nicht allein. KI-Systeme zur Verarbeitung natürlicher Sprache können eingehende E-Mails anhand ihres Inhalts klassifizieren und an zuständige Stellen weiterleiten, wie Sundberg et al. (2024) in einer Fallstudie zu öffentlichem E-Mail-Management zeigen. Für GaLaBau-Betriebe bleibt jedoch offen, unter welchen technischen, prozessualen und organisatorischen Bedingungen eine solche Lösung im realen Betriebsalltag praxistauglich wird. Die vorliegende Arbeit untersucht deshalb nicht nur, ob eine Klassifikation technisch möglich ist, sondern welche Voraussetzungen erfüllt sein müssen, damit eine KI-gestützte E-Mail-Automation betrieblich anschlussfähig wird.

Eine explorative Vorerhebung unter sechs GaLaBau-Betrieben aus dem Umfeld des Digitalisierungsausschusses des Fachverbands Hessen-Thüringen (März 2026, n=6) konkretisiert diese Ausgangslage. Datenschutzbedenken, unklarer Nutznachweis und wahrgenommener Einführungsaufwand wurden am häufigsten als Ablehnungsgründe genannt. Vier von sechs Betrieben signalisierten jedoch, eine automatisierte Lösung einzusetzen, wenn eine klare, messbare Zeitersparnis nachgewiesen werden kann. Dieser Befund ist zugleich Problembestätigung und Gestaltungsauftrag. KI-gestützte E-Mail-Automation im Handwerk muss als Implementierungsaufgabe verstanden werden, die Anforderungen an Prozesse, Datenqualität, Datenschutz und betriebliche Verantwortlichkeiten verbindet. Die Ergebnisse der Vorerhebung werden in Kapitel 4.1.1 ausgeführt, wo sie als empirische Grundlage für die Designentscheidungen des Prototyps dienen.

1.2 Zielsetzung und Forschungsfrage

Ziel dieser Arbeit ist die prototypische Implementierung und Evaluation einer KI-gestützten E-Mail-Automation in einem mittelständischen GaLaBau-Handwerksbetrieb. Dazu wird in Zusammenarbeit mit einem anonymisierten Partnerunternehmen ein funktionsfähiger Prototyp entwickelt, der eingehende E-Mails klassifiziert, Anhänge analysiert und zunächst Outlook-Kategorie-Labels setzt, ohne operative Weiterleitungen

automatisch auszulösen. Die automatische Weiterleitung an zuständige Personen oder Funktionspostfächer bildet eine technisch vorbereitete Ausbaustufe. Ihre produktive Aktivierung wird in dieser Arbeit nicht vorausgesetzt, sondern anhand der Labelphase bewertet. Der Beitrag ist technischer und organisatorischer Natur zugleich. Auf der technischen Seite werden Architektur- und Implementierungsentscheidungen für einen datenschutzorientiert gestalteten, lokal betriebenen KI-Workflow dokumentiert, der ohne Transfer von E-Mail-Inhalten an externe KI-Dienstleister auskommt. Auf der organisatorischen Seite wird untersucht, welche Voraussetzungen erfüllt sein müssen, damit ein solcher Prototyp unter realen Bedingungen praxistauglich wird. Die dieser Arbeit zugrunde liegende Forschungsfrage lautet:

Wie lässt sich eine KI-gestützte E-Mail-Automation im GaLaBau-Handwerk prototypisch implementieren, und welche technischen, prozessualen und organisatorischen Voraussetzungen bestimmen ihre Praxistauglichkeit?

Die Frage gliedert sich in zwei Teilaspekte. Erstens wird untersucht, wie der Prototyp technisch aufgebaut und in die bestehende Systemlandschaft eingebettet werden kann. Zweitens wird analysiert, welche Grenzen und Voraussetzungen sich in der praktischen Erprobung ergeben, insbesondere mit Blick auf Prozessklarheit, Kategorienlogik, Stammdatenqualität, Datenschutzanforderungen und wirtschaftliche Verhältnismäßigkeit. Beide Dimensionen werden in Kapitel 5 auf Basis eigener Erhebungs- und Evaluationsdaten beantwortet.

1.3 Aufbau der Arbeit

Die vorliegende Arbeit gliedert sich in sechs Kapitel. Kapitel 2 erarbeitet den theoretischen Rahmen in vier aufeinander aufbauenden Blöcken. Zunächst wird der Digitalisierungsstand im GaLaBau-Handwerk beschrieben und die spezifischen strukturellen, kulturellen und kompetenzbezogenen Hemmnisse der Branche werden herausgearbeitet. Daran schließen sich die technologischen Grundlagen künstlicher Intelligenz an, von klassischen Verfahren des maschinellen Lernens (ML) über Deep Learning bis hin zu Large Language Models (LLMs) und deren Fähigkeit, Aufgaben über präzise formulierte Prompts zu bearbeiten. Der dritte Block behandelt E-Mail-Bearbeitung als betrieblichen Prozess und kognitiven Stressor, wobei Robotic Process Automation (RPA) und KI-gestützte Klassifizierung als komplementäre

Automatisierungsansätze gegenübergestellt werden. Der vierte Block analysiert Implementierungsbedingungen KI-gestützter Automation, insbesondere Prozess- und Datentauglichkeit, Pilotierung, Kontrolllogik, Datenschutzorientierung und wirtschaftliche Verhältnismäßigkeit. Kapitel 3 beschreibt das Forschungsdesign als gestaltungsorientierte Einzelfallstudie und erläutert die eingesetzten Erhebungsinstrumente: einen explorativen Vorerhebungsfragebogen, ein Prozessmapping beim Partnerunternehmen, ein Mail-Tagebuch, eine Labelauswertung von 57 E-Mails sowie eine qualitative Nachbesprechung mit der IT-Kontaktperson des Partnerunternehmens. Kapitel 4 dokumentiert den iterativen Entwicklungsprozess von den frühen Prototypversionen bis zum finalen n8n-Workflow mit lokalem Ollama-Sprachmodell llama3.1:8b und begründet die zentralen Designentscheidungen zu Kategorisierungslogik, Anhangsanalyse, Geocapture-Anbindung und Datenschutzarchitektur. Kapitel 5 wertet die erhobenen Daten aus und leitet daraus Implementierungskriterien für praxistaugliche KI-Automation im Handwerk ab. Kapitel 6 beantwortet die Forschungsfrage, benennt Limitationen und zeigt weiteren Forschungsbedarf auf.

2 Theoretischer Rahmen

Der theoretische Rahmen verknüpft vier Perspektiven, die für die Beurteilung des Prototyps erforderlich sind, den Digitalisierungsstand im GaLaBau-Handwerk, die technologischen Grundlagen KI-gestützter Textverarbeitung, E-Mail-Automation als betrieblichen Prozess und die Implementierungsbedingungen KI-gestützter Systeme. Damit wird begründet, warum technische Machbarkeit, Prozessklarheit, Datenqualität, Kontrolle, Datenschutz und wirtschaftlicher Nutzen in der späteren Evaluation gemeinsam betrachtet werden.

2.1 Digitalisierung im Handwerk

Digitalisierung im Handwerk bildet den Ausgangspunkt für die Beurteilung, ob eine KI-gestützte E-Mail-Automation im GaLaBau-Kontext praxistauglich implementiert werden kann. Der Abschnitt klärt den Begriff der Digitalisierung, ordnet den Prototyp als prozessuale Automatisierung ein und zeigt, welche Hemmnisse und Kompetenzerfordernisse die spätere Bewertung der Lösung prägen.

2.1.1 Begriff und Transformationsebenen

Digitalisierung wird in Wissenschaft und Praxis nicht einheitlich verwendet. Schallmo et al. (2017) stellen fest, dass die Begriffe Digitalisierung und Digitisierung (digitalisation und digitisation) in Veröffentlichungen und Unternehmenskontexten oft synonym gebraucht werden, obwohl sie konzeptionell auf unterschiedlichen Transformationsebenen angesiedelt sind. Diese begriffliche Uneinheitlichkeit erschwert den wissenschaftlichen Diskurs und kann dazu führen, dass unterschiedlich weitreichende Eingriffe in Prozesse und Organisationsstrukturen unter demselben Begriff zusammengefasst werden. Für die vorliegende Arbeit wird daher eine analytische Differenzierung in drei Ebenen zugrunde gelegt, die in der betriebswirtschaftlichen Forschung gebräuchlich ist und eine methodisch präzise Einordnung des Prototyps ermöglicht.

Digitization bezeichnet die technische Überführung analoger Informationen in digitale Formate, etwa die Umwandlung papierbasierter Dokumente in digitale Dateien oder die Erfassung von Messdaten in Datenbanksystemen. Diese Ebene betrifft ausschließlich die Repräsentation von Informationen und verändert Prozesse oder Strukturen noch nicht grundlegend. Die zweite Ebene, Digitalization, umfasst die prozessuale Neugestaltung bestehender betrieblicher Abläufe durch den gezielten Einsatz digitaler Technologien mit

dem Ziel, Effizienz, Geschwindigkeit oder Qualität zu steigern. Digital Transformation schließlich beschreibt Vial (2019) auf Basis einer Synthese von 282 Fachbeiträgen als „a process that aims to improve an entity by triggering significant changes to its properties through combinations of information, computing, communication, and connectivity technologies“ (Vial, 2019, S. 121). In organisatorischer Hinsicht umfasst diese Ebene die umfassende strategische Neuausrichtung eines Unternehmens, bei der bestehende Geschäftsmodelle, Wertschöpfungsketten und Kundenbeziehungen grundlegend verändert werden.

Für die vorliegende Arbeit ist diese Differenzierung aus zwei Gründen relevant. Erstens verdeutlicht sie, dass Digitalisierung kein singuläres Ereignis ist, sondern ein schrittweiser Prozess mit mehreren Transformationsebenen. Eine digitale Ablage von Dokumenten ist deshalb anders zu bewerten als eine Automatisierung, die Abläufe, Zuständigkeiten und Entscheidungspunkte verändert. Zweitens ermöglicht die Differenzierung eine methodisch präzise Einordnung des Prototyps. Dieser ist der zweiten Ebene zuzuordnen, da er einen bestehenden Geschäftsprozess, nämlich die eingehende E-Mail-Bearbeitung, durch KI-gestützte Automatisierung optimiert, ohne das Geschäftsmodell des anonymisierten Partnerunternehmens grundlegend zu verändern. Die dritte Ebene, Digital Transformation, übersteigt den Rahmen einer Einzelfallstudie und wird in Kapitel 3 bei der Abgrenzung des Untersuchungsrahmens explizit ausgeschlossen.

2.1.2 Digitalisierungsstand im Handwerk und GaLaBau

Das Digitalisierungspotenzial für kleine und mittlere Unternehmen (KMU) ist auf internationaler Ebene anerkannt und empirisch gut belegt. Aktuelle Analysen der Organisation for Economic Co-operation and Development (OECD) zeigen, dass die Adoption künstlicher Intelligenz (KI) durch KMU zwar zunimmt, strukturell jedoch erheblich hinter Großunternehmen zurückbleibt. Im OECD-Durchschnitt nutzen 40 Prozent der Großunternehmen KI-Technologien, gegenüber lediglich 11,9 Prozent der Kleinunternehmen. Als treibende Faktoren werden Produktivitätssteigerungen in operativen Prozessen, Kostensenkungen durch Automatisierung repetitiver Aufgaben sowie verbesserte Entscheidungsgrundlagen identifiziert (OECD, 2025). Für KMU in arbeitsintensiven Branchen wie dem Handwerk ist dieses Potenzial besonders ausgeprägt, da manuelle, wiederkehrende Prozesse in diesen Betrieben einen erheblichen Anteil der verfügbaren Kapazitäten binden (Beuchel et al., 2024; Helms et al., 2024).

Die tatsächliche Umsetzung dieses Potenzials bleibt im deutschen Handwerk deutlich zurück. Beuchel et al. (2024) stellen in einer Status-quo-Analyse fest, dass Digitalisierungsrelevanz und Umsetzungsstand über Gewerke hinweg stark variieren und letztlich betriebspezifisch zu betrachten sind, insbesondere kleine, digitalisierungsferne Betriebe sehen sich dabei im strukturellen Nachteil. Besonders ausgeprägt ist der Rückstand beim Einsatz von KI. Über 80 Prozent der Handwerksbetriebe, die KI bereits einsetzen, wenden dies erst seit höchstens zwei Jahren an, was auf eine frühe und noch fragmentierte Adoptionsphase hindeutet (Schwaewe et al., 2024; Komus et al., 2024). Helms et al. (2024) bestätigen, dass die digitale Transformation das Handwerk insgesamt nur langsam durchdringt und Chancen zur Prozessverbesserung und Mitarbeitendenentlastung bislang ungenutzt bleiben. Sagala und Öri (2024) zeigen in einem systematischen Review mit 109 Primärstudien, dass KMU digitale Vorteile aufgrund fehlenden strategischen und praktischen Wissens nicht vollständig ausschöpfen können und dass die Kernerfolgskriterien digitaler Transformation bislang unzureichend geklärt sind.

Für den GaLaBau-Sektor sind diese Befunde in besonderem Maße relevant. Beuchel et al. (2024) zeigen, dass administrative Automatisierung eines der zentralen Digitalisierungsanliegen gerade in kleinen Handwerksbetrieben ist, da sie das Ziel verfolgen, Ressourcen für handwerkliche Kerntätigkeiten freizusetzen. Im GaLaBau-Kontext betrifft dies vor allem repetitive Büroprozesse wie die Bearbeitung von Eingangspost, die Vorsortierung von Anfragen aus der Kundschaft und die Weiterleitung projektbezogener Korrespondenz an zuständige Bauleitungen. Diese Aufgaben müssen in kleinen Betrieben ohne eigenständige Bürostruktur von wenigen Personen parallel zu telefonischer Kundenkommunikation und operativem Tagesgeschäft bewältigt werden. Für den GaLaBau-Kontext zeigt sich damit eine zentrale Ausgangslage der vorliegenden Arbeit. Der Bedarf an administrativer Entlastung ist erkennbar, während die strukturellen Voraussetzungen für digitale Umsetzung zugleich begrenzt bleiben.

2.1.3 Hemmnisse, Kompetenzen und Designimplikationen

Die folgenden Abschnitte verdichten zentrale Voraussetzungen und Barrieren digitaler Transformation im Handwerk. Zunächst werden strukturelle und kulturelle Hemmnisse herausgearbeitet, anschließend digitale Kompetenzanforderungen betrachtet. Daraus werden Designimplikationen abgeleitet, die für die spätere Entwicklung einer KI-gestützten E-Mail-Automation im GaLaBau-Kontext relevant sind.

2.1.3.1 Strukturelle und kulturelle Hemmnisse

Die Diskrepanz zwischen Potenzial und Umsetzung erklärt sich durch ein konsistentes Bündel struktureller, kultureller und kompetenzbasierter Hemmnisse. Kocak und Pawlowski (2022) identifizieren in einer Mixed-Methods-Studie mit zehn Fachpersonen aus dem Handwerk insgesamt 44 Digitalisierungsbarrieren in vier Kategorien. Als bedeutsamste Einzelbarriere nennen die befragten Personen fehlende qualifizierte Fachkräfte. Acht der zehn Befragten stufen diese Barriere als signifikant ein. Danach folgen mangelnde Digitalkompetenzen sowie unklare wirtschaftliche Nutzenperspektiven. Solange kein konkreter Nutznachweis vorliegt, bleibt die Investitionsbereitschaft gering, da das Scheitern von Digitalisierungsprojekten für kleine Betriebe unmittelbar existenzrelevant sein kann (Kocak und Pawlowski, 2022).

Neben diesen operativen Hemmnissen wirken kulturelle Faktoren als eigenständige Barrieren. Helms et al. (2024) zeigen, dass fehlendes Verständnis, Unsicherheiten bei der Einführung und ein Mangel an Vertrauen in die Sinnhaftigkeit digitaler Transformation zu ihrer sehr langsamen Durchdringung im Handwerk führen. Kallmuenzer et al. (2025) bestätigen dies in einer qualitativen Studie mit KMU-Führungskräften. Risikoscheue Unternehmenskultur und das Festhalten an gewachsenen Systemen erweisen sich als zentrale kulturelle Barrieren, die selbst bei vorhandener Technologie einen Wandel verhindern können. Auf individueller Ebene verstärkt Widerstand gegen Veränderungen unter Mitarbeitenden diese Haltung zusätzlich (Kallmuenzer et al., 2025).

Die beschriebenen Hemmnisse führen zu einem strukturellen Ressourcenkonflikt. Digitalisierung kann operative Belastung langfristig senken, benötigt dafür jedoch zunächst zusätzliche Kapazitäten. Diese Kapazitäten sind in kleinen Betrieben häufig durch das Tagesgeschäft gebunden. In Betrieben ohne eigene IT-Abteilung entsteht so ein selbstverstärkender Kreislauf, der Transformationsprojekte dauerhaft verschiebt (Helms et al., 2024). Beuchel et al. (2024) konkretisieren diesen Mechanismus. Insbesondere bei kleinen Betrieben sind Betriebsinhabende und Führungskräfte in der Regel so stark in das operative Tagesgeschäft eingebunden, dass keine Kapazitäten für strategische Transformationsplanung verbleiben. Digitalisierungsthemen werden in der Folge mehrheitlich durch äußere Anreize oder Zwänge angegangen, nicht aus eigenem Antrieb (Beuchel et al., 2024).

2.1.3.2 Kompetenzanforderungen

Eine spezifische Dimension dieses Konflikts betrifft die Kompetenzebene. Kocak und Pawlowski (2023) zeigen in einer auf den Handwerkssektor fokussierten Studie, dass Digitalkompetenzen und insbesondere ein digitales Mindset bei Führungskräften und Mitarbeitenden die entscheidende Voraussetzung für eine erfolgreiche digitale Transformation darstellen. Als kritischste Kompetenzdimensionen im Handwerk identifizieren sie technologische Fähigkeiten, professionelle Fachkenntnisse, die Fähigkeit zur Informationsverarbeitung sowie die Bereitschaft zur Anpassung an digitale Veränderungen. Ohne diese Grundlage scheitern selbst technisch ausgereifte Lösungen häufig an mangelnder organisatorischer Anschlussfähigkeit. Die Bereitschaft zur Nutzung neuer Systeme hängt wesentlich davon ab, ob die beteiligten Personen den Wandel als sinnvoll, beherrschbar und in bestehende Prozesse integrierbar wahrnehmen. Diese Verbindung aus Technologie, Organisation und Prozessgestaltung wird in Kapitel 2.4 als Implementierungsvoraussetzung KI-gestützter Automation systematisch aufgegriffen. Seppänen et al. (2025) bestätigen diese Befunde empirisch in 22 KMU-Fallstudien. Digitale Fähigkeiten von Mitarbeitenden wirken als zentrale Voraussetzung der Transformation, erzeugen jedoch gleichzeitig Herausforderungen durch Kompetenzlücken, Veränderungsresistenz und Abhängigkeit von gewachsenen Systeminfrastrukturen.

2.1.3.3 Designimplikationen für den GaLaBau-Kontext

Die aktuelle Forschung zeigt, dass eine Überwindung der beschriebenen Hemmnisse gelingt, wenn spezifische organisatorische Voraussetzungen gezielt aufgebaut werden. Arroyabe et al. (2024) belegen in einer Analyse europäischer KMU, dass digitale Fähigkeiten der Organisation der stärkste Einflussfaktor für KI-Adoption sind. Für Unternehmen mit stark ausgeprägten digitalen Fähigkeiten prognostiziert die Studie eine KI-Adoptionswahrscheinlichkeit von 51,8 Prozent (Arroyabe et al., 2024). Auf Führungsebene zeigen Held et al. (2026), dass Dynamic Capabilities in KMU digitale Führung und digitale Kultur signifikant stärken. Für den GaLaBau-Kontext verweist dies darauf, dass KI-gestützte Automation nicht nur technisch implementiert, sondern auch organisatorisch getragen werden muss. Schwaewe et al. (2024) bestätigen in einer systematischen Übersicht von 106 Fachartikeln, dass erfolgreiche KI-Adoption von einem Zusammenspiel aus kompatibler Infrastruktur, aufgebaute Wissen, kultureller Offenheit und regulatorischem Rahmen abhängt. Dass dieser regulatorische Aspekt für Handwerksbetriebe besonders wirksam ist, zeigt eine branchenübergreifende Studie unter

Handwerksbeteiligung. 66 Prozent der befragten Unternehmen nennen Rechtssicherheit als die größte Einzelherausforderung beim KI-Einsatz (Schwaeke et al., 2024; Komus et al., 2024). Für den GaLaBau-Kontext ist dieser Befund relevant, weil KI-gestützte E-Mail-Automation personenbezogene Kommunikationsdaten verarbeitet und daher nicht nur technisch, sondern auch rechtlich und organisatorisch abgesichert werden muss. Skare et al. (2023) belegen in einer empirischen Analyse europäischer KMU ergänzend, dass die digitale Transformation betrieblicher Verwaltungsprozesse die Produktivität der Belegschaft steigert und gleichzeitig Kostenstrukturen verbessert, was die betriebswirtschaftliche Begründung für Digitalisierungsinvestitionen empirisch untermauert. Přívara et al. (2025) zeigen in einer quantitativen Studie, dass eine tragfähige IT-Infrastruktur Organisationen dabei unterstützt, schneller auf veränderte Marktbedingungen zu reagieren. Für ressourcenbeschränkte Kleinstunternehmen ist dies besonders relevant, weil digitale Lösungen nur dann skalierbar sind, wenn vorhandene Systeme, Daten und Schnittstellen ausreichend anschlussfähig sind.

Für den GaLaBau-Kontext folgen aus diesen Befunden drei grundlegende Designanforderungen an jeden praxistauglichen KI-Ansatz. Erstens muss die Lösung ohne tiefe IT-Vorkenntnisse implementierbar sein, da entsprechende Kompetenzen strukturell fehlen und Digitalisierungsprojekte andernfalls an der Einstiegshürde scheitern. Zweitens muss sie einen sichtbar prüfbaren Nutznachweis erzeugen, da dieser nach den Befunden von Kocak und Pawlowski (2022) die zentrale Bedingung für Investitionsbereitschaft und Akzeptanz in ressourcenbeschränkten Betrieben darstellt. Drittens muss die Lösung datenschutzorientiert gestaltet und vor produktiver Nutzung rechtlich sowie organisatorisch dokumentiert werden, um die von Schwaeke et al. (2024) identifizierte regulatorische Barriere zu adressieren. Alle drei Anforderungen sind grundlegend für den in Kapitel 4 beschriebenen Prototyp. Die technologischen Grundlagen, die eine solche Automation ermöglichen, werden im folgenden Abschnitt dargelegt.

2.2 Künstliche Intelligenz

Künstliche Intelligenz bildet die technologische Grundlage des in Kapitel 4 beschriebenen Prototyps. Für die Arbeit ist vor allem relevant, warum Sprachmodelle E-Mail-Inhalte ohne eigenes Trainingsprojekt einordnen können und weshalb ihre Ergebnisse dennoch kontrolliert werden müssen. Der Abschnitt erläutert deshalb die Entwicklung von maschinellern Lernen zu Large Language Models, betriebliche

Anwendungspotenziale sowie Grenzen wie Halluzinationen und mangelnde Nachvollziehbarkeit.

2.2.1 Von maschinellern Lernen zu Large Language Models

Für den Prototyp ist nicht künstliche Intelligenz im Allgemeinen relevant, sondern die Entwicklung hin zu Sprachmodellen, die unstrukturierte E-Mail-Inhalte klassifizieren können. Der Abschnitt zeichnet diese Entwicklung in zwei Schritten nach: Zunächst werden KI, maschinelles Lernen und Deep Learning begrifflich abgegrenzt, anschließend werden Large Language Models als technische Grundlage der späteren Klassifikation eingeordnet.

2.2.1.1 Von KI zu Deep Learning

Künstliche Intelligenz bezeichnet Systeme, die eine Umgebung wahrnehmen, auf Basis dieser Wahrnehmung eigenständige Entscheidungen treffen und dabei ein definiertes Ziel verfolgen (Russell und Norvig, 2021). Diese Eigenschaft grenzt KI-Systeme grundlegend von klassischer Software ab. Während herkömmliche Programme ausschließlich vorher festgelegte Befehle ausführen, sind KI-Systeme in der Lage, unbekannte Situationen zu interpretieren und auf dieser Grundlage zu handeln. Für den E-Mail-Kontext ist diese Abgrenzung unmittelbar relevant, da die Kategorisierung eingehender Nachrichten nach Inhalt und Absenderkontext aufgrund der sprachlichen Varianz betrieblicher Kommunikation nicht vollständig regelbasiert abbildbar ist.

Maschinelles Lernen (im Folgenden: ML) erweitert dieses Grundprinzip, indem Systeme statistische Muster aus Trainingsdaten eigenständig erlernen, anstatt auf explizit programmierten Entscheidungsregeln aufzubauen (Sarker, 2021). Übertragen auf E-Mail-Klassifikation bedeutet dies: ein Modell, das klassifizierte E-Mails als Trainingsmaterial erhält, kann auf dieser Basis neue Nachrichten kategorisieren, ohne dass jede mögliche Formulierung im Voraus antizipiert werden müsste. Die statistische Natur dieses Lernprozesses hat jedoch eine grundlegende Konsequenz. ML-Modelle erkennen Korrelationen zwischen Mustern und Kategorien, ohne die inhaltliche Bedeutung von Sprache tatsächlich zu verstehen. Für stark standardisierte Kommunikationsformen kann dies ausreichen. Für variantenreiche betriebliche E-Mails mit wechselnden Absendern, impliziten Fachbegriffen und kontextabhängigen Formulierungen stoßen klassische ML-Ansätze jedoch an Grenzen, wenn sie auf vorab definierten, manuell ausgewählten Sprachmerkmalen beruhen.

Deep Learning vertieft diesen Ansatz durch künstliche neuronale Netze mit mehreren Verarbeitungsschichten, die relevante Sprachmerkmale vollständig selbstständig aus Rohdaten extrahieren, anstatt auf vorab definierten Merkmalen aufzubauen (Sarker, 2021). Diese Fähigkeit zur autonomen Merkmalsbildung ist das technische Fundament moderner Sprachmodelle. Da keine manuelle Auswahl relevanter Merkmale erforderlich ist, können Deep-Learning-Architekturen aus rohen Textdaten semantische Repräsentationen erzeugen. Diese Repräsentationen gehen weit über die Möglichkeiten früherer regelbasierter Klassifikation hinaus und haben den Weg zu leistungsfähigen Sprachmodellen geebnet.

2.2.1.2 Large Language Models: Emergenz und Kontextlernen

Für textbezogene Anwendungen wird diese Entwicklung im Bereich des Natural Language Processing (NLP) relevant. NLP bezeichnet die maschinelle Verarbeitung natürlicher Sprache und umfasst Aufgaben wie Textklassifikation, Informationsextraktion und Zusammenfassung (Alsuhaymi, 2025). Large Language Models (LLMs) sind eine moderne Ausprägung dieses Feldes: Sie bauen auf Deep-Learning-Architekturen auf, werden auf umfangreichen Textsammlungen trainiert und können Aufgaben in natürlicher Sprache ohne aufgabenspezifisches Vortraining lösen (Alsuhaymi, 2025). Ab einer kritischen Modellgröße entstehen dabei sogenannte Emergenzeffekte. Fähigkeiten wie logisches Schlussfolgern oder kontextbezogenes Einordnen, die nicht explizit trainiert wurden, sondern aus der Skalierung des Trainings resultieren (Wei et al., 2022). Für den betrieblichen Einsatz bedeutet dies, dass ein LLM Aufgaben übernehmen kann, für die es kein spezifisches Training erhalten hat, weil die Kombination aus allgemeinem Verständnis sprachlicher Zusammenhänge und einer präzisen Aufgabenbeschreibung ausreicht.

Eine für den betrieblichen Einsatz besonders bedeutsame Eigenschaft stellt in diesem Zusammenhang das sogenannte In-Context Learning dar. Durch die Formulierung einer Aufgabe in der textuellen Eingabe an das Modell, dem Prompt, können LLMs klassifizieren, zusammenfassen oder einordnen, ohne zuvor auf den spezifischen Anwendungsfall trainiert worden zu sein (Alsuhaymi, 2025). Im Unterschied zu herkömmlichen ML-Ansätzen, für die das Zusammenstellen und Annotieren eines Trainingsdatensatzes wochen- oder monatelange Vorbereitung erfordert, genügt beim In-Context Learning eine präzise formulierte Aufgabenbeschreibung im Prompt. Diese Aufgabenbeschreibung enthält Kategoriedefinitionen, Abgrenzungsregeln und bei Bedarf

wenige Beispiele. Für KMU ohne eigene KI-Expertise hat dies eine unmittelbare praktische Konsequenz. Die Implementierung einer E-Mail-Klassifizierung erfordert keine datenwissenschaftliche Fachkraft und keine aufwendige Modellentwicklung. Erforderlich ist zunächst eine klar strukturierte Aufgabenbeschreibung, die das Modell in die Lage versetzt, neue E-Mails zuverlässig einzuordnen.

2.2.2 Betriebliche Anwendung und Grenzen

Nachdem Abschnitt 2.2.1 die technische Entwicklung bis zu LLMs dargestellt hat, richtet Abschnitt 2.2.2 den Blick auf deren betriebliche Nutzung. Für die vorliegende Arbeit lassen sich LLMs als textbasierte Ausprägung generativer KI einordnen, da sie aus sprachlichen Eingaben neue Textausgaben, Zusammenfassungen, Klassifikationen oder Entscheidungsvorschläge erzeugen können. Generative KI bezeichnet KI-Verfahren, die auf Basis gelernter Datenmuster neue, sinnvolle Inhalte wie Text, Bilder oder Audio erzeugen können (Feuerriegel et al., 2023). Damit bildet generative KI den anwendungsbezogenen Rahmen für die folgenden Ausführungen zu Produktivitätspotenzialen und Einsatzgrenzen.

2.2.2.1 Produktivitätspotenzial und Anwendungsfelder

Im betrieblichen Einsatz zeigen generative KI-Systeme empirisch messbare Produktivitätssteigerungen, insbesondere bei wissensintensiven Tätigkeiten wie Texterstellung, Klassifizierung und Entscheidungsvorbereitung (Ibrahim et al., 2024; Feuerriegel et al., 2023). Feuerriegel et al. (2023) quantifizieren in einer kontrollierten Studie mit Wissensarbeitenden, dass die Nutzung generativer KI die Bearbeitungszeit vergleichbarer Aufgaben signifikant reduziert, wobei besonders weniger erfahrene Nutzende überproportional profitierten, da KI-Unterstützung einen Teil des Wissensunterschieds zwischen erfahrenen Fachkräften und Einsteigenden ausgleicht. Handler et al. (2024) belegen darüber hinaus, dass LLMs strukturierte Entscheidungsunterstützung in Unternehmen leisten können, indem sie eingehende Informationen kategorisieren, priorisieren und weiterleiten. Diese Funktionen entsprechen dem in Kapitel 4 beschriebenen Routingprozess (automatisierte Zuordnung und Weiterleitung eingehender E-Mails). Wei und Pardo (2022) zeigen in einer qualitativen Fallstudie, dass KMU zur Integration von KI-Technologien typischerweise auf externe Plattformlösungen angewiesen sind, da eigenständige Entwicklungsressourcen fehlen und Zugang zu technologischem Wissen den entscheidenden Adoptionsfaktor darstellt. Ardito et al. (2025) belegen zudem für

europäische KMU empirisch, dass KI-Adoption positiv mit Umsatzwachstum korreliert und in Kombination mit komplementären digitalen Technologien besonders wirksam ist.

Ein für die E-Mail-Verarbeitung technisch relevanter Aspekt betrifft die gemeinsame Verarbeitung der gesamten Nachricht. Moderne LLMs können Betreff und Nachrichtentext zusammen analysieren, sodass kontextuell relevante Hinweise erhalten bleiben (Alsuhaymi, 2025). Diese Fähigkeit zur Verarbeitung zusammenhängender Eingaben ist für die betriebliche Klassifizierung besonders wertvoll. Formulierungen, Anredestile und implizite Hinweise auf den Absendertyp können die Kategorisierungsentscheidung maßgeblich beeinflussen. Eine E-Mail mit dem Betreff „Projekt Musterstraße“ kann je nach Textinhalt und Anredeform eine Angebotsanfrage einer neu anfragenden Person, eine Statusmeldung aus der Bauleitung oder eine Eingangsrechnung eines Subunternehmens sein. Erst die Analyse aller verfügbaren Kontextinformationen ermöglicht eine zuverlässige Zuordnung (Alsuhaymi, 2025).

2.2.2.2 Grenzen und Erklärbarkeit

Der produktive Einsatz generativer KI ist jedoch an eine wesentliche Einschränkung gebunden. LLMs neigen zu sogenannten Halluzinationen, sprachlich kohärenten und überzeugend formulierten Ausgaben, die faktisch inkorrekt sind (Schneider et al., 2024). Dieser Effekt entsteht, weil LLMs auf statistischen Wahrscheinlichkeitsverteilungen über Sprachsequenzen basieren und nicht über ein gesichertes Faktenwissen verfügen, das Modell generiert die wahrscheinlichste Fortsetzung eines Textes, nicht zwangsläufig die inhaltlich korrekte. Bei ambivalenten oder ungewöhnlichen E-Mails, die keinem definierten Kategorieschema eindeutig entsprechen, kann dies zu plausibel klingenden, aber fehlerhaften Klassifizierungen führen. Für operative Geschäftsprozesse hat dies direkte Konsequenzen: Eine falsch klassifizierte E-Mail kann beispielsweise eine Angebotsanfrage an einen falschen Zielbereich lenken und dadurch betriebliche Folgeprobleme verursachen. Deshalb bleibt eine menschliche Prüfung erforderlich. Im Prototyp wird diese Kontrolle durch die Label-Phase umgesetzt: Das System leitet E-Mails zunächst nicht produktiv weiter, sondern versieht sie mit Kategorien, die anschließend fachlich überprüft werden. Erst auf dieser Grundlage kann entschieden werden, ob und unter welchen Bedingungen eine spätere Vollautomation verantwortbar ist (Kapitel 4.2.2).

Die mangelnde Nachvollziehbarkeit von LLM-Entscheidungen verstärkt diese Problematik. Brasse et al. (2023) beschreiben das sogenannte Black-Box-Dilemma. Da die internen Verarbeitungsprozesse für Nutzende nicht einsehbar sind, entsteht die Gefahr, dass entweder blindes Vertrauen in Systemausgaben entsteht oder die Technologie pauschal abgelehnt wird. Der Erfolg der gesellschaftlichen und wirtschaftlichen Integration von KI-Systemen hängt maßgeblich vom Vertrauen der Nutzenden ab. Glikson und Woolley (2020) zeigen in einem Review von 20 Jahren empirischer KI-Vertrauensforschung, dass die Fähigkeit zur erfolgreichen KI-Integration in Organisationen entscheidend davon abhängt, ob Mitarbeitende der Technologie vertrauen. Riedl (2022) ordnet Vertrauen ebenfalls als zentrale Bedingung ein. Beide Extreme sind für einen operativen Einsatz problematisch, da blindes Vertrauen unkontrollierte Fehler ermöglicht, während pauschale Ablehnung das Nutzungspotenzial der Technologie vollständig zunichte macht. An dieser Stelle setzt Explainable AI (im Folgenden: XAI) an. XAI bezeichnet Ansätze, die KI-Entscheidungen zumindest in Grundzügen transparent und nachvollziehbar machen, um diesen Vertrauensmangel zu überbrücken (Arrieta et al., 2020). Arrieta et al. (2020) zeigen, dass Erklärbarkeit eine Grundvoraussetzung für die Akzeptanz von KI-Systemen darstellt, da sie Vertrauen nicht auf blinder Übernahme, sondern auf nachvollziehbarer Systemlogik aufbaut. Benk et al. (2025) bestätigen in einer Analyse von 24 Jahren KI-Vertrauensforschung, dass die Vertrauenswürdigkeit eines KI-Systems multidimensional von Eigenschaften wie Zuverlässigkeit, Robustheit und Transparenz abhängt. Im Kontext des Prototyps wird XAI durch einen numerischen Konfidenzwert operationalisiert. Unterschreitet die Klassifizierungssicherheit einen definierten Schwellenwert, signalisiert das System explizit Unsicherheit und übergibt die Entscheidung an die menschliche Kontrollinstanz, anstatt eine möglicherweise fehlerhafte Kategorie auszugeben. Für eine betriebliche Implementierung reicht diese technische Kontrolllogik jedoch nicht aus, sie muss mit geeigneten Prozessen, Datenstrukturen und organisatorischen Rahmenbedingungen verbunden werden.

LLMs verfügen damit über die technischen Voraussetzungen, Texte zu klassifizieren, Intentionen zu erkennen und Entscheidungen strukturiert zu unterstützen. Das folgende Kapitel zeigt, wie diese Fähigkeiten auf das E-Mail-Management als automatisierbaren Geschäftsprozess übertragen werden können und welche Implementierungsansätze für KMU im GaLaBau-Kontext geeignet sind.

2.3 Prozessautomatisierung im E-Mail-Management

E-Mail-Management bildet die Verbindung zwischen technischer Klassifikation und betrieblicher Entlastung. Der Abschnitt beschreibt E-Mail zunächst als zeitaufwendigen Verwaltungsprozess und ordnet die Belastungswirkung anschließend mithilfe der Action Regulation Theory ein. Darauf aufbauend werden regelbasierte und KI-gestützte Automatisierungsansätze gegenübergestellt, aus denen die Prozesslogik des in Kapitel 4 beschriebenen Prototyps abgeleitet wird.

2.3.1 E-Mail als Geschäftsprozess

E-Mail-Bearbeitung wird im Folgenden nicht nur als Kommunikationsform, sondern als wiederkehrender Verwaltungsprozess betrachtet. Zunächst wird gezeigt, warum Sichtung, Einordnung und Weiterleitung eingehender Nachrichten zeitlich und organisatorisch relevant sind. Anschließend wird die Action Regulation Theory herangezogen, um die besondere Belastungswirkung von Unterbrechungen und unklaren Zuständigkeiten einzuordnen.

2.3.1.1 Zeitaufwand und Belastungsstruktur

E-Mail-Bearbeitung stellt in modernen Büroumgebungen einen zeitlich bedeutsamen Verwaltungsprozess dar. Studien zu Wissensarbeit, also Tätigkeiten mit hohem Anteil an Informationsverarbeitung, Kommunikation und Entscheidungsvorbereitung, zeigen einen erheblichen Zeitaufwand für das Lesen, Sortieren und Beantworten eingehender Nachrichten (Martin et al., 2025; Lanctot und Duxbury, 2025). Erschwerend kommt hinzu, dass betriebliche Postfächer in der Regel keine homogenen Nachrichtentypen empfangen. Stattdessen gehen dort unterschiedlichste Nachrichten ein, darunter Anfragen der Kundschaft, Lieferscheine, Bewerbungen, Projektkorrespondenz und Werbe-E-Mails. Diese müssen vor der inhaltlichen Bearbeitung zunächst manuell eingeordnet werden. Übertragen auf den GaLaBau-Kontext verschärft sich diese Belastung in Betrieben ohne spezialisierte IT- oder Prozessmanagementstrukturen. Sortier-, Rückfrage- und Weiterleitungsaufgaben konzentrieren sich dann auf wenige Personen im Büroalltag, die gleichzeitig Kundenkommunikation, Auftragsabwicklung und weitere administrative Aufgaben bewältigen müssen. Das quantitative Ausmaß allein erklärt jedoch nicht, warum E-Mail-Bearbeitung als besonders belastend wahrgenommen wird, hierfür bedarf es eines psychologischen Erklärungsrahmens.

2.3.1.2 Action Regulation Theory

Die Action Regulation Theory (im Folgenden: ART) liefert diesen Rahmen. ART beschreibt, wie Menschen zielgerichtete Handlungen in hierarchisch organisierten Sequenzen planen und ausführen. Übergeordnete Ziele werden in Teilziele gegliedert, die wiederum in konkrete Handlungsschritte überführt werden, wobei das Arbeitsgedächtnis die jeweilige Position in dieser Hierarchie aktiv aufrechterhalten muss (Zacher und Frese, 2018). Dabei werden drei Kategorien von Regulationshindernissen unterschieden. **Regulationsbehinderungen** liegen vor, wenn ein Ziel zwar klar definiert ist, seine Umsetzung jedoch blockiert wird. Von **Regulationsüberforderung** wird gesprochen, wenn die Anforderungen die verfügbaren Kapazitäten übersteigen. **Regulationserschwernisse** entstehen hingegen durch Unterbrechungen, die eine laufende Handlungssequenz stören und deren Wiederaufnahme zusätzlichen kognitiven Aufwand erfordert (Zacher und Frese, 2018). Die ART gilt als empirisch gut replizierter Rahmen zur Erklärung arbeitsbedingter Stressreaktionen und bildet die konzeptionelle Grundlage für aktuelle E-Mail-Forschung.

E-Mail-Unterbrechungen lassen sich als Regulationserschwernis einordnen. Eingehende Nachrichten treffen zu unvorhersehbaren Zeitpunkten ein und unterbrechen laufende Arbeitssequenzen, wodurch ein konkurrierendes Handlungsziel entsteht, das unmittelbare Aufmerksamkeit fordert (Kern et al., 2024). Die Wiederaufnahme der unterbrochenen Aufgabe erfordert, dass das Arbeitsgedächtnis den vorherigen Bearbeitungsstand vollständig rekonstruiert. Dieser Vorgang ist kognitiv aufwendig, begünstigt bei häufigen Unterbrechungen kumulative Erschöpfung und kann die Qualität nachfolgender Entscheidungen beeinträchtigen. Hinzu kommt ein strukturell bedingter Erwartungsdruck. Mitarbeitende fühlen sich zunehmend gezwungen, E-Mails zeitnah und auch außerhalb regulärer Arbeitszeiten zu beantworten, was den Erholungsprozess zusätzlich hemmt und den Erschöpfungskreislauf verlängert (Lanctot und Duxbury, 2025).

Die Forschungslage zu diesen Effekten ist konsistent und über Kontexte hinweg robust. Ein systematischer Review über 25 Jahre E-Mail-Forschung zeigt, dass Fragmentierungseffekte durch unstrukturierte E-Mail-Bearbeitung über verschiedene Branchen, Organisationsgrößen und kulturelle Kontexte hinweg reproduzierbar nachgewiesen wurden, wobei die ART als übergreifendes Rahmenmodell besonders gut geeignet ist, die zugrundeliegenden Mechanismen zu erklären (Russell et al., 2024). Der Review hebt dabei hervor, dass individuelle Bewältigungsstrategien allein, etwa das bewusste Prüfen von E-Mails zu festen Zeitpunkten, den strukturellen Stressor nicht

beseitigen, sondern lediglich abmildern, nachhaltige Entlastung erfordert organisatorische oder technische Lösungen, die die Unterbrechungshäufigkeit strukturell reduzieren. Damit ist die Bearbeitung von E-Mails nicht allein ein Effizienzproblem, sondern ein systemisches Belastungsphänomen, das durch fehlende inhaltliche Vorsortierung verstärkt wird und gezielte Automatisierungsansätze begründet.

2.3.2 Robotic Process Automation und KI-gestützte Klassifizierung

Für E-Mail-Automation kommen sowohl regelbasierte als auch KI-gestützte Ansätze in Betracht. Der folgende Abschnitt grenzt zunächst die Logik und Grenzen von Robotic Process Automation ab, bevor empirische Praxisbeispiele KI-gestützter Klassifizierung betrachtet werden. Dadurch wird deutlich, warum der in Kapitel 4 entwickelte Prototyp regelbasierte Workflow-Orchestrierung mit sprachmodellbasierter Inhaltsklassifikation verbindet.

2.3.2.1 Logik und Grenzen

Robotic Process Automation (im Folgenden: RPA) stellt einen etablierten Automatisierungsansatz dar, der ohne tiefen Eingriff in bestehende IT-Systeme auskommt. RPA-Software operiert auf der Präsentationsschicht vorhandener Anwendungen und ahmt regelbasierte Nutzungsinteraktionen nach, sodass bestehende Software wie E-Mail-Clients oder ERP-Systeme automatisiert bedienbar wird, ohne dass Schnittstellen auf Datenbankebene erforderlich sind (Herm et al., 2023). Für kleine und mittlere Unternehmen (KMU) ohne eigene IT-Infrastruktur ist dieser Ansatz besonders attraktiv, da die Implementierung auf vorhandenen Systemen aufsetzt und keine grundlegende Infrastrukturveränderung erfordert. Der Einsatz birgt jedoch erhebliche Risiken. Herm et al. (2023) zeigen in einer branchenübergreifenden Analyse, dass bis zur Hälfte aller RPA-Projekte ohne angemessene Vorbereitung scheitert, wobei ungeeignete Prozessauswahl und fehlende Change-Management-Maßnahmen als häufigste Ursachen identifiziert werden. Das von Herm et al. (2023) entwickelte elfstufige Framework adressiert dieses Risiko durch eine strukturierte Abfolge von Prozesseignungsprüfung, Pilotimplementierung und kontrollierter Skalierung. Für KMU empfiehlt sich auf dieser Grundlage ein sequenzieller Einstieg mit einem einzelnen, klar definierten Prozess, bevor weitere Bereiche in die Automation einbezogen werden.

Im E-Mail-Kontext hat sich RPA als technisches Referenzmodell für die Sortierung nach strukturierten Metadaten bewährt. Khare et al. (2022) beschreiben, wie RPA-Bots eingehende E-Mails anhand von Absenderadressen, Betreffzeilen und Anhangformaten

identifizieren und an definierte Zieladressen weiterleiten können, ohne dass manuelle Eingriffe erforderlich sind. Die Grenzen dieses regelbasierten Ansatzes zeigen sich jedoch beim für GaLaBau-Betriebe typischen Kommunikationsprofil besonders deutlich. Ein und dieselbe absendende Stelle, etwa ein Subunternehmen oder ein Baustofflieferbetrieb, kann im selben Zeitraum Lieferscheine, Rechnungen und projektbezogene Rückfragen an dasselbe Postfach senden, sodass die Absenderadresse als alleiniges Routingkriterium versagt. Eine Angebotsanfrage, eine Beschwerde und eine projektbezogene Statusmeldung können zudem identische Metadaten aufweisen und lassen sich ohne inhaltliches Textverständnis nicht zuverlässig unterscheiden.

2.3.2.2 Empirischer Praxisnachweis

KI-gestützte Klassifizierungsverfahren adressieren genau diese Einschränkung. NLP ermöglicht die kontextabhängige Analyse von E-Mail-Inhalten jenseits einfacher Stichwortsuche. Deep-Learning-Modelle identifizieren dabei nicht nur Themenkategorien, sondern unterscheiden auch zwischen Handlungsintentionen wie Informationsanfragen, Terminvorschlägen oder verbindlichen Zusagen (Wang et al., 2019). In Kombination mit der in Kapitel 2.2 beschriebenen LLM-Technologie entsteht so ein Klassifizierungssystem, das regelbasiertes Routing und semantisches Textverständnis verbindet und damit die Schwächen reiner RPA-Ansätze strukturell überwindet. Der in Kapitel 4 beschriebene Prototyp greift genau auf dieses Zusammenspiel zurück. LLM-basierte Inhaltsanalyse übernimmt die Kategorienzuordnung, während ein nachgelagerter n8n-Workflow zunächst Labels setzt und eine spätere regelbasierte Weiterleitung technisch vorbereitet.

Einen empirischen Praxisnachweis für NLP-basierte E-Mail-Klassifikation in einer Behördenumgebung liefern Sundberg et al. (2024). Eine Fallstudie bei einer schwedischen Bundesbehörde zeigt, dass eingehende E-Mails auf Basis ihres Inhalts automatisch kategorisiert und an zuständige Stellen weitergeleitet werden können, wobei ein auf historischen E-Mail-Daten trainiertes Sprachmodell für die Klassifikationsentscheidung eingesetzt wurde (Sundberg et al., 2024). Ein konkretes Anwendungsbeispiel aus der kommerziellen Praxis dokumentiert eine Fallstudie des KI-Dienstleisters Kortical¹. Bei einer britischen Großbank, die jährlich über 500.000 E-Mails

¹ Kortical: Customer services email automation with AutoML. Online-Fallstudie eines KI-Dienstleisters. <https://kortical.com/case-studies/email-automation/> (zuletzt abgerufen am 03.06.2026).

manuell bearbeitete, ordnete ein mit weniger als 300 Beispiel-E-Mails trainiertes Modell nach Anbieterangaben rund 57 Prozent der eingehenden Nachrichten mit einer berichteten Genauigkeit von 95 Prozent automatisch der zuständigen Abteilung zu. Unsichere Fälle wurden weiterhin manuell bearbeitet. Da es sich um einen Anbieterbericht mit wirtschaftlichem Eigeninteresse und nicht um eine begutachtete Untersuchung handelt, dient diese Fallstudie ausschließlich als Praxisillustration und in Kapitel 5 als grobe Vergleichsfolie, nicht als direkt übertragbarer Zielwert. Die absolute Übertragbarkeit ist einzuschränken, da Praxisberichte aus industriellen oder serviceorientierten Großunternehmen ein höheres E-Mail-Volumen und eine abweichende Kategorienstruktur aufweisen als ein GaLaBau-KMU. Für die vorliegende Arbeit ist daher nicht nur relevant, ob Klassifikation technisch möglich ist, sondern unter welchen Prozess- und Datenbedingungen eine solche Klassifikation im Handwerkskontext praxistauglich wird.

Die empirische Literatur zeigt damit, dass E-Mail-Bearbeitung einen relevanten kognitiven Belastungsfaktor darstellt und sowohl regelbasierte als auch KI-gestützte Automatisierungsansätze strukturelle Entlastung ermöglichen können. Zugleich bleibt technische Machbarkeit nur eine Voraussetzung erfolgreicher Implementierung. Für die vorliegende Arbeit ist daher zu klären, unter welchen Prozess-, Daten- und Organisationsbedingungen KI-gestützte E-Mail-Automation im Handwerkskontext praxistauglich werden kann. Diese Implementierungsbedingungen werden im folgenden Kapitel systematisch hergeleitet.

2.4 Implementierungsbedingungen im KMU-Kontext

Die Praxistauglichkeit KI-gestützter E-Mail-Automation hängt nicht allein von der Klassifikationsleistung ab. Für den späteren Bewertungsrahmen sind ebenso Prozessstabilität, Datenqualität, Kontrolllogik, Datenschutz und wirtschaftliche Verhältnismäßigkeit relevant. Der Abschnitt leitet diese Kriterien als Implementierungsbedingungen im KMU-Kontext her.

2.4.1 Prozess- und Datentauglichkeit

KI-gestützte Automation setzt voraus, dass der zu automatisierende Prozess ausreichend stabil, wiederholbar und formal beschreibbar ist. Für Robotic Process Automation (RPA) zeigen Herm et al. (2023), dass Prozessauswahl und Implementierungsvorbereitung zentrale Erfolgsbedingungen sind, weil ungeeignete Prozessauswahl und fehlende

Change-Management-Maßnahmen zu den häufigsten Ursachen gescheiterter Automatisierungsvorhaben zählen. Das von Herm et al. (2023) entwickelte Implementierungsframework sieht deshalb eine schrittweise Abfolge aus Prozessidentifikation, Eignungsprüfung, Pilotimplementierung, Business Case, Rollout und Skalierung vor. Für KI-gestützte E-Mail-Automation bedeutet dies, dass nicht die technische Klassifikation allein im Zentrum steht, sondern die vorgelagerte Frage, ob der betriebliche Bearbeitungsprozess formal genug beschrieben werden kann.

Die Auswahl geeigneter Automatisierungsprozesse hängt wesentlich von Prozessstabilität, Prozessstandardisierung und klaren Input- und Output-Daten ab. Farinha et al. (2024) identifizieren unter anderem Machbarkeit, genaue Prozessbeschreibung, strukturierte Ein- und Ausgangsdaten, Prozessstabilität, Standardisierung, Testdaten und Wiederholbarkeit als relevante Kriterien der RPA-Prozessauswahl. Diese Kriterien sind auf KI-gestützte E-Mail-Automation übertragbar, obwohl die Klassifikationsentscheidung nicht rein regelbasiert erfolgt. Auch ein LLM kann nur dann zuverlässig in einen Workflow eingebunden werden, wenn Zielkategorien, Eskalationsregeln und Weiterleitungsziele fachlich eindeutig definiert sind. Unklare Zuständigkeiten werden durch KI daher nicht aufgelöst, sondern in automatisierte Unsicherheit überführt.

Prozessstandardisierung ist dabei nicht mit vollständiger Vereinheitlichung gleichzusetzen. Romero et al. (2015) zeigen, dass der angemessene Standardisierungsgrad kontextabhängig ist und eine Balance zwischen Einheitlichkeit und zulässiger Variabilität erfordert. Afflerbach et al. (2016) ergänzen, dass Standardisierung als wirtschaftliche Abwägung zu verstehen ist, weil Verbesserungen bei Kosten, Zeit und Qualität mit Einbußen an Flexibilität verbunden sein können. Für E-Mail-Automation im Handwerk folgt daraus, dass nicht jedes Eingangsszenario automatisiert werden muss. Praxistauglich ist vielmehr ein differenzierter Automatisierungsgrad, bei dem häufige, stabile und risikoarme Fälle automatisiert werden, während seltene oder mehrdeutige Fälle bewusst im manuellen Prozess verbleiben.

Neben der Prozesslogik bildet Datenqualität eine eigenständige Voraussetzung für Automatisierbarkeit. Wang und Strong (1996) zeigen, dass Datenqualität aus Sicht der Datennutzenden nicht nur Korrektheit umfasst, sondern auch Vollständigkeit, Zugänglichkeit, Interpretierbarkeit, Relevanz und Passung zum Nutzungskontext. Diese breitere Sicht ist für KI-gestützte E-Mail-Automation besonders relevant, weil eine

funktionierende technische Schnittstelle keinen Nutzen erzeugt, wenn Projektnummern, Zuständigkeiten oder Stammdaten im Fachsystem nicht eindeutig gepflegt sind. Datenqualität wird damit zur praktischen Grenze technischer Automation. Ein Workflow kann nur jene Entscheidungen zuverlässig ausführen, die in den zugrundeliegenden Datenstrukturen ausreichend eindeutig repräsentiert sind.

2.4.2 Pilotierung, Kontrolle und Wirtschaftlichkeit

KI-gestützte Automation sollte vor einer produktiven Vollautomation in einer kontrollierten Pilotphase erprobt werden. Herm et al. (2023) beschreiben den Pilot als Phase, in der technische und finanzielle Machbarkeit überprüft, der automatisierte Ablauf mit dem bisherigen Prozess verglichen und Akzeptanz im Unternehmen aufgebaut wird. Für KI-gestützte E-Mail-Automation ist eine solche Pilotphase besonders bedeutsam, weil Klassifikationsfehler nicht nur technische Abweichungen darstellen, sondern direkt operative Fehlleitungen erzeugen können. Merhi (2022) verweist in diesem Zusammenhang auf hohe Scheiterquoten bei KI- und Big-Data-Projekten und führt diese unter anderem auf fehlende Pilotstrategien und mangelnde Vorbereitung zurück. Eine vorgeschaltete Labelphase, in der das System Kategorien setzt, ohne automatisch weiterzuleiten, reduziert dieses Risiko und ermöglicht zugleich die systematische Beobachtung von Fehlerbildern, Grenzfällen und fehlenden Kategorien. Robertson et al. (2025) zeigen ergänzend, dass KI-Implementierungen neben technischen Anpassungen einen umfassenden Change-Management-Ansatz erfordern, da Organisationskultur, Führung und Prozessstruktur die Integrationsfähigkeit maßgeblich bestimmen.

Kontrolllogik ist eine Kernbedingung für verantwortbare KI-Automation. Die in Kapitel 2.2 dargestellten Grenzen von LLMs, insbesondere Halluzinationen und mangelnde Nachvollziehbarkeit, machen menschliche Überprüfung in unsicheren Fällen erforderlich. Im Automatisierungskontext bedeutet dies, dass ein System nicht nur richtige Entscheidungen treffen soll, sondern auch seine eigenen Einsatzgrenzen anzeigen muss. Konfidenzschwellen, manuelle Prüfpfade und klar definierte Ausschlussregeln verhindern, dass unklare Fälle unter dem Anschein technischer Sicherheit automatisiert weiterverarbeitet werden (Arrieta et al., 2020). Damit wird Kontrolle nicht als nachträgliche Korrektur verstanden, sondern als Bestandteil der Prozessarchitektur.

Datenschutz ist im KMU-Kontext als Implementierungsbestandteil und nicht als nachgelagerte Formalität zu behandeln. Automatisierte Verarbeitung von E-Mail-Inhalten kann personenbezogene Daten, Bewerbungsunterlagen, Rechnungen und

projektbezogene Informationen betreffen. Sartor und Lagioia (2020) zeigen, dass automatisierte Datenverarbeitung im Kontext der Datenschutz-Grundverordnung (DSGVO) Anforderungen an Transparenz, Rechtmäßigkeit und menschliche Überprüfbarkeit auslösen kann. Für KI-gestützte E-Mail-Automation folgt daraus, dass Architekturentscheidungen, Dienstleisterauswahl, Auftragsverarbeitung, Zugriffskonzepte und Dokumentation vor produktiver Nutzung geklärt werden müssen. Der lokale Betrieb von KI-Modellen kann Datenschutzrisiken reduzieren, ersetzt jedoch keine organisatorische Datenschutzprüfung.

Die wirtschaftliche Verhältnismäßigkeit entscheidet darüber, ob technische Machbarkeit in einen sinnvollen betrieblichen Einsatz überführt werden kann. Reijers und Liman Mansar (2005) zeigen für Business Process Redesign, dass Prozessverbesserung anhand von Zeit, Kosten, Qualität und Flexibilität bewertet werden sollte und zwischen diesen Dimensionen Zielkonflikte bestehen können. Eine reine Betrachtung direkter Bearbeitungszeit greift daher zu kurz. Für E-Mail-Automation können auch Durchlaufzeit bis zur richtigen Zustellung, Fehlervermeidung, Koordinationsaufwand und Entlastung bei unklaren Zuständigkeiten relevante Nutzendimensionen sein. Gleichzeitig muss der Aufwand für Implementierung, Betrieb, Prompt-Pflege, Datenschutzdokumentation und Prozessanpassung berücksichtigt werden.

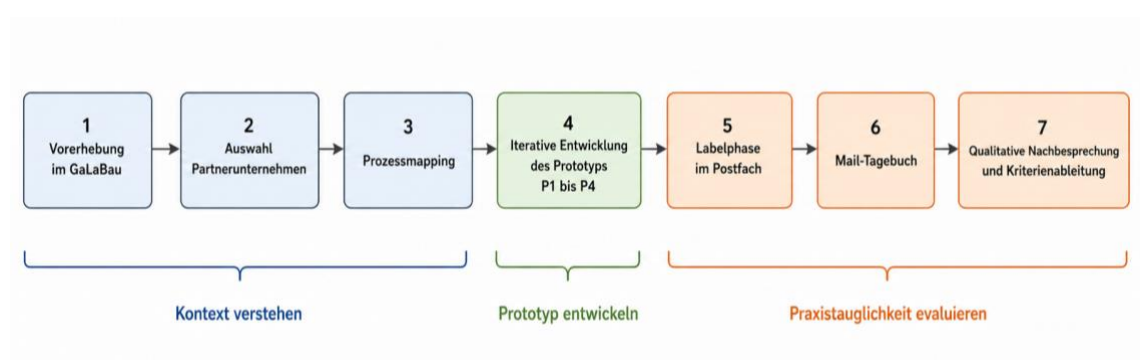
Aus den dargestellten Bedingungen ergibt sich ein theoretischer Kriterienrahmen für die spätere Evaluation. Praxistaugliche KI-gestützte Automation benötigt einen stabil beschreibbaren Prozess, ein trennscharfes Kategorienschema, nutzungstaugliche Datenstrukturen, einen angemessenen Automatisierungsgrad, eine kontrollierte Pilotphase, eine datenschutzorientierte Architektur und eine wirtschaftlich begründbare Nutzenlogik. Diese Kriterien werden in Kapitel 5 nicht erneut theoretisch hergeleitet, sondern auf den entwickelten Prototyp angewendet.

3 Methodik

Kapitel 3 beschreibt, wie die Forschungsfrage methodisch operationalisiert wird. Im Mittelpunkt steht ein gestaltungsorientiertes Einzelfalldesign, das die Entwicklung eines KI-gestützten E-Mail-Prototyps mit einer formativen Evaluation seiner Praxistauglichkeit verbindet. Die Erhebungsinstrumente dienen dabei nicht nur der Datensammlung, sondern bilden die Brücke zwischen Anwendungskontext, Prototypentwicklung und Bewertung in Kapitel 5.

Das Forschungs- und Entwicklungsdesign folgt einer sequenziellen Logik, in der zunächst der Anwendungskontext verstanden, anschließend der Prototyp entwickelt und abschließend seine Praxistauglichkeit evaluiert wird. Abbildung 1 visualisiert diese Abfolge und zeigt, welche Datengrundlagen in die einzelnen Arbeitsschritte eingehen.

Abbildung 1: Design der Forschungsmethode



Anmerkung: Eigene Darstellung.

Die Sequenz in Abbildung 1 operationalisiert die Forschungsfrage in drei Schritten. Die Vorerhebung erfasst zunächst Bedarf, Akzeptanzbarrieren und Gestaltungshinweise im GaLaBau-Kontext. Das Prozessmapping überführt diese Ausgangslage in eine konkrete betriebliche Prozesslogik und bildet damit die Grundlage für die Prototypentwicklung in Kapitel 4. Mail-Tagebuch, Labelauswertung und qualitative Nachbesprechung liefern anschließend die Evidenz, anhand derer in Kapitel 5 technische, prozessuale, datenbezogene und wirtschaftliche Grenzen der Praxistauglichkeit beurteilt werden.

3.1 Forschungsdesign und methodischer Rahmen

Die Arbeit folgt einem gestaltungsorientierten Forschungsziel, da ein KI-gestützter E-Mail-Automatisierungsprototyp unter realen betrieblichen Bedingungen entwickelt und

evaluiert wird. Ein technisches Artefakt wird entwickelt, im Praxiskontext erprobt und anhand definierter Kriterien evaluiert (Hevner et al., 2004). Entscheidend ist dabei die Frage, ob das entwickelte Artefakt seine angestrebte Wirkung erzeugt und unter welchen Bedingungen es in vergleichbaren Kontexten einsetzbar ist. Die eingesetzten Erhebungsinstrumente dienen daher sowohl dem Nachweis von Klassifikationsleistung als auch der Identifikation von Implementierungsgrenzen.

Als methodischer Rahmen dient die Einzelfallstudie nach Yin (2014), der die Fallstudie definiert als „an empirical inquiry that investigates a contemporary phenomenon (the 'case') in depth and within its real-world context“ (Yin, 2014, S. 16). Diese Strategie eignet sich insbesondere dann, wenn die Grenzen zwischen dem untersuchten Phänomen und seinem organisatorischen Kontext nicht klar zu ziehen sind. Diese Bedingung trifft auf die Einführung eines KI-Systems in den spezifischen Büroalltag eines Betriebs aus dem Garten- und Landschaftsbau (GaLaBau) zu. Das vorliegende Mehrinstrumentendesign aus Vorerhebung, Prozessmapping, Mail-Tagebuch, Labelauswertung und qualitativer Nachbesprechung entspricht dieser Anforderung an den Einsatz mehrerer Evidenzquellen.

Als Untersuchungsfall wurde ein anonymisiertes Partnerunternehmen mit rund 300 Mitarbeitenden und klar getrennten Abteilungen ausgewählt. Die Auswahlkriterien umfassen ein konkretes, klar beschreibbares Anwendungsproblem, eine datenschutzrechtliche Anforderung an lokale KI-Verarbeitung und die aktive Bereitschaft zur Mitwirkung in der Forschungs Kooperation. Das Anwendungsproblem bestand in einem heterogenen Info@-Postfach mit täglich variablem Kommunikationsmix, die lokale KI-Verarbeitung bildete eine reale Gestaltungsbedingung für die technische Umsetzung. Der Betrieb wird nicht als statistisch repräsentativer Durchschnittsfall verstanden, sondern als instruktiver Einzelfall eines mittelständischen, organisatorisch differenzierten GaLaBau-Unternehmens. Analytische Generalisierbarkeit im Sinne Flyvbjergs (2006) wird daher nur unter expliziten Transferbedingungen beansprucht. Vergleichbare Aussagen sind für Betriebe plausibel, die ein zentrales Info@-Postfach, heterogene Eingangsmails, begrenzte IT-Ressourcen, datenschutzbezogene Anforderungen und eine ähnliche Abhängigkeit von Fachsystem- und Stammdatenqualität aufweisen. Statistische Generalisierbarkeit wird bei einer Einzelfallstudie nicht angestrebt und ist auch nicht das Ziel dieser Arbeit.

3.2 Erhebungsinstrumente

Die Erhebungsinstrumente bilden unterschiedliche Phasen des Forschungs- und Entwicklungsprozesses ab. Die Vorerhebung, die Projektanbahnung und das Prozessmapping dienen dazu, Anwendungskontext, Fallzugang und betriebliche Prozesslogik zu verstehen. Mail-Tagebuch, Labelauswertung und Nachbesprechung liefern anschließend die Datengrundlage für die formative Bewertung der Praxistauglichkeit.

3.2.1 Vorerhebung: Fragebogen

Der erste empirische Erhebungsschritt schloss an die Präsentation einer frühen allgemeinen Prototypidee im Digitalisierungsausschuss an und diente als explorative Vorerhebung zur Identifikation von Bedarfslagen und Akzeptanzbarrieren in der Zielgruppe. Befragt wurden sechs GaLaBau-Betriebe aus dem Umfeld des Digitalisierungsausschusses des Landesverbands Hessen-Thüringen. Der Fragebogen umfasste Fragen zu E-Mail-Aufkommen, wahrgenommenem Zeitaufwand, Einstellung zu Automatisierung sowie Gründen für Zurückhaltung. Als Antwortformat kam eine fünfstufige Likert-Skala zum Einsatz, die auf dem von Boone und Boone (2012) beschriebenen Grundprinzip ordinaler Likert-Items basiert. Aussagen werden von 1 (stimme gar nicht zu) bis 5 (stimme voll zu) bewertet, wobei einzelne Aussagen thematisch separat ausgewertet werden, da keine zusammenführbare Gesamtdimension gemessen wird (Boone und Boone, 2012). Fragebogen und aggregierte Ergebnisse sind in Anhang A1 dokumentiert.

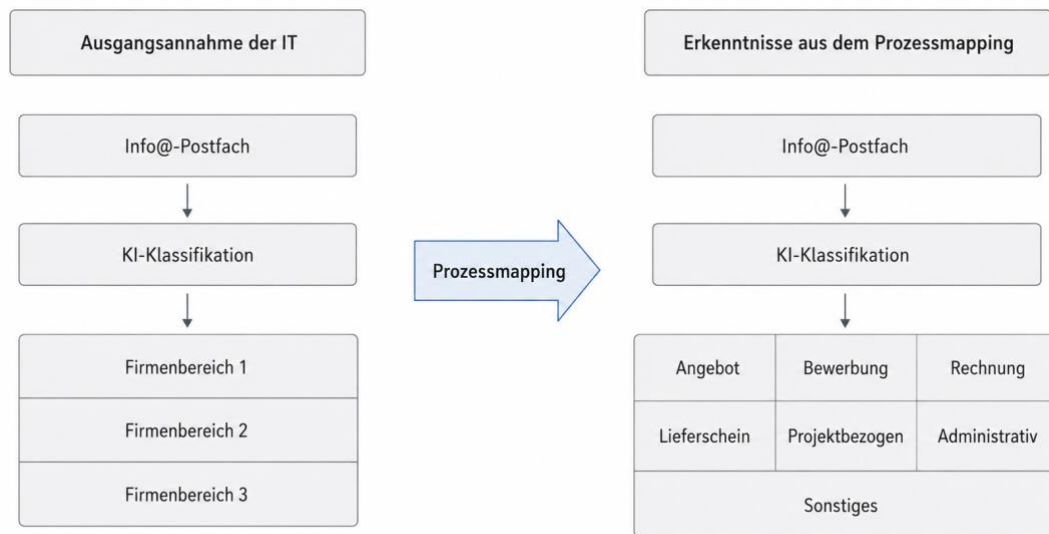
Diese Erhebung basiert auf einer Convenience-Stichprobe, da die Betriebe über ein bestehendes Verbandsnetzwerk rekrutiert wurden und keine Zufallsauswahl vorlag. $n=6$ ist für statistische Repräsentativität unzureichend, reicht jedoch für eine explorative Bedarfsanalyse aus. Ihr primärer Zweck besteht darin, die weitere fallbezogene Prototypentwicklung und die spätere Evaluation an realen Einsatzbedingungen auszurichten. Die Befunde werden in Kapitel 4.1.1 ausgeführt, wo sie als Designlegitimation dienen.

3.2.2 Prozessmapping beim Partnerunternehmen

Das Prozessmapping diente dazu, die tatsächliche Bearbeitungslogik des zentralen Info@-Postfachs vor der fallbezogenen technischen Umsetzung zu rekonstruieren. Ausgangspunkt war zunächst eine von der IT-Abteilung eingebrachte Annahme:

Eingehende E-Mails sollten im Wesentlichen einem von drei organisatorischen Firmenbereichen zugeordnet und anschließend dorthin weitergeleitet werden. Diese Ausgangsannahme war für eine erste Automationslogik plausibel, reduzierte den realen Bearbeitungsprozess jedoch auf eine einfache Routingentscheidung. Abbildung 2 stellt dar, wie diese Annahme durch das Prozessmapping erweitert wurde.

Abbildung 2: Veränderung der Prozessannahme durch das Prozessmapping



Anmerkung: Eigene Darstellung.

Im Austausch mit operativen Anwendenden aus dem Empfang zeigte sich, dass der reale Info@-Prozess deutlich mehr Fallunterscheidungen enthält als die ursprüngliche Drei-Bereiche-Logik. Neben Angeboten, Bewerbungen, Rechnungen, Lieferscheinen, projektbezogenen Nachrichten und administrativen E-Mails entstanden weitere Unterentscheidungen, etwa bei Bewerbungsarten, projektbezogenen Dokumenten, Rückmeldungen auf Preisanfragen und Fällen ohne eindeutige Zuständigkeit. Die Abbildung verdeutlicht, dass die technische Aufgabe über die reine Zuordnung zu drei Zielbereichen hinausging. Vielmehr musste ein Eingangspostfach mit unterschiedlichen Nachrichtentypen systematisch abgebildet und verarbeitet werden.

Das Prozessmapping ist methodisch relevant, weil es den Übergang von einer technischen Automationsidee zu einem betrieblich anschlussfähigen Prototyp markiert. Die sichtbar gewordene Prozesskomplexität beeinflusste die Kategorisierungslogik, die Anzahl möglicher Routing-Endpunkte und die Entscheidung, zunächst eine Labelphase vor eine

Vollautomation zu stellen. Damit bildet das Prozessmapping eine zentrale Evidenzquelle für die spätere Bewertung der Praxistauglichkeit. Das anonymisierte Ergebnisprotokoll des Prozessmappings ist in Anhang A5.3 dokumentiert.

3.2.3 Mail-Tagebuch

Das Mail-Tagebuch erfasste den Zeitaufwand für die manuelle Bearbeitung des zentralen Info@-Postfachs vor produktiver Aktivierung einer Vollautomation. An vier Arbeitstagen wurden Bearbeitungsdauer, qualitative Aufwandseinschätzung und besondere Auffälligkeiten festgehalten, die Erhebungsdaten sind in Kapitel 5.1 dargestellt und als Rohdaten in Anhang A4 dokumentiert.

Das Diary-Verfahren ermöglicht die strukturierte Erfassung alltagsnaher Arbeitsprozesse durch die betroffenen Personen selbst und ist für die Messung zeitbezogener, repetitiver Aufgaben methodisch geeignet (Ohly et al., 2010). Für Kapitel 5 dient das Mail-Tagebuch jedoch nicht als Vorher-Nachher-Nachweis einer realisierten Effizienzsteigerung, sondern als Bezugsgröße für die wirtschaftliche Verhältnismäßigkeit der Vollautomation. Eine methodische Einschränkung besteht in der Selbstaufzeichnung. Messungenauigkeiten durch subjektive Zeitschätzung sind möglich und werden im Evaluationskapitel explizit eingeordnet.

3.2.4 Labelauswertung und Nachbesprechung

Die Labelphase bildet eine eigenständige Datenerhebung zwischen technischer Prototypentwicklung und möglicher Vollautomation. In dieser Phase klassifiziert der Prototyp eingehende E-Mails und setzt Outlook-Kategorie-Labels, ohne automatische Weiterleitungen auszulösen. Die Labels wurden nicht in einer separaten Auswertungsliste erzeugt, sondern direkt im produktiven Postfach angezeigt. Aus dem laufenden Info@-Postfach wurden 57 E-Mails ausgewertet. Spam-Mails wurden nicht berücksichtigt, da sie auch im betrieblichen Alltag nicht Teil der regulären Info@-Bearbeitung sind. Für jede betrachtete E-Mail wurde geprüft, ob die Zuordnung fachlich richtig, falsch beziehungsweise uneindeutig oder mit dem bestehenden Kategorienschema nicht abbildbar war. Als fachlich richtige Zuordnung galt eine Klassifikation, wenn das gesetzte Label dem späteren fachlich richtigen Weiterleitungsziel entsprach.

Ergänzend wurde eine qualitative Nachbesprechung mit der IT-Kontaktperson des Partnerunternehmens durchgeführt. Diese Kontaktperson hatte vor der Nachbesprechung Rückmeldungen aus dem Empfang eingeholt, der den Prozess im Alltag bearbeitet. In der

Nachbesprechung wurden anschließend alle 57 ausgewerteten E-Mails gemeinsam gesichtet, nicht nur ausgewählte Fehl- oder Grenzfälle. Die qualitative Einordnung diente dazu, quantitative Klassifikationsergebnisse nicht isoliert als Modellleistung zu interpretieren, sondern im Zusammenhang mit tatsächlicher Weiterverarbeitung, Prozesslogik, Kategorienabgrenzung und Datenqualität zu bewerten. Eine unabhängige Mehrpersonenvalidierung der Referenzbewertung erfolgte nicht. Die Bewertung stellt daher eine praxisnahe, aber explorative Referenzeinordnung dar und wird in Kapitel 5 entsprechend vorsichtig interpretiert. Das Protokoll der Nachbesprechung ist in Anhang A5.4 dokumentiert.

3.3 Evaluationskonzept und Limitationen

Die Evaluation folgt einer formativen Implementierungslogik. Der Prototyp wird nicht ausschließlich danach bewertet, ob eine produktive Vollautomation aktiviert wurde, sondern danach, welche Erkenntnisse die Labelphase über die Praxistauglichkeit des Systems liefert. Praxistauglichkeit bezeichnet in dieser Arbeit die Eignung eines Prototyps, unter realen betrieblichen Bedingungen verlässlich, anschlussfähig, rechtlich dokumentiert und wirtschaftlich vertretbar in einen bestehenden Arbeitsprozess integriert zu werden. Der Begriff wird damit enger gefasst als allgemeine technische Machbarkeit, aber weiter als produktive Vollautomatisierung: Auch ein nicht vollständig automatisierter Prototyp kann praxistaugliche Teilleistungen enthalten, wenn seine Einsatzgrenzen klar bestimmt sind.

Zur Operationalisierung dieses Begriffs werden sechs Bewertungsdimensionen unterschieden. Die technische Funktionsfähigkeit prüft, ob die zentralen Workflow-Bestandteile im Prototyp grundsätzlich umgesetzt wurden. Die Klassifikationsqualität betrachtet die fachliche Richtigkeit der gesetzten Labels. Die Prozessanschlussfähigkeit bewertet, ob Kategorien, Zuständigkeiten und Routing-Endpunkte betrieblich trennscharf genug sind. Die Daten- und Systemanschlussfähigkeit bezieht sich auf die Nutzbarkeit von Projektnummern, Zuständigkeitsdaten und Mitarbeitendeninformationen für API-Abfragen. Ergänzend werden Datenschutzorientierung und wirtschaftliche Verhältnismäßigkeit berücksichtigt. Tabelle 1 ordnet diesen Dimensionen die jeweilige Datengrundlage und die daraus folgende Aussagegrenze zu.

Tabelle 1: Bewertungsdimensionen und Aussagegrenzen der Evaluation

Dimension	Datengrundlage	Aussagegrenze
Technische Funktionsfähigkeit	Prototypdokumentation	Keine Aussage über stabilen Dauerbetrieb
Klassifikationsqualität	Labelauswertung von 57 E-Mails	Keine klassische Genauigkeitsstudie, da keine unabhängige Mehrpersonvalidierung vorliegt
Prozessanschlussfähigkeit	Prozessmapping und Nachbesprechung	Keine vollständige Organisationsanalyse aller Abteilungen
Daten- und Systemanschlussfähigkeit	Geocapture-Integration und Nachbesprechung	Nur am genutzten Fachsystem und vorhandenen Datenbestand geprüft
Datenschutzorientierte Einbettung	Projektdokumentation und Prototypbeschreibung	Keine rechtsgutachterliche Prüfung der datenschutzrechtlichen Konformität
Wirtschaftliche Verhältnismäßigkeit	Mail-Tagebuch, Prozessmapping und Nachbesprechung	Kein belastbarer Return-on-Investment-Nachweis ohne Produktivpilot

Anmerkung: Eigene Darstellung.

Tabelle 1 macht sichtbar, dass die Evaluation mehrere Aussageebenen trennt. Die Arbeit bewertet technische Umsetzung, Klassifikation, Prozessanschluss, Datenqualität, Datenschutzorientierung und Wirtschaftlichkeit nicht als einheitliches Erfolgsurteil, sondern als miteinander verbundene Prüfdimensionen. Dadurch bleibt erkennbar, welche Aussagen aus der Labelphase belastbar abgeleitet werden können und welche lediglich als Transferbedingung für vergleichbare Betriebe gelten.

Die wichtigsten methodischen Grenzen ergeben sich aus der explorativen Anlage der Untersuchung. Die kleine Vorerhebungsstichprobe von $n=6$ erlaubt keine statistische Generalisierung; die analytische Übertragbarkeit auf vergleichbare KMU im GaLaBau wird in Kapitel 5 deshalb explizit mit Transferbedingungen versehen. Hinzu kommen Messunsicherheiten durch die Selbstaufzeichnung im Mail-Tagebuch, die Begrenzung der Labelauswertung auf den Kommunikationsmix des Erhebungszeitraums und die fehlende unabhängige Mehrpersonvalidierung der Referenzbewertung. Auch die qualitative Nachbesprechung wurde primär mit der IT-Kontaktperson geführt, wobei Rückmeldungen aus dem Empfang einbezogen wurden. Der Evaluationszeitraum fällt zudem in die GaLaBau-Hochsaison im Frühjahr, wodurch E-Mail-Volumen und Kommunikationsmix beeinflusst sein können. Diese Einschränkungen mindern die

Aussagekraft der Befunde nicht grundlegend, begrenzen aber die Präzision quantitativer Leistungskennzahlen.

Das gewählte Forschungsdesign rahmt die im folgenden Kapitel beschriebene Prototypentwicklung methodisch ein und liefert das Instrumentarium für die Evaluation in Kapitel 5.

4 Prototypentwicklung

Das vorliegende Kapitel beschreibt die Entwicklung des mit KI gestützten E-Mail-Automatisierungsprototyps. Im ersten Abschnitt werden Projektanbahnung, Vorerhebung und betrieblicher Ausgangspunkt dargestellt. Anschließend wird erläutert, wie sich die ursprüngliche Annahme einer einfachen Weiterleitungslogik durch das Prozessmapping beim Partnerunternehmen veränderte. Abschnitt 4.2 beschreibt den finalen Prototyp P4 in seinen technischen, logischen und datenschutzrechtlichen Dimensionen. Die Abgrenzung zu Kapitel 5 ist dabei explizit. Kapitel 4 beantwortet die Frage, was das entwickelte System ist und wie es entstand. Kapitel 5 bewertet, welche Erkenntnisse sich aus der Erprobung für die Praxistauglichkeit ableiten lassen.

4.1 Entwicklungskontext

Der Entwicklungskontext zeigt, wie aus einer allgemeinen Prototypidee ein fallbezogener Use Case beim Partnerunternehmen entstand. Dafür werden zunächst Vorerhebung und Projektanbahnung eingeordnet, bevor die iterative Entwicklung von P1 bis P4 als Lernprozess beschrieben wird. Die Projektanbahnung und die zugehörigen Gesprächsprotokolle sind in Anhang A5 dokumentiert.

4.1.1 Ausgangssituation und Vorerhebung

Die Grundidee einer KI-gestützten E-Mail-Verarbeitung entstand zunächst unabhängig vom späteren Partnerunternehmen. Ausgangspunkt war die Frage, ob wiederkehrende E-Mail-Sortieraufgaben in mehreren kleineren GaLaBau-Betrieben durch einen übertragbaren Prototyp unterstützt werden könnten. Die Idee wurde im Digitalisierungsausschuss des Fachverbands Garten-, Landschafts- und Sportplatzbau Hessen-Thüringen vorgestellt. Da das unmittelbare Interesse der Betriebe verhalten blieb, wurde anschließend eine explorative Vorerhebung durchgeführt, um Ablehnungsgründe und Einsatzbedingungen systematisch zu erfassen.

Im März 2026 wurden sechs GaLaBau-Betriebe aus dem Umfeld des Digitalisierungsausschusses des Fachverbands Hessen-Thüringen befragt, die methodische Beschreibung dieser Erhebung findet sich in Kapitel 3.2.1. Die zentralen Befunde zeigten ein latentes, jedoch kein akutes Problembewusstsein. 83 Prozent der Betriebe empfangen täglich 10 bis 30 E-Mails, 50 Prozent wenden dafür täglich ein bis zwei Stunden auf. Der wahrgenommene Belastungsgrad lag im Durchschnitt bei 2,5 auf

einer fünfstufigen Skala. Die zugrunde liegende Skala reichte von 1 als geringer Belastung bis 5 als hoher Belastung, der Mittelwert verweist daher auf moderaten Leidensdruck, nicht auf akuten Handlungsbedarf. Die Adoptionsbereitschaft war bedingt positiv. Vier von sechs Betrieben gaben an, eine Automationslösung einzusetzen, wenn eine klare Zeitersparnis nachgewiesen wird. Für die Prototypentwicklung bedeutete dies, dass der Nutzen nicht nur behauptet, sondern in der späteren Evaluation prüfbar gemacht werden musste. Zugleich mussten Einführungsaufwand und Datenschutz von Beginn an als Gestaltungsbedingungen berücksichtigt werden.

Die genannten Ablehnungsgründe lieferten ebenso bedeutsame Gestaltungshinweise. Als häufigster Hemmnisfaktor wurde der organisatorische Einführungsaufwand genannt, gefolgt von Datenschutzbedenken sowie unklarem Nutzen. Der mittlere Digitalisierungsstand von 3 von 5 bei gleichzeitig hoher Technologieoffenheit von 4,2 von 5 bestätigte, dass die Umsetzungshürde nicht in einer ablehnenden Grundhaltung, sondern in konkreten Einführungsbarrieren liegt. Parallel zur Vorerhebung und zur Suche nach einem Fallpartner entstand über das Verbandsumfeld der Kontakt zum Prozessmanager des späteren anonymisierten Partnerunternehmens. Die Projektanbahnung erfolgte zunächst über Gespräche mit diesem Prozessmanager und wurde anschließend durch ein Erstgespräch mit der IT-Abteilung konkretisiert. Der Kommunikationsweg ist in Anhang A5.1 dokumentiert.

4.1.2 Iterative Prototypentwicklung und Prozessmapping

Der finale Prototyp P4 ist das Ergebnis einer iterativen Entwicklung, in der jede Vorstufe eine konkrete Anforderung für die Folgeversion produzierte. P1 entstand zunächst als allgemeiner Ansatz zur KI-gestützten Kategorisierung eingehender E-Mails. In den Gesprächen mit dem Prozessmanager des Partnerunternehmens wurde dieser Ansatz auf die Organisationsstruktur des Unternehmens bezogen und zu P2 weiterentwickelt. Da die betriebliche Anschlussfähigkeit einer Antwortunterstützung noch nicht eindeutig war, wurde im Erstgespräch mit der IT-Abteilung ein neuer Use Case fokussiert: die automatische Weiterleitung eingehender E-Mails aus dem zentralen Info@-Postfach an drei angenommene Firmenbereiche. Tabelle 2 verdichtet diese Entwicklung nicht als reine Chronologie, sondern als Lernprozess aus Use-Case-Schärfung, technischen Anforderungen und daraus folgenden Designentscheidungen.

Tabelle 2: Prototypentwicklung

Prototyp	Zeitraum	Kernfunktion	Erkenntnisgewinn / Ablehnungsgrund
P1	Oktober 2025	Kategorisierung eingehender E-Mails nach Typ	Allgemeiner Ansatz für mehrere kleinere Betriebe, konkreter Partner-Use-Case noch nicht definiert
P2	März 2026	Erkennung von Antwortnotwendigkeit, Antwortentwurf	Aus Gesprächen mit dem Prozessmanager entstanden; betriebliche Anschlussfähigkeit als zentraler Untersuchungsfall noch nicht eindeutig
P3	April 2026	Automatische Weiterleitung an drei angenommene Zielbereiche über externen KI-Dienst	Use Case aus dem Erstgespräch mit der IT-Abteilung; Datenschutzanforderungen und reale Prozesslogik noch nicht ausreichend geklärt
P4	April/Mai 2026	Lokaler n8n-Workflow mit Ollama, erweitertem Kategorienschema, Labelphase und vorbereiteter Routinglogik	Ergebnis aus Datenschutzanforderung, Prozessmapping und Entscheidung für kontrollierte Pilotierung statt direkter Vollautomation

Anmerkung: Eigene Darstellung.

Die Tabelle zeigt zwei zentrale Entwicklungslinien. Erstens verschob sich der fachliche Fokus von einer allgemeinen E-Mail-Unterstützung hin zu einem konkreten Routingproblem im zentralen Info@-Postfach. Diese Entwicklung ist in der Projektanbahnung und im Erstgespräch mit der IT-Abteilung dokumentiert (Anhang A5.1 und A5.2). Zweitens wurde die zunächst einfache Drei-Bereiche-Logik durch das Prozessmapping mit dem Empfang deutlich erweitert. Der reale Info@-Prozess bestand nicht aus drei einfachen Zielbereichen, sondern aus unterschiedlichen Nachrichtentypen, Unterentscheidungen und potenziellen Routing-Endpunkten (Anhang A5.3). Aus der anfänglichen Dreiteilung entstand damit eine Kategorisierungslogik mit Angebot, Bewerbung, Rechnung, Lieferschein, projektbezogen, administrativ, sonstig und manuell sowie mehreren nachgelagerten Unterfällen. Der finale Prototyp P4 verbindet deshalb lokale KI-Inferenz, Labelphase, optionale Vollautomation und Anbindung an das betriebliche Fachsystem. Die technische Dokumentation des finalen Prototyps ist in den Anhängen A2 und A3 dokumentiert.

4.2 Finaler Prototyp (P4)

Der finale Prototyp P4 bündelt die Ergebnisse der vorangegangenen Entwicklungs- und Prozessklärungsschritte in einer kontrollierten Workflow-Architektur. Die folgenden Abschnitte beschreiben zunächst technische Architektur und Datenfluss, anschließend Kategorisierungslogik und Einführungsstrategie, Geocapture-Anbindung sowie Datenschutz und Dokumentation. Damit wird nachvollziehbar, wie aus der erweiterten Prozesslogik ein prototypisch umsetzbarer n8n-Workflow mit lokaler KI-Inferenz entstand.

4.2.1 Technische Architektur und Datenfluss

Der finale Prototyp P4 ist in Anlehnung an das Input-Processing-Output-Framework (IPO) aufgebaut (Lee et al., 2023). Eingehende E-Mails bilden den Input, die LLM-basierte Klassifikation durch ein lokal laufendes Sprachmodell den Processing-Schritt und die Label-Setzung beziehungsweise potenzielle Weiterleitung den Output. Tabelle 3 zeigt die dafür eingesetzten Systemkomponenten und ordnet zugleich ein, welche Datenschutzrelevanz sie für den Workflow besitzen.

Tabelle 3: Technische Komponenten des Prototyps

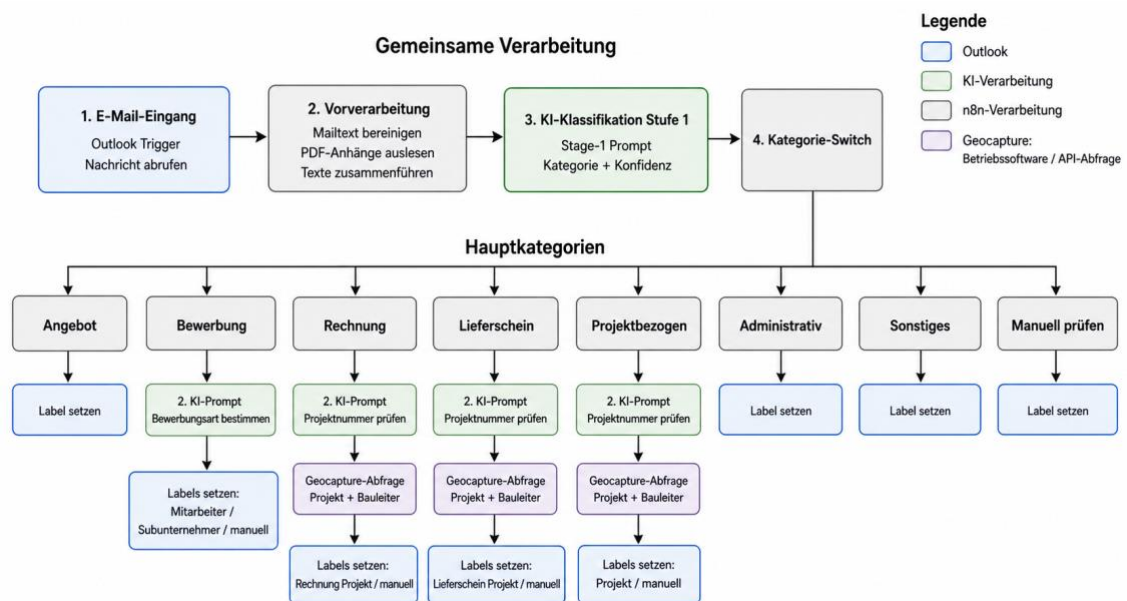
Komponente	Funktion	Datenschutz-Relevanz
Microsoft Outlook (Microsoft 365, M365)	E-Mail-Eingang, Label-Setzung, Weiterleitung	Bereits bestehende Infrastruktur des anonymisierten Partnerunternehmens
n8n (selbstgehostet auf virtuellem privatem Server, VPS)	Workflow-Orchestrierung, Routing-Logik, API-Aufrufe	Keine Datenweitergabe an Dritte
Ollama mit llama3.1:8b (lokal auf VPS)	KI-Modell-Inferenz, strukturierte JSON-Klassifikation	Vollständig lokal, kein externer KI-Dienst erforderlich
Geocapture-API (Fachsystem für Projekt- und Baustellendaten)	Zuordnung Projektnummer zu zuständiger Bauleitung	Interne API, kein Drittlandtransfer

Anmerkung: Eigene Darstellung.

Tabelle 3 verdeutlicht, dass der Prototyp nicht aus einem einzelnen KI-Modell besteht, sondern aus einer verknüpften Workflow-Architektur. Outlook bildet den betrieblichen Eingangskanal, n8n steuert die Prozesslogik, Ollama übernimmt die lokale Modellinferenz und die Geocapture-API stellt die Verbindung zu projektbezogenen Stammdaten her. Geocapture wird in dieser Arbeit als betriebliches Fachsystem des Partnerunternehmens verstanden, in dem projektbezogene Stammdaten verwaltet werden.

Für den Prototyp ist vor allem die über eine interne API mögliche Abfrage von Projektnummern, Projektzuständigkeiten und Mitarbeitendenstammdaten relevant. Die Geocapture-Anbindung dient damit nicht der KI-Klassifikation selbst, sondern der nachgelagerten Routingvorbereitung: Wenn eine E-Mail eine verwertbare Projektreferenz enthält, kann der Workflow prüfen, welche zuständige Person dem Projekt im Fachsystem zugeordnet ist. Abbildung 3 ergänzt diese Komponentenübersicht um die Prozesssicht auf den Datenfluss.

Abbildung 3: Vereinfachte Darstellung des Prototyps



Anmerkung: Eigene Darstellung.

Der Datenfluss gliedert sich in fünf Hauptschritte. Eine neu eingehende E-Mail löst in Outlook einen n8n-Trigger aus. Betreff, Absender, Nachrichtentext und vorhandene Anhänge werden geladen und in einem Code-Node bereinigt. Anschließend wird ein strukturierter Prompt an das lokal laufende Ollama-Modell übergeben. Das Modell gibt eine JSON-Antwort zurück, die neben der Hauptkategorie auch einen Konfidenzwert sowie weitere Klassifikationsfelder enthält. Auf Basis dieser Antwort setzt n8n in der Labelversion eine Outlook-Kategorie oder trifft im Vollautomationsworkflow eine Routingentscheidung. Der VPS dient dabei als Ausführungsumgebung für n8n und Ollama. Abbildung 3 verdeutlicht damit, dass die Labelphase nicht am Ende einer isolierten Modellentscheidung steht, sondern Teil eines kontrollierten Workflows mit Vorverarbeitung, Klassifikation, Sicherheitslogik und möglicher Systemanbindung ist.

Die Modellwahl fiel für die praktische Erprobung auf llama3.1:8b aus der LLaMA-Modellfamilie, ein lokal über Ollama ausführbares Open-Source-Sprachmodell. Die LLaMA-Reihe gehört zu den meistgenutzten offen verfügbaren Sprachmodellen, die als lokale Modellinferenz ohne externe API-Anbindung betrieben werden können (Schneider et al., 2024). Ausschlaggebend für die Wahl war die technische Umsetzbarkeit auf der verfügbaren VPS-Infrastruktur sowie die Möglichkeit, E-Mail-Inhalte ohne Übermittlung an externe KI-Dienste zu verarbeiten. Diese Entscheidung adressiert die in Kapitel 2.1 und 2.4 dargestellten Ressourcen- und Datenschutzerfordernissen von KMU, bringt jedoch zugleich Grenzen mit sich. Ein kleineres lokales Modell reduziert Infrastrukturaufwand und Datenschutzrisiken, kann aber bei komplexen Klassifikationsaufgaben, uneindeutigen Kategorien und langen Dokumentenkontexten an Leistungsgrenzen stoßen. Emergenzeffekte, also Fähigkeiten, die erst ab einer bestimmten Modellgröße entstehen, können bei kleineren Parameterzahlen ausbleiben (Wei et al., 2022).

4.2.2 Kategorisierungslogik und Einführungsstrategie

Die Klassifikation des Prototyps erfolgt in zwei aufeinanderfolgenden Stufen. In Stufe 1 ordnet das Sprachmodell jede eingehende E-Mail einer von acht Hauptkategorien zu: ANGEBOT, BEWERBUNG, RECHNUNG, LIEFERSCHEIN, PROJEKTBEZOGEN, ADMINISTRATIV, SONSTIGES sowie MANUELL. Die Kategorie MANUELL fungiert dabei als Sicherheitsnetz. Liegt der vom Modell ausgegebene Konfidenzwert unter dem definierten Schwellenwert von 0,80, wird die Klassifikation automatisch zu MANUELL überschrieben und die E-Mail zur menschlichen Nachbearbeitung markiert. Damit bleibt der Mensch in allen Fällen einbezogen, in denen das Modell keine hinreichend sichere Einordnung leisten kann. Der Mechanismus bildet die in Kapitel 2.4 beschriebene Kontrolllogik technisch ab, ohne bereits operative Weiterleitungsentscheidungen zu automatisieren.

Für Kategorien mit dokumentenbasiertem Kontext schließt sich eine optionale Stufe 2 an. Bei BEWERBUNG, RECHNUNG, LIEFERSCHEIN und PROJEKTBEZOGEN werden Anhänge, insbesondere PDF-Dokumente, separat analysiert. Bei Bewerbungen dient diese Analyse der Unterscheidung zwischen Bewerbungen von Mitarbeitenden und Bewerbungen durch Subunternehmen. Bei Rechnungen, Lieferscheinen und projektbezogenen Nachrichten sucht das Modell nach Projektnummern, Auftragsnummern, Baustellenadressen oder anderen Projektreferenzen. Bei

RECHNUNG und LIEFERSCHEIN gilt dabei eine Sicherheitsregel. Ohne nachgewiesene Projektnummer erfolgt grundsätzlich keine automatische Weiterleitung, um Fehlleitungen bei finanzwirksamen Dokumenten auszuschließen.

Die Einführung des Systems in den Betrieb folgt einer bewusst zweistufigen Strategie. In Phase 1 klassifiziert der n8n-Workflow eingehende E-Mails und setzt Outlook-Kategorie-Labels, ohne automatisch weiterzuleiten. Mitarbeitende sehen die KI-Einordnung, treffen die Weiterleitungsentscheidung jedoch selbst und können die Systemleistung risikofrei beobachten. In Phase 2 liegt ein Vollautomationsworkflow vor, der automatisches Routing an zuständige Personen oder Funktionspostfächer vorsieht. Die Aktivierung dieser Phase ist als nachgelagerte Go-live-Entscheidung angelegt und wird auf Grundlage der Evaluation in Kapitel 5 bewertet.

4.2.3 Geocapture-Anbindung und Routing

Die Geocapture-Anbindung erweitert den Prototyp über reine E-Mail-Klassifikation hinaus. Geocapture ist eine im Bau- und GaLaBau-Kontext eingesetzte Software zur digitalen Erfassung und Verwaltung mobiler Arbeits- und Projektdaten. Das System unterstützt unter anderem projektbezogene Zeiterfassung, Ressourcen- und Baustellenbezug sowie Schnittstellen zu weiteren Fachsystemen. Im untersuchten Fall war Geocapture für den Prototyp relevant, weil dort projektbezogene Stammdaten und Zuständigkeitsinformationen über eine Schnittstelle abgefragt werden konnten. Die Anbindung wurde daher nicht für die KI-Klassifikation selbst genutzt, sondern für die nachgelagerte Routingvorbereitung.

Wenn eine E-Mail eine Projektnummer oder Projektreferenz enthielt, sollte der Workflow über Geocapture prüfen, welche Projektleitung beziehungsweise zuständige Person dem Vorgang zugeordnet ist. Bei Rechnungen, Lieferscheinen und projektbezogenen E-Mails sollte zunächst eine Projektreferenz aus Mailtext oder Anhang erkannt werden. Anschließend ruft der Workflow über einen API-Schlüssel den passenden Projektdatensatz im betrieblichen Fachsystem ab. Aus diesem Projektdatensatz wird die ID der zuständigen Projektleitung gelesen; in einem zweiten API-Schritt wird daraus der Mitarbeitendendatensatz geladen, um die E-Mail-Adresse der zuständigen Person für das Routing zu ermitteln.

Diese Logik erweitert den Prototyp von einer reinen Inhaltsklassifikation zu einer datenbasierten Routingvorbereitung. Die KI vergibt dabei nicht nur eine Kategorie, sondern soll die E-Mail mit bestehenden betrieblichen Datenstrukturen verknüpfen.

Voraussetzung ist, dass Projektnummern eindeutig vorhanden sind, Zuständigkeiten im Fachsystem gepflegt werden und die technische Relation zwischen Projekt, bearbeitender Rolle und Mitarbeitendendatensatz für API-Abfragen nutzbar ist. Die Evaluation in Kapitel 5 bewertet, in welchem Umfang diese Voraussetzung im untersuchten Fall erfüllt war.

4.2.4 Datenschutz und Dokumentation

Das datenschutzorientierte Architekturprinzip des Prototyps besteht darin, E-Mail-Inhalte nicht an externe KI-Dienste oder cloudbasierte Modellanbieter zu übermitteln. Alle Klassifikationsschritte laufen auf dem selbstgehosteten VPS, auf dem n8n und Ollama betrieben werden. Die Entscheidung für lokale KI-Verarbeitung adressiert damit die in der Vorerhebung und im Verlauf von P3 identifizierten Datenschutzbedenken.

Als bestehende externe Infrastruktur kommt Microsoft 365 zum Einsatz, das bereits vor Einführung des Prototyps vom anonymisierten Partnerunternehmen genutzt wurde. Der Prototyp verändert den datenschutzrechtlichen Status quo in diesem Punkt nicht grundlegend. E-Mails, die zuvor von Mitarbeitenden in Outlook gesichtet und weitergeleitet wurden, durchlaufen denselben Dienst wie bisher. Zusätzlich wurden für die prototypische Umsetzung organisatorische Datenschutzfragen geklärt, darunter eine Geheimhaltungsvereinbarung, die Dokumentation der Datenflüsse, ein Auftragsverarbeitungsvertrag mit dem Hosting-Anbieter sowie Anpassungen interner Datenschutzdokumentation. Die zugehörige Datenschutzdokumentation ist in Anhang A6 gebündelt. Damit zeigt der Prototyp, dass datenschutzorientierte KI-Automation nicht allein durch lokale Modellinferenz entsteht, sondern durch das Zusammenspiel aus Architektur, Dienstleisterauswahl, Zugriffskontrolle und Dokumentation.

Die in diesem Kapitel beschriebene Prototypentwicklung zeigt, wie sich der iterative Weg von P1 zu P4 in Architektur, Kategorisierungslogik, Einführungsstrategie und Systemanbindung niederschlug. Die lokale Architektur, die zweistufige Einführungsstrategie, die konfidenzbasierte Eskalationslogik, die Anhangsanalyse und die Geocapture-Anbindung bilden die technische Grundlage der anschließenden Evaluation. Im folgenden Kapitel 5 wird bewertet, welche technischen, prozessualen und organisatorischen Voraussetzungen aus dieser Erprobung für praxistaugliche KI-Automation im Handwerk abgeleitet werden können.

5 Evaluation

Kapitel 5 bewertet den entwickelten Prototyp unter realen betrieblichen Bedingungen. Die formative Evaluation der Labelphase bildet dabei die Grundlage, um technische Funktionsfähigkeit, fachliche Anschlussfähigkeit, Datenqualität, Datenschutzdokumentation und wirtschaftliche Verhältnismäßigkeit zusammenzuführen. Praxistauglichkeit wird gemäß Kapitel 3.3 als mehrdimensionale Eignung verstanden. Ein Prototyp muss technisch funktionieren, fachlich anschlussfähig sein, mit verfügbaren Daten arbeiten können, datenschutzorientiert dokumentiert sein und in einem vertretbaren Verhältnis zum betrieblichen Nutzen stehen. Aus der Datengrundlage, der Labelauswertung und der qualitativen Nachbesprechung werden deshalb technische, prozessuale und organisatorische Voraussetzungen abgeleitet, die für praxistaugliche KI-Automation im Handwerk erfüllt sein müssen.

5.1 Datengrundlage der Evaluation

Die Evaluation stützt sich auf mehrere eigene Datenquellen, die in Kapitel 3.2 methodisch eingeführt wurden. Das Prozessmapping bildet die Grundlage für die Bewertung der Prozessanschlussfähigkeit. Das Mail-Tagebuch dokumentiert den manuellen Bearbeitungsaufwand des zentralen Info@-Postfachs und dient als Bezugsgröße für die wirtschaftliche Verhältnismäßigkeit. Die Labelphase umfasst 57 E-Mails, deren KI-Labels direkt im betrieblichen Postfach gesetzt und anschließend fachlich geprüft wurden. Ergänzend erfolgte eine qualitative Nachbesprechung mit der IT-Kontaktperson des Partnerunternehmens. Diese hatte zuvor Rückmeldungen aus dem Empfang eingeholt; in der gemeinsamen Auswertung wurden alle betrachteten E-Mails auf Label, tatsächliche Weiterverarbeitung und fachlich richtiges Weiterleitungsziel geprüft. Diese Kombination ermöglicht eine Bewertung, die über eine reine Genauigkeitskennzahl hinausgeht und die Prozess- und Datenbedingungen des betrieblichen Einsatzes einbezieht.

Das Mail-Tagebuch zeigt, dass der direkte Bearbeitungsaufwand im untersuchten Zeitraum begrenzt war. Über vier Erhebungstage wurden insgesamt 75 Minuten für die manuelle Bearbeitung des Info@-Postfachs dokumentiert. Am ersten Tag betrug der Aufwand 30 Minuten und wurde als erhöht eingeordnet, da zusätzlich eine Spam-Identifizierung durchgeführt wurde. An den drei Folgetagen lag der Aufwand jeweils bei

15 Minuten und wurde als normal bewertet. Tabelle 4 stellt diese Rohwerte dar, die zugrunde liegenden Tagebuchdaten sind in Anhang A4 dokumentiert.

Tabelle 4: Mail-Tagebuch Auswertung

Tag	Datum	Zeitaufwand	Aufwandseinschätzung	Besonderheit
Tag 1	29.04.	30 Minuten	erhöht	Spam-Identifizierung
Tag 2	30.04.	15 Minuten	normal	keine
Tag 3	04.05.	15 Minuten	normal	keine
Tag 4	05.05.	15 Minuten	normal	keine

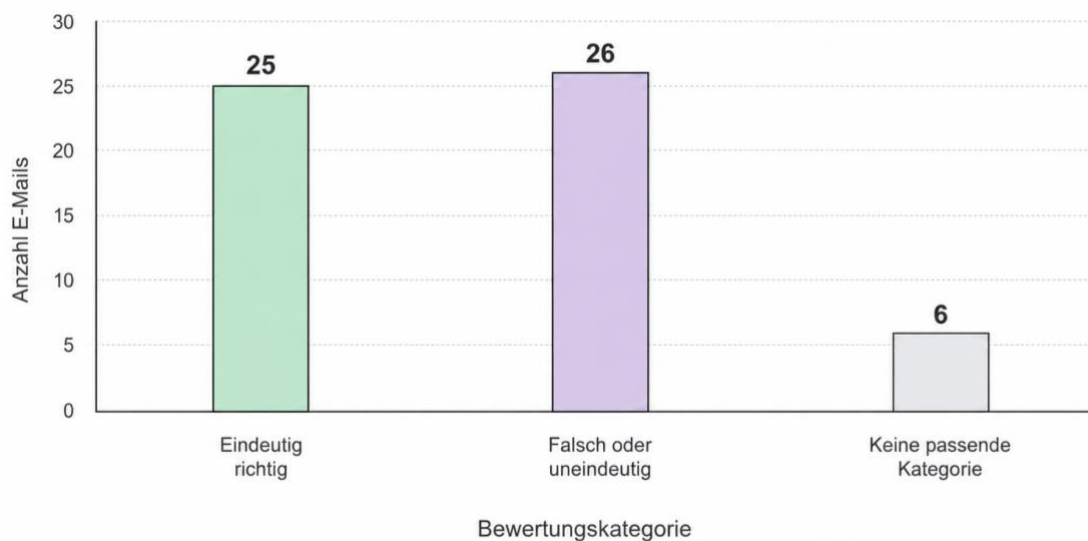
Anmerkung: Eigene Darstellung.

Aus dieser Erhebung ergibt sich ein Durchschnitt von 18,75 Minuten pro Tag. Da der erste Tag durch die zusätzliche Spam-Identifizierung beeinflusst war, bildet der reguläre Aufwand von 15 Minuten pro Tag die plausiblere Bezugsgröße für die wirtschaftliche Bewertung. Tabelle 4 misst damit keinen Effizienzgewinn des Prototyps, sondern dokumentiert den manuellen Ausgangsaufwand, gegen den eine mögliche Vollautomation zu bewerten wäre. Die reine Zeitersparnis stellt im untersuchten Fall nur ein begrenztes Nutzenargument dar. Für die Praxistauglichkeit des Prototyps müssen daher zusätzlich Durchlaufzeiten, Fehlervermeidung, Zuständigkeitsklarheit und Entlastung bei unklaren Fällen berücksichtigt werden. Diese breitere Bewertungslogik entspricht dem in Kapitel 2.4 entwickelten Kriterienrahmen.

5.2 Ergebnisse der Labelphase

Die Labelphase zeigte eine begrenzte, aber aufschlussreiche Klassifikationsleistung. Von 57 ausgewerteten E-Mails wurden 25 eindeutig korrekt klassifiziert. Weitere 26 E-Mails wurden als falsch oder uneindeutig bewertet. Sechs E-Mails konnten keiner vorhandenen Kategorie sinnvoll zugeordnet werden und bilden damit eine separate dritte Gruppe. Spam-Mails wurden aus der Auswertung ausgeschlossen, da sie auch im regulären Arbeitsalltag nicht Teil der operativen Info@-Bearbeitung sind. Maßstab der Bewertung war nicht allein die semantische Hauptkategorie, sondern das fachlich richtige spätere Weiterleitungsziel, das in der Labelphase durch die gesetzten Postfachlabels abgebildet wurde. Abbildung 4 visualisiert diese drei Bewertungsgruppen:

Abbildung 4: Ergebnis der Labelauswertung



Anmerkung: Eigene Darstellung.

Die Zahlen dürfen nicht als klassische Genauigkeitsstudie interpretiert werden. Die Gruppe falsch oder uneindeutig umfasst sowohl eindeutige Fehlklassifikationen als auch fachliche Grenzfälle. Diese Fälle wurden bewusst nicht als korrekt gezählt, weil für eine produktive Vollautomation nur eindeutig richtige Weiterleitungsziele tragfähig wären. In mehreren Fällen war die Richtung der Klassifikation nachvollziehbar, die endgültige Einordnung jedoch vom konkreten Weiterleitungsprozess abhängig. Eine nachträgliche quantitative Aufteilung der 26 falsch beziehungsweise uneindeutig bewerteten E-Mails in eindeutige Modellfehler, fachliche Grenzfälle und Prozessunklarheiten war nicht belastbar möglich, da hierfür eine systematische Mehrpersonvalidierung mit den operativ beteiligten Rollen erforderlich gewesen wäre. Die Kategorie wird deshalb konservativ zusammengeführt ausgewiesen und qualitativ interpretiert. Das Auswertungsprotokoll der Labelphase ist in Anhang A5.4 dokumentiert.

Besonders relevant waren die Kategorien MANUELL und SONSTIGES. Beide Kategorien erwiesen sich als nicht ausreichend trennscharfe Sammelbereiche. MANUELL sollte ursprünglich nur Fälle auffangen, in denen das Modell keine sichere Zuordnung treffen konnte. SONSTIGES sollte irrelevante oder nicht weiterzuleitende E-Mails bündeln. In der praktischen Auswertung fanden sich in beiden Kategorien jedoch auch Nachrichten, die fachlich einer eigenständigen Kategorie oder einem bestehenden Prozess hätten zugeordnet werden können. Dazu zählten insbesondere Werbung, Lieferanteninformationen und Rückmeldungen auf Preisanfragen. Für eine produktive

Vollautomation hätte das Kategorienschema daher weiter differenziert und fachlich validiert werden müssen. Der Befund konkretisiert die in Kapitel 2.4 beschriebene Anforderung, dass Automatisierungsprozesse stabile Kategorien sowie klare Eingangs- und Ausgangslogiken benötigen.

5.3 Prozessuale und datenbezogene Grenzen

Die wichtigste Erkenntnis der Evaluation liegt in der Abhängigkeit zwischen KI-Klassifikation und betrieblicher Prozessklarheit. Der Prototyp konnte E-Mails technisch klassifizieren und Labels setzen, doch die fachliche Entscheidung war nicht in allen Fällen eindeutig operationalisierbar. Wenn Mitarbeitende im Empfang selbst Rückfragen stellen müssen, welche Person oder Abteilung zuständig ist, kann diese Entscheidung nicht ohne Weiteres in eine stabile Routinglogik überführt werden. Der Engpass lag damit nicht ausschließlich im KI-Modell, sondern in der fehlenden formalen Entscheidungslogik des Prozesses. Damit bestätigt der Fall die in Kapitel 2.4 hergeleitete Anforderung, dass KI-Automation einen angemessenen Standardisierungsgrad benötigt, ohne notwendige fachliche Flexibilität vollständig zu verdrängen. Diese Grenze wurde bereits im Prozessmapping sichtbar und in der Labelauswertung erneut bestätigt. Die zugehörigen Ergebnisprotokolle sind in Anhang A5.3 und A5.4 dokumentiert.

Die Geocapture-Anbindung verdeutlichte diese Abhängigkeit zusätzlich. Der Workflow war technisch in der Lage, Projektnummern oder Projektreferenzen aus E-Mails und Anhängen zu extrahieren und anschließend eine API-Abfrage an das Fachsystem auszuführen. In der praktischen Erprobung zeigte sich jedoch, dass die Weiterverarbeitung von der Qualität der Stammdaten abhängt. Wenn Projektnummern nicht eindeutig gepflegt sind, wenn zu einem Vorgang mehrere Referenzen existieren oder wenn die Zuständigkeit im System nicht über den abrufbaren Mitarbeitendendatensatz verknüpft ist, kann der Workflow keine verlässliche Weiterleitung erzeugen. Die technische Schnittstelle war daher nicht der alleinige Engpass. Die Grenzen lagen vor allem in der Prozessklarheit und ergänzend in der Stammdatenqualität des angeschlossenen Fachsystems. Die in Kapitel 2.4 beschriebene Datenqualität wird im Fall damit zur konkreten Implementierungsgrenze.

Auch die Eingangsdaten aus den E-Mails selbst begrenzten die Automatisierung. Mails von Kundschaft und Lieferunternehmen enthielten nicht immer die Informationen, die für eine eindeutige Klassifikation erforderlich waren. Fehlende Projektnummern, unklare Betreffzeilen, allgemeine Formulierungen oder mehrdeutige Anhänge erschweren

sowohl menschliche als auch KI-gestützte Zuordnung. Für KI-Systeme verschärft sich dieses Problem, da sie nur auf Basis der verfügbaren Informationen und der definierten Kategorien entscheiden können. Unvollständige Eingangsdaten führen daher zu unsicheren oder unbrauchbaren Klassifikationen.

5.4 Wirtschaftliche Praxistauglichkeit

Die technische Umsetzbarkeit des Prototyps wurde grundsätzlich gezeigt. Der n8n-Workflow konnte E-Mails abrufen, bereinigen, durch ein lokal betriebenes Ollama-Modell klassifizieren, Outlook-Labels setzen, Anhänge analysieren und im Vollautomationsworkflow Geocapture-Abfragen vorbereiten. Damit war die technische Grundlage für eine KI-gestützte E-Mail-Automation vorhanden. Die produktive Aktivierung der Vollautomation wurde dennoch nicht weiterverfolgt, weil ein wirtschaftlich positiver Effekt auf Basis der verfügbaren Erhebungsdaten nicht belastbar nachgewiesen werden konnte.

Der zunächst modellierte Prozess beruhte auf der Annahme einer einfachen Verteilung an drei organisatorische Firmenbereiche. Nach dem Prozessmapping und der Labelphase ergab sich jedoch eine deutlich komplexere Struktur mit zahlreichen Kategorien, Unterkategorien und möglichen Routing-Endpunkten. Perspektivisch wären etwa 16 bis 18 Endpunkte notwendig gewesen, um den Prozess fachlich vollständig abzubilden. Dem stand ein regulärer manueller Bearbeitungsaufwand von etwa 15 Minuten pro Tag gegenüber. Unter diesen Bedingungen konnte eine Vollautomation nicht allein über direkte Zeitersparnis begründet werden. Wirtschaftlich plausibel wäre sie erst dann, wenn zusätzlich Durchlaufzeiten, Fehlerrisiken oder Koordinationsaufwand messbar reduziert werden.

Eine zusätzliche, in der Erhebung zunächst nicht systematisch erfasste Kennzahl ist die Durchlaufzeit vom Eingang der E-Mail bis zur Ankunft am richtigen Arbeitsplatz. In der Nachbesprechung zeigte sich, dass einzelne E-Mails zwar nur wenige Minuten aktive Bearbeitungszeit verursachten, durch Rückfragen und unklare Zuständigkeiten jedoch erst deutlich später abschließend zugeordnet wurden. Für zukünftige Evaluationen von E-Mail-Automation sollte deshalb nicht nur die Bearbeitungsdauer pro Nachricht gemessen werden, sondern auch die Zeit bis zur fachlich richtigen Zustellung.

Die Modellwahl beeinflusst die Praxistauglichkeit ebenfalls. Das lokal eingesetzte Modell llama3.1:8b ermöglichte eine datenschutzorientierte Verarbeitung auf begrenzter Infrastruktur. Gleichzeitig deuteten die Ergebnisse darauf hin, dass ein kleineres lokales

Modell bei komplexen Kategorien, langen Anhängen und uneindeutigen Prozessregeln an Grenzen stößt. Eine leistungsfähigere Modellvariante, besseres Prompt-Tuning oder eine Reduktion der Automatisierungsziele könnten die Ergebnisse verbessern. Aus Sicht der Praxistauglichkeit wäre insbesondere ein reduzierter Automatisierungsgrad sinnvoll. Statt alle E-Mails automatisiert zu behandeln, sollte der Workflow auf gut strukturierte, wiederkehrende und risikoarme Fälle fokussiert werden. Unklare Fälle bleiben dabei bewusst manuell.

5.5 Kriterienrahmen

Aus der Evaluation lassen sich mehrere Implementierungskriterien ableiten, die den in Kapitel 2.4 entwickelten theoretischen Rahmen auf den untersuchten Fall übertragen. Die vorangegangenen Abschnitte zeigen, dass die Praxistauglichkeit KI-gestützter E-Mail-Automation nicht allein von der technischen Funktionsfähigkeit des Workflows abhängt. Entscheidend ist vielmehr das Zusammenspiel aus klaren Zuständigkeiten, trennscharfen Kategorien, nutzbaren Stammdaten, datenschutzorientierter Einbettung und wirtschaftlicher Verhältnismäßigkeit. Tabelle 5 verdichtet diese Befunde zu einem fallbezogenen Prüfraster für vergleichbare Automationsvorhaben im Handwerk.

Tabelle 5: Implementierungskriterien für KI-gestützte E-Mail-Automation

Kriterium	Konsequenz für KI-gestützte Automation
Prozessklarheit	Vor der technischen Umsetzung müssen Eingangsfälle, Entscheidungsregeln und Weiterleitungsziele eindeutig beschrieben werden.
Kategorienschema	Kategorien müssen vollständig, trennscharf und regelmäßig anhand realer E-Mails validiert werden.
Datenqualität	Projektnummern, Zuständigkeiten und Mitarbeitendenstammdaten müssen gepflegt und eindeutig über das Fachsystem abrufbar sein.
Automatisierungsgrad	Der Workflow sollte zunächst auf häufige, stabile und risikoarme Fälle begrenzt werden; unklare Fälle bleiben bewusst manuell.
Modell- und Infrastrukturwahl	Modellgröße, Prompt-Qualität, Datenschutzanforderungen und verfügbare Infrastruktur müssen gemeinsam bewertet werden.
Datenschutzorientierung	Datenflüsse, lokale Verarbeitung, Auftragsverarbeitung und Geheimhaltung müssen vor produktiver Nutzung dokumentiert sein.
Pilotierung	Ein produktiver Einsatz sollte erst nach kontrollierter Labelphase, Fehleranalyse und Anpassung des Kategorienschemas erfolgen.

Kriterium	Konsequenz für KI-gestützte Automation
Wirtschaftliche Verhältnismäßigkeit	Eine Vollautomation muss über reine Bearbeitungszeit hinaus mit Durchlaufzeit, Fehlervermeidung oder Koordinationsentlastung begründet werden.

Anmerkung: Eigene Darstellung.

Tabelle 5 ist nicht als universelles Reifegradmodell zu verstehen, sondern als fallbezogene Verdichtung der Evaluationsergebnisse. Die Kriterien greifen ineinander: Prozessklarheit beeinflusst die Qualität des Kategorienschemas, das Kategorienschema bestimmt den sinnvollen Automatisierungsgrad, und Datenqualität entscheidet darüber, ob ein technisch vorbereiteter Workflow betrieblich nutzbar wird. Die Evaluation zeigt damit, dass der Prototyp nicht an der grundsätzlichen technischen Machbarkeit scheiterte. Vielmehr machte die Labelphase sichtbar, welche Voraussetzungen vor einer produktiven KI-Automation erfüllt sein müssen. Für den untersuchten Fall sind vor allem Prozessklarheit, Kategorienqualität, Stammdatenpflege, ein realistischer Automatisierungsumfang und eine datenschutzorientierte Einbettung ausschlaggebend.

6 Fazit

Das Fazit fasst die zentralen Ergebnisse der Fallstudie zusammen, beantwortet die Forschungsfrage und ordnet die Grenzen sowie den weiteren Forschungsbedarf ein.

6.1 Beantwortung der Forschungsfrage

Die vorliegende Arbeit untersuchte die prototypische Implementierung einer KI-gestützten E-Mail-Automation im GaLaBau-Handwerk. Ausgangspunkt war die Frage, ob ein wiederkehrender manueller Sortier- und Weiterleitungsprozess durch ein lokal betriebenes KI-System technisch unterstützt und perspektivisch automatisiert werden kann. Im Verlauf der Fallstudie zeigte sich jedoch, dass die Praxistauglichkeit eines solchen Systems nicht allein durch die technische Funktionsfähigkeit bestimmt wird. Entscheidend ist das Zusammenspiel aus technischer Architektur, Prozessklarheit, Datenqualität, Datenschutz und wirtschaftlicher Verhältnismäßigkeit.

Die Forschungsfrage kann daher differenziert beantwortet werden. Eine KI-gestützte E-Mail-Automation lässt sich im GaLaBau-Handwerk prototypisch implementieren, wenn bestehende E-Mail-Infrastrukturen, Workflow-Orchestrierung und lokale Sprachmodell-Inferenz miteinander verbunden werden. Der entwickelte Prototyp zeigte diese technische Machbarkeit. Eingehende E-Mails konnten aus Microsoft Outlook abgerufen, durch einen n8n-Workflow verarbeitet, mit dem lokal betriebenen Ollama-Modell llama3.1:8b klassifiziert und mit Outlook-Labels versehen werden. Darüber hinaus wurden Anhänge analysiert, Projektreferenzen gesucht und eine Geocapture-Anbindung für potenzielles Routing an zuständige Personen vorbereitet. Damit wurde gezeigt, dass KI-gestützte Klassifikation und workflowbasierte Prozessautomation auch unter datenschutzorientierten Infrastrukturbedingungen grundsätzlich umsetzbar sind.

Die Evaluation zeigte zugleich, dass technische Umsetzbarkeit nicht mit betrieblicher Praxistauglichkeit gleichgesetzt werden darf. Praxistauglichkeit wurde in dieser Arbeit als mehrdimensionale Eignung verstanden, die technische Funktionsfähigkeit, Prozessanschlussfähigkeit, Datenqualität, Datenschutzdokumentation und wirtschaftliche Verhältnismäßigkeit verbindet. In der Labelphase wurden 57 E-Mails ausgewertet. Davon wurden 25 E-Mails eindeutig korrekt klassifiziert, 26 E-Mails falsch oder uneindeutig zugeordnet und sechs E-Mails konnten keiner bestehenden Kategorie sinnvoll zugeordnet werden. Diese Ergebnisse verweisen nicht ausschließlich auf die Leistungsgrenzen des verwendeten Sprachmodells. Vielmehr wurde sichtbar, dass die

zugrunde liegende Kategorisierungslogik noch nicht ausreichend trennscharf war. Besonders die Kategorien MANUELL und SONSTIGES erwiesen sich als Sammelbereiche, in denen auch E-Mails landeten, die fachlich einer eigenständigen Kategorie oder einem bestehenden Prozess hätten zugeordnet werden können.

Der zentrale Erkenntnisgewinn liegt deshalb in der prozessualen Dimension der Implementierung. Die zunächst angenommene Weiterleitung an drei organisatorische Firmenbereiche erwies sich im Prozessmapping als zu stark vereinfacht. Der reale Info@-Prozess enthielt deutlich mehr Kategorien, Unterentscheidungen und potenzielle Routing-Endpunkte, als die ursprüngliche Drei-Bereiche-Logik abbilden konnte. Zusätzlich zeigte die qualitative Nachbesprechung, dass selbst die operativ beteiligten Personen nicht in allen Fällen einer einheitlichen Entscheidungslogik folgten. Wenn Zuständigkeiten im Arbeitsalltag situativ durch Rückfragen geklärt werden müssen, kann eine KI diese Logik nicht zuverlässig automatisieren. Eine Automatisierung setzt daher voraus, dass der Prozess vorab fachlich geklärt und in eindeutige Entscheidungsregeln überführt wird.

Eine weitere Voraussetzung betrifft die Datenqualität angeschlossener Fachsysteme. Die Geocapture-Integration war technisch möglich, hing aber von der Pflege der Stammdaten ab. Projektnummern, Projektzuständigkeiten und Mitarbeitendenstammdaten müssen eindeutig und maschinenlesbar miteinander verknüpft sein, damit ein Workflow automatisiert weiterleiten kann. Die Fallstudie zeigte, dass eine funktionierende Schnittstelle allein keinen Nutzen erzeugt, wenn die zugrunde liegenden Daten keine verlässliche Entscheidung ermöglichen. Für KI-Automation im Handwerk bedeutet dies, dass Fachsysteme und Stammdatenpflege nicht als Randbedingung behandelt werden dürfen, sondern Teil der Automatisierungsvoraussetzungen sind.

Auch die wirtschaftliche Bewertung fiel differenziert aus. Das Mail-Tagebuch ergab einen regulären manuellen Bearbeitungsaufwand von etwa 15 Minuten pro Tag. Gleichzeitig hätte eine fachlich vollständige Vollautomation perspektivisch 16 bis 18 Routing-Endpunkte sowie weitere Kategorien wie Werbung oder Antwort auf Preisanfrage benötigt. Unter diesen Bedingungen konnte ein wirtschaftlich positiver Effekt nicht belastbar nachgewiesen werden. Der Nutzen einer KI-Automation sollte deshalb nicht nur über aktive Bearbeitungszeit gemessen werden. Relevanter können Durchlaufzeit, Fehlervermeidung und Entlastung bei unklaren Zuständigkeiten sein. Diese Kennzahlen wurden in der vorliegenden Arbeit sichtbar, aber noch nicht systematisch gemessen.

Aus der Fallstudie lässt sich ableiten, dass praxistaugliche KI-Automation im Handwerk mehrere Voraussetzungen erfüllen muss. Zunächst braucht es ein klares Prozessmapping, in dem Eingangsfälle, Entscheidungen und Verantwortlichkeiten abgegrenzt werden. Darauf aufbauend muss ein trennscharfes Kategorienschema entstehen, das Rest- und Prüfkategorien nicht als Auffangbecken für ungelöste Prozesslogik nutzt. Zudem müssen Stammdaten und Fachsysteme so gepflegt sein, dass automatisierte Abfragen zuverlässige Ergebnisse liefern. Der Automatisierungsgrad sollte realistisch begrenzt werden. Nicht alle E-Mails müssen automatisiert werden, oft ist ein fokussierter Teilautomationsansatz für häufige, risikoarme und gut strukturierte Fälle sinnvoller. Datenschutz, Auftragsverarbeitung, Zugriffskonzepte und Dokumentation müssen bereits in der Entwicklung berücksichtigt werden. Schließlich sollte jede Vollautomation durch eine Label- oder Pilotphase vorbereitet werden, in der Fehlerbilder, Grenzfälle und Go-live-Kriterien systematisch beobachtet werden.

Die Arbeit leistet damit einen Beitrag an der Schnittstelle von KI, Prozessautomatisierung und Digitalisierung im Handwerk. Ihr Ergebnis ist kein pauschaler Erfolgsnachweis einer vollautomatischen E-Mail-Bearbeitung, sondern ein differenzierter Implementierungsbefund. Der Prototyp zeigte, dass die technische Grundlage vorhanden ist, während die Grenzen vor allem in Prozessstruktur, Datenqualität und organisatorischer Einbettung liegen. Gerade dieser Befund ist für Handwerksbetriebe relevant, weil er zeigt, dass KI-Projekte nicht mit der Auswahl eines Modells beginnen sollten, sondern mit der Klärung der zu automatisierenden Entscheidung.

6.2 Limitationen

Die Aussagekraft der Arbeit ist durch den Einzelfallcharakter begrenzt. Die Vorerhebung umfasste sechs Betriebe, die Labelauswertung 57 E-Mails und die qualitative Nachbesprechung erfolgte primär über die IT-Kontaktperson des Partnerunternehmens. Rückmeldungen aus dem Empfang wurden einbezogen, eine unabhängige Mehrpersonvalidierung der Referenzbewertung fand jedoch nicht statt. Die Ergebnisse sind daher nicht statistisch generalisierbar.

Weitere Limitationen ergeben sich aus dem Evaluationsdesign. Die Vollautomation wurde nicht produktiv aktiviert, sodass keine Nachhermessung unter laufenden Betriebsbedingungen durchgeführt werden konnte. Das Mail-Tagebuch erfasste den direkten Bearbeitungsaufwand, jedoch keine systematische Durchlaufzeit bis zur richtigen Zustellung. Auch die Modellbewertung bleibt auf das lokal eingesetzte

llama3.1:8b, den konkreten Kommunikationsmix des Erhebungszeitraums und das vorhandene Kategorienschema beschränkt. Eine vorsichtige analytische Übertragung ist dennoch für vergleichbare Handwerksbetriebe plausibel, wenn diese ein zentrales Info@-Postfach, heterogene Eingangsmails, begrenzte IT-Ressourcen, datenschutzbezogene Anforderungen und ähnliche Abhängigkeiten von Prozess- und Stammdatenqualität aufweisen.

6.3 Ausblick und weiterer Forschungsbedarf

Für weiterführende Forschung und Praxis ergeben sich mehrere Ansatzpunkte. Künftige Evaluationen sollten nicht nur aktive Bearbeitungszeit, sondern auch die Durchlaufzeit vom Eingang der E-Mail bis zur Ankunft am richtigen Arbeitsplatz messen. Zudem sollte untersucht werden, wie sich größere oder spezialisiertere Sprachmodelle auf die Klassifikationsqualität auswirken, wenn Datenschutzanforderungen und Betriebskosten zugleich berücksichtigt werden. Besonders relevant erscheint ein reduzierter Produktivpilot, der erst nach validiertem Prozessmapping und stabilisiertem Kategorienschema klar strukturierte Kategorien automatisiert und alle unsicheren Fälle manuell belässt. Unter diesen Bedingungen könnte der entwickelte Prototyp weiterentwickelt werden, nicht als Vollautomation aller Eingangsmails, sondern als kontrollierte Teilautomation, die dort unterstützt, wo Prozesse, Daten und Verantwortlichkeiten bereits ausreichend geklärt sind.

Literaturverzeichnis

- Afflerbach, P., Bolsinger, M., & Röglinger, M. (2016). An economic decision model for determining the appropriate level of business process standardization. *Business Research*, 9(2), 335–375. <https://doi.org/10.1007/s40685-016-0035-6>
- Alsuhaymi, D. (2025). A survey of large language models: Evolution, architectures, and applications in natural language processing. *Electronics*, 14(18), Article 3580. <https://doi.org/10.3390/electronics14183580>
- Ardito, L., Filieri, R., Raguseo, E., & Vitari, C. (2025). Artificial intelligence adoption and revenue growth in European SMEs: Synergies with IoT and big data analytics. *Internet Research*, 35(4), 1508–1534. <https://doi.org/10.1108/INTR-02-2024-0195>
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
- Arroyabe, J. C. F., Arroyabe, M. F., Fernández de Arroyabe, C. I., & Arroyabe, I. L. (2024). AI adoption in European SMEs: The role of digital capabilities, strategic orientation and sector. *Technology in Society*, 79, Article 102733. <https://doi.org/10.1016/j.techsoc.2024.102733>
- Benk, M., Kerstan, S., von Wangenheim, F., & Ferrario, A. (2025). Multidimensional trustworthiness of AI systems: A 24-year bibliometric analysis. *AI & Society*, 40, 2083–2101. <https://doi.org/10.1007/s00146-024-02059-y>
- Beuchel, S., Jantos, L., & Meub, L. (2024). Digitalisierung im Handwerk zwischen Zettelwirtschaft und KI: Eine Status Quo-Analyse (ifh Forschungsbericht 23). Volkswirtschaftliches Institut für Mittelstand und Handwerk an der Universität Göttingen. <https://doi.org/10.47952/gro-publ-220>

- Boone, H. N., & Boone, D. A. (2012). Analyzing Likert data. *Journal of Extension*, 50(2), Article 48. <https://tigerprints.clemson.edu/joe/vol50/iss2/48>
- Brasse, J., Broder, H. R., Förster, M., Klier, M., & Sigler, I. (2023). Explainable artificial intelligence in information systems: A review of the status quo and future research directions. *Electronic Markets*, 33, Article 26. <https://doi.org/10.1007/s12525-023-00644-5>
- Farinha, C., Cardoso, J. L., & Barbosa, M. (2024). Criteria and methods for prioritization and selection of RPA automation candidates: A Delphi study. *Information Systems and e-Business Management*, 22, 153–174. <https://doi.org/10.1007/s10257-023-00645-z>
- Feuerriegel, S., Hartmann, J., Janiesch, C., & Scholten, P. (2023). Generative AI. *Business & Information Systems Engineering*, 66(1), 111–126. <https://doi.org/10.1007/s12599-023-00834-7>
- Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2), 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Handler, M., Handler, N., & Fettke, P. (2024). Large language models as decision support systems in business: Applications and implications. *International Journal of Information Management*, 79, Article 102811. <https://doi.org/10.1016/j.ijinfomgt.2024.102811>
- Held, P., Heubeck, T., & Meckl, R. (2026). Boosting SMEs' digital transformation: The role of dynamic capabilities in cultivating digital leadership and digital culture. *Review of Managerial Science*, 20, 1687–1715. <https://doi.org/10.1007/s11846-025-00919-5>

- Helms, M., Bosbach, J., Finster, R., Friesecke, M., Beuchel, S., Bartelt-Urich, K., Meub, L., Thiem, S., Pulanco Umel, A., Jantos, L., Berger, H., Robra-Bissantz, S., & Lattemann, C. (2024). Handwerk – aus Tradition innovativ: Der Mensch im Mittelpunkt der Digitalisierung. *HMD Praxis der Wirtschaftsinformatik*, 61, 1380–1391. <https://doi.org/10.1365/s40702-024-01123-x>
- Herm, L. V., Janiesch, C., Helm, A., Imgrund, F., Hofmann, A., & Winkelmann, A. (2023). A framework for implementing robotic process automation projects. *Information Systems and e-Business Management*, 21(3), 1–36. <https://doi.org/10.1007/s10257-022-00553-8>
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105. <https://doi.org/10.2307/25148625>
- Ibrahim, M., Caschera, M. C., & Ferri, F. (2024). Artificial intelligence and productivity: A literature review on the impact of generative AI in knowledge work. *Frontiers in Artificial Intelligence*, 7, Article 1496518. <https://doi.org/10.3389/frai.2024.1496518>
- Kallmünzer, A., Mikhaylov, A., Chelaru, M., & Czakon, W. (2025). Adoption and performance outcome of digitalization in small and medium-sized enterprises. *Review of Managerial Science*, 19, 2011–2038. <https://doi.org/10.1007/s11846-024-00744-2>
- Kern, M., Ohly, S., Ďuranová, L., & Friedrichs, J. (2024). Drowning in emails: Investigating email classes and work stressors. *Frontiers in Psychology*, 15, Article 1439070. <https://doi.org/10.3389/fpsyg.2024.1439070>
- Khare, A., Singh, S., Mishra, R., Prakash, S., & Dixit, P. (2022). E-Mail assistant – Automation of e-mail handling and management using RPA. In *Proceedings of the 2022 International Conference on Decision Aid Sciences and Applications*

(DASA) (S. 1386–1390). IEEE.
<https://doi.org/10.1109/DASA54658.2022.9765017>

Kocak, S., & Pawlowski, J. (2022). Digitalization breakers: A mixed method for finding the concrete barriers in the craft sector. In PACIS 2022 Proceedings. Pacific Asia Conference on Information Systems (PACIS). AIS Electronic Library.
<https://aisel.aisnet.org/pacis2022/>

Kocak, S., & Pawlowski, J. (2023). Digital transformation in the craft sector: Identifying digital competencies and attitudes for managers and employees. In Proceedings of the 2023 Computers and People Research Conference (SIGMIS-CPR '23) (S. 1–6). ACM. <https://doi.org/10.1145/3579168.3632720>

Komus, A., Selsam, C., Schwarzkopf, K., Ukena, R., Krause, C., Heibel, L., Gohl, A., Loch, M., Pietrusziak, K., Bentzinger, L., Wilke, L., Diederichs, T., Dilly, M., Ergart, A., Gravelius, J., Langenbach, P., Uhrig, J., Rusche, E., & Pietsch, P. (2024). KI in Unternehmen: Strategien, Trends & Herausforderungen. Hochschule Koblenz.
<https://handwerkdigital.de/deulocal/textbilder/images/Seiten/ki-studie-2023/KI%20in%20Unternehmen%20-%20Strategien%2C%20Trends%20und%20Herausforderungen.pdf>

Lanctot, A., & Duxbury, L. (2025). You've got mail – whether you want it or not: An emic investigation into how email use can be managed. Computers in Human Behavior Reports, 18, Article 100618.
<https://doi.org/10.1016/j.chbr.2025.100618>

Lee, M. C. M., Scheepers, H., Lui, A. K. H., & Ngai, E. W. T. (2023). The implementation of artificial intelligence in organizations: A systematic literature review. Information & Management, 60(5), Article 103816.
<https://doi.org/10.1016/j.im.2023.103816>

Mark, G., Iqbal, S. T., Czerwinski, M., Johns, P., Sano, A., & Lutchyn, Y. (2016). Email duration, batching and self-interruption: Patterns of email use on productivity and

- stress. In Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16) (S. 1717–1728). ACM. <https://doi.org/10.1145/2858036.2858262>
- Martin, G., Fuhrer, R., & Poussing, N. (2025). The burden of email: Analyzing the relationship between email volume, well-being, and productivity. *Information & Management*, 62(1), Article 104089. <https://doi.org/10.1016/j.im.2024.104089>
- Merhi, M. I. (2022). Towards a framework for big data and analytics governance. *International Journal of Information Management*, 69, Article 102545. <https://doi.org/10.1016/j.ijinfomgt.2022.102545>
- OECD. (2025). AI adoption by small and medium-sized enterprises (OECD Discussion Paper for the G7). Organisation for Economic Co-operation and Development. <https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/sme-digitalisation/ai-adoption-by-smes.pdf>
- Ohly, S., Sonnentag, S., Niessen, C., & Zapf, D. (2010). Diary studies in organizational research: An introduction and some practical recommendations. *Journal of Personnel Psychology*, 9(2), 79–93. <https://doi.org/10.1027/1866-5888/a000009>
- Prívará, A., Rahman, A., Ognjanović, I., Hasnaoui, A., & Hachard, V. (2025). From survival to growth: Digital strategies for micro and small enterprises. *Strategic Change*, 34, 1–11. <https://doi.org/10.1002/jsc.2666>
- Reijers, H. A., & Liman Mansar, S. (2005). Best practices in business process redesign: An overview and qualitative evaluation of successful redesign heuristics. *Omega*, 33(4), 283–306. <https://doi.org/10.1016/j.omega.2004.04.012>
- Riedl, R. (2022). Is trust in artificial intelligence systems related to user personality? Review of empirical evidence and future research directions. *Electronic Markets*, 32, 2021–2051. <https://doi.org/10.1007/s12525-022-00594-4>

- Robertson, J., Botha, E., Oosthuizen, K., & Montecchi, M. (2025). Artificial intelligence implementation in organizations: A systematic review of critical success factors. *Journal of Business Research*, 189, Article 115198. <https://doi.org/10.1016/j.jbusres.2025.115198>
- Romero, H. L., Dijkman, R. M., Grefen, P. W. P. J., & van Weele, A. J. (2015). Factors that determine the extent of business process standardization and the subsequent effect on business performance. *Business & Information Systems Engineering*, 57(4), 261–275. <https://doi.org/10.1007/s12599-015-0386-0>
- Russell, E., Jackson, T. W., Fullman, M., & Chamakiotis, P. (2024). Getting on top of work-email: A systematic review of 25 years of research to understand effective work-email activity. *Journal of Occupational and Organizational Psychology*, 97(1), 74–103. <https://doi.org/10.1111/joop.12462>
- Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4. Aufl.). Pearson.
- Sagala, G. H., & Öri, D. (2024). Toward SMEs digital transformation success: A systematic literature review. *Information Systems and e-Business Management*, 22, 667–706. <https://doi.org/10.1007/s10257-024-00682-2>
- Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. *SN Computer Science*, 2, Article 420. <https://doi.org/10.1007/s42979-021-00815-1>
- Sartor, G., & Lagioia, F. (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence (STOA Study PE 641.530). *Europäisches Parlament*. <https://doi.org/10.2861/293>
- Schallmo, D., Williams, C. A., & Boardman, L. (2017). Digital transformation of business models – Best practice, enablers, and roadmap. *International Journal of Innovation Management*, 21(8), Article 1740014. <https://doi.org/10.1142/S136391961740014X>

- Schneider, J., Meske, C., & Kuss, P. (2024). Foundation models: A new paradigm for artificial intelligence. *Business & Information Systems Engineering*, 66(3), 367–379. <https://doi.org/10.1007/s12599-024-00851-0>
- Schwaeke, J., Peters, A., Kanbach, D. K., Kraus, S., & Jones, P. (2024). The new normal: The status quo of AI adoption in SMEs. *Journal of Small Business Management*, 63(3), 1297–1331. <https://doi.org/10.1080/00472778.2024.2379999>
- Seppänen, M., Haapanen, L., & Kunttu, I. (2025). Digital transformation of SMEs: Skill gaps, change resistance, and strategic leadership. *Digital Business*, 5, Article 100106. <https://doi.org/10.1016/j.digbus.2025.100106>
- Skare, M., de Obesso, M. M., & Ribeiro-Navarrete, S. (2023). Digital transformation and European small and medium enterprises: A comparative study using digital economy and society index data. *International Journal of Information Management*, 68, Article 102594. <https://doi.org/10.1016/j.ijinfomgt.2022.102594>
- Sundberg, L., & Holmström, J. (2024). Artificial intelligence in public sector digital transformation: The case of NLP-based email management. *Journal of Strategic Information Systems*, 33(4), Article 101848. <https://doi.org/10.1016/j.jsis.2024.101848>
- Vial, G. (2019). Understanding digital transformation: A review and a research agenda. *Journal of Strategic Information Systems*, 28(2), 118–144. <https://doi.org/10.1016/j.jsis.2019.01.003>
- Wang, L., Hosseini, R., Awadallah, A. H., Bennett, P. N., & Quirk, C. (2019). Context-aware intent identification in email conversations. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (S. 1075–1084). ACM. <https://doi.org/10.1145/3331184.3331284>

- Wang, R. Y., & Strong, D. M. (1996). Beyond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, 12(4), 5–33.
<https://doi.org/10.1080/07421222.1996.11518099>
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., Yogatama, D., Bosma, M., Zhou, D., Metzler, D., Chi, E. H., Hashimoto, T., Vinyals, O., Liang, P., Dean, J., & Fedus, W. (2022). Emergent abilities of large language models. *Transactions on Machine Learning Research*.
<https://openreview.net/forum?id=yzkSU5zdwD>
- Wei, W., & Pardo, C. (2022). Implementing AI in SMEs: Key challenges and pathways to success. *Industrial Marketing Management*, 107, 384–397.
<https://doi.org/10.1016/j.indmarman.2022.10.008>
- Yin, R. K. (2014). *Case study research: Design and methods* (5. Aufl.). Sage.
- Zacher, H., & Frese, M. (2018). Action regulation theory: Foundations, current knowledge, and future directions. In D. S. Ones, N. Anderson, C. Viswesvaran, & H. K. Sinangil (Hrsg.), *The SAGE Handbook of Industrial, Work and Organizational Psychology* (Bd. 2, S. 32–49). Sage.

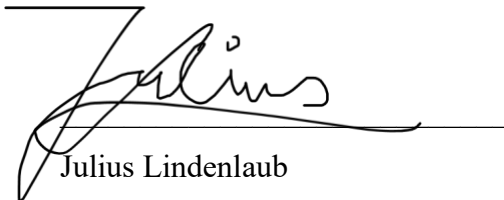
Eidesstattliche Erklärung

EIDESSTATTLICHE ERKLÄRUNG

Ich, Julius Lindenlaub, erkläre hiermit an Eides statt, dass ich die vorliegende Bachelorarbeit selbständig angefertigt habe. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht.

Ich habe nur die erlaubten und dokumentierten Hilfsmittel benutzt und den Einsatz generativer KI-Systeme, wie vom Studiengang gefordert, offengelegt. Ich verantworte die Auswahl, Übernahme und sämtliche Ergebnisse des von mir verwendeten KI-generierten Outputs vollumfänglich selbst.

Die Arbeit wurde bisher weder in gleicher noch in ähnlicher Form einer anderen Prüfungsbehörde vorgelegt und auch noch nicht veröffentlicht.



Julius Lindenlaub

Erfurt, 07 / Juni / 2026

Transparenzerklärung

Ich, Julius Lindenlaub, erkläre hiermit, dass die vorliegende Arbeit selbstständig angefertigt wurde. Die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche, gemäß der am Studiengang bestehenden Richtlinien, kenntlich gemacht.

Dokumentation der Verwendung von KI-gestützten Tools

Im Rahmen der Arbeit wurden folgende KI-gestützte Tools für folgende Zwecke verwendet.

KI-gestütztes Tool	Art der Unterstützung	Beitrag zur Arbeit
ChatGPT und Claude	Unterstützung in der frühen Ideenfindung, Strukturierung und Erörterung möglicher Themenrichtungen, Forschungszugänge und Umsetzungsvarianten.	Die Tools dienten als Reflexions- und Strukturierungshilfe in der Orientierungsphase. Die Auswahl des Themas, die Eingrenzung der Forschungsfrage und die inhaltlichen Entscheidungen wurden eigenständig getroffen.
Claude Code und Codex	Unterstützung bei der technischen Entwicklung der Prototypen, insbesondere bei Code-Entwürfen, Fehlersuche, Anpassung von Workflows und Umsetzung einzelner Funktionen.	Die Tools unterstützten die praktische Programmierarbeit. Der Prototyp wurde nicht automatisiert erstellt, sondern im Rahmen eigener Entwicklungsentscheidungen geprüft, angepasst und in den Untersuchungskontext eingeordnet.
Ollama mit llama3.1:8b	Einsatz als lokal betriebenes Sprachmodell im entwickelten Prototyp zur KI-gestützten Klassifikation eingehender E-Mails.	Das Modell war Bestandteil des Untersuchungsgegenstands. Seine Ausgaben wurden im Rahmen der Labelphase ausgewertet und dienten als Grundlage für die Evaluation der Praxistauglichkeit.
ImageGen	Unterstützung bei der Erstellung beziehungsweise grafischen Aufbereitung einzelner Abbildungen für die Arbeit.	Die erzeugten Grafiken wurden für die visuelle Darstellung von Forschungsdesign, Workflow oder Evaluationsergebnissen genutzt und inhaltlich an den tatsächlichen Aufbau der Arbeit angepasst.
ChatGPT und Claude	Unterstützung bei Strukturierung, Formulierungsfeedback, sprachlicher Überarbeitung und Konsistenzprüfung einzelner Textabschnitte. Außerdem wurden diese Tools als Hilfen bei Zitation verwendet	Die Tools dienten als Feedback- und Überarbeitungshilfe. Die Textinhalte wurden eigenständig verfasst; Hinweise der Tools wurden geprüft und nur bei fachlicher Passung berücksichtigt. Durch die Hilfe bei Zitation, war kein zusätzliches Programm notwendig

Sämtliche Inhalte, die mithilfe von KI-gestützten Tools generiert und als direkte oder indirekte Zitate in die Arbeit übernommen wurden, sind gemäß der vom Studiengang vorgegebenen Zitierweise klar als solche gekennzeichnet.

Anhang

Anhangsverzeichnis

Anhang	Seite	Inhalt
A1	ii	Fragebogen und Ergebnisse der Vorerhebung
A2	iv	n8n-Workflow-Konfiguration
A3	v	Workflow-Screenshot und zentrale Code- beziehungsweise Prompt- Ausschnitte
A4	ix	Mail-Tagebuch-Rohdaten
A5	x	Protokolle und Projektanbahnung mit dem Partnerunternehmen
A6	xviii	Datenschutzdokumentation

A1 Fragebogen und Ergebnisse der Vorerhebung

Anhang A1 enthält den eingesetzten Fragebogen sowie die aggregierten Ergebnisse der explorativen Vorerhebung unter sechs GaLaBau-Betrieben. Dokumentiert sind E-Mail-Aufkommen, wahrgenommener Bearbeitungsaufwand, Belastungseinschätzung, Einstellung zu Automatisierung, Ablehnungsgründe sowie Bedingungen für einen möglichen Einsatz.

Code	Frage / Variable
F1	Zeitstempel
F2	Wie viele Mails erhalten Sie durchschnittlich pro Tag (schätzungsweise)?
F3	Wie hoch schätzen Sie den täglichen Zeitaufwand für die Bearbeitung von E-Mails im Büro?
F4	Wie stark empfinden Sie E-Mail-Bearbeitung als Belastung im Büroalltag?
F5	Eine automatisierte Sortierung oder Vorverarbeitung von Mails könnte unsere Büroarbeit erleichtern.
F6	Ich kann mir grundsätzlich vorstellen eine solche digitale Lösung im Büro einzusetzen.
F7	Warum ist eine solche Lösung für Sie aktuell eher weniger interessant? (Mehrfachauswahl)
F8	Unter welchen Bedingungen wäre eine solche Lösung für Sie interessant? (Mehrfachauswahl)
F9	Wie würden Sie den Digitalisierungsstand Ihres Unternehmens insgesamt einschätzen?
F10	Wie offen ist ihr Unternehmen grundsätzlich für neue digitale Technologien?
F11	Wie viele Mitarbeiter beschäftigt Ihr Unternehmen?
F12	Haben Sie noch Anmerkungen, welche nicht durch die Fragen widerspiegelt werden konnten? Gerne auch Anregungen für Automatisierungen oder Themengebiete, die für Sie gerade sehr interessant/ wichtig sind.

TN	B1	B2	B3	B4	B5	B6
F1	19.03.2026, 17:40	20.03.2026, 09:09	20.03.2026, 12:50	24.03.2026, 11:39	25.03.2026, 12:56	26.03.2026, 13:03
F2	10-30	weniger als 10	10-30	10-30	10-30	10-30
F3	1-2 Stunden	30-60 min	30-60 min	1-2 Stunden	1-2 Stunden	30-60 min
F4	2 - eher gering	1 - gar nicht belastend	3 - mittel	3 - mittel	3 - mittel	5 - stimme voll zu
F5	2 - stimme eher nicht zu	2 - stimme eher nicht zu	4 - stimme eher zu	3 - neutral	3 - neutral	5 - stimme voll zu
F6	2 - stimme eher nicht zu	2 - eher unklar	4 - stimme eher zu	3 - neutral	4 - stimme eher zu	5 - stimme voll zu
F7	Jede Mail erfordert eine diskrete Antwort, da es Mails zu laufenden Projekten sind. Es kommen keine Mails rein, die standardisiert beantwortet werden könnten.	Nutzen für mich unklar	organisatorischer Aufwand für Einführung zu hoch	aktuell keine Priorität; Andere Umsetzungen, Softwareveränderung, [Branchensoftware], etc. haben gerade Priorität	wir haben kein großes Mail Aufkommen, der Aufwand im Büro ist überschaubar; organisatorischer Aufwand für Einführung zu hoch	Datenschutzbedenken
F8	klare Zeitersparnis	aktuell grundsätzlich kein Interesse	- klare Zeitersparnis - einfache Bedienung - zuverlässige Ergebnisse - Empfehlung durch andere Unternehmer/Kollegen - transparente und sichere Datenverarbeitung	aktuell grundsätzlich kein Interesse	- klare Zeitersparnis - einfache Bedienung - zuverlässige Ergebnisse	klare Zeitersparnis
F9	3 - mittel	3 - mittel	3 - mittel	3 - mittel	3 - mittel	3 - mittel
F10	5 - sehr offen	4 - eher offen	3 - neutral	4 - eher offen	4 - eher offen	5 - sehr offen
F11	11-20	6-10	11-20	mehr als 50	21-50	11-20
F12						

A2 n8n-Workflow-Konfiguration

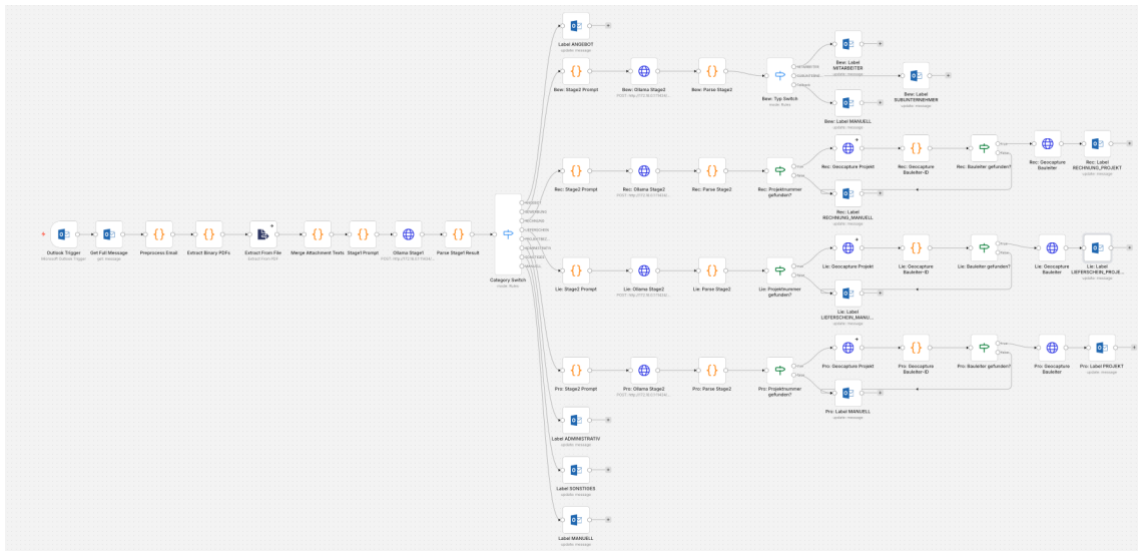
Der vollständige Workflow wurde in n8n umgesetzt. Aufgrund des Umfangs wird der vollständige JSON-Export nicht vollständig abgedruckt. Stattdessen dokumentiert Anhang A2 die zentrale Workflow-Struktur und die verwendeten Hauptkomponenten.

Komponente	Node(s)	Funktion
E-Mail-Eingang	Outlook Trigger, Get Full Message	Ruft ungelesene E-Mails inklusive Anhänge ab
Vorverarbeitung	Preprocess Email	Bereinigt den Nachrichtentext und extrahiert Metadaten
Anhangsanalyse	Extract Binary PDFs, Extract From File	Liest PDF-Anhänge aus
Hauptklassifikation	Stage1 Prompt, Ollama Stage1	Ordnet E-Mails einer Hauptkategorie zu
Entscheidungslogik	Parse Stage1 Result, Category Switch	Prüft Konfidenzwerte und leitet in die passende Prozesslogik
Detaillklassifikation	Stage2 Prompts	Prüft Bewerbung, Rechnung, Lieferschein und projektbezogene Nachrichten genauer
Systemintegration	Geocapture Nodes	Sucht Projektdaten und zuständige Personen
Ergebnis	Outlook Label Nodes	Vergibt passende Kategorien in Outlook

A3 Workflow-Screenshot und zentrale Code- beziehungsweise Prompt-Ausschnitte

Anhang A3 enthält den Screenshot des finalen n8n-Workflows sowie zentrale Code- und Prompt-Ausschnitte. Zugangsdaten, IDs und API-Tokens wurden anonymisiert.

Screenshot des n8n-Workflows



Vorverarbeitung eingehender E-Mails

Der Ausschnitt zur Vorverarbeitung zeigt, wie Metadaten wie Betreff, Absender und Anhänge ausgelesen werden. Außerdem wird der HTML-Inhalt der E-Mail bereinigt und auf eine maximale Länge reduziert.

```
JavaScript
const msg = $input.item.json;
const binary = $input.item.binary || {};

const attachmentKeys = Object.keys(binary).filter(k => k.startsWith('attachment_'));
const hasAttachments = attachmentKeys.length > 0;
const attachmentNames = attachmentKeys
  .map(k => binary[k]?.fileName || k)
  .join(', ');

let body = msg.body?.content || '';
body = body
  .replace(/<[^>]*>/g, ' ')
  .replace(/\\s+/g, ' ')
  .trim();

body = body.substring(0, 3000);

return [{
  json: {
    messageId: msg.id,
    subject: msg.subject || '',
    from: msg.from?.emailAddress?.address || '',
    fromName: msg.from?.emailAddress?.name || '',
    body,
    hasAttachments,
    attachmentCount: attachmentKeys.length,
    attachmentNames
  },
  binary
}];
```

Hauptklassifikation der E-Mail

Die Hauptklassifikation erfolgt über einen KI-Prompt. Die E-Mail wird anhand von Betreff, Absender, Nachrichtentext und optionalem PDF-Auszug einer vordefinierten Kategorie zugeordnet. Der Prompt fordert ausschließlich eine JSON-Ausgabe an, damit das Ergebnis im weiteren Workflow automatisiert verarbeitet werden kann.

```
JavaScript
const d = $input.item.json;
const pdfSnippet = d.attachmentText
  ? d.attachmentText.substring(0, 1500)
  : '';

const systemPrompt = `
Du bist ein E-Mail-Klassifizierungsassistent fuer ein Garten- und Landschaftsbauunternehmen.
Analysiere eingehende E-Mails und ordne sie EXAKT einer Kategorie zu.
Bei Unsicherheit: MANUELL.

Kategorien:
ANGEBOT – Neue Projektanfragen, Kundenanfragen fuer Leistungen
BEWERBUNG – Bewerbungen von Personen oder Unternehmen
RECHNUNG – Rechnungen, Abschlags- oder Schlussrechnungen
LIEFERSCHEIN – Lieferscheine oder Liefernachweise
PROJEKTBEZOGEN – Mails zu bestehenden Projekten
ADMINISTRATIV – Bank, Versicherung, Steuerberater, Behoerden
SONSTIGES – Newsletter, Werbung, irrelevante Info-Mails
MANUELL – Keine sichere Zuordnung moeglich

Regeln:
– confidence < 0.80 immer: manual_review_required = true
– Antworte NUR mit validem JSON, kein Text davor oder danach.

JSON-Format:
{
  "category": "ANGEBOT|BEWERBUNG|RECHNUNG|LIEFERSCHEIN|PROJEKTBEZOGEN|ADMINISTRATIV|SONSTIGES|MANUELL",
  "confidence": 0.0,
  "project_number_found": false,
  "project_number": null,
  "manual_review_required": false,
  "reason": ""
};

const userMessage = `
Betreff: ${d.subject}
Absender: ${d.fromName} <${d.from}>
Anhaenge vorhanden: ${d.hasAttachments}
Anhang-Dateinamen: ${d.attachmentNames}

E-Mail-Text:
${d.body}

${pdfSnippet ? 'PDF-Auszug (Anhang):\n' + pdfSnippet : ''}
`;

return [{
  json: {
    ...d,
    requestBody: {
      model: 'llama3.1:8b',
      stream: false,
      options: { temperature: 0 },
      messages: [
        { role: 'system', content: systemPrompt },
        { role: 'user', content: userMessage }
      ]
    }
  }
}];
```

Verarbeitung des KI-Ergebnisses

Nach der KI-Klassifikation wird die Antwort geparkt. Falls kein gültiges JSON zurückgegeben wird oder die Konfidenz unterhalb des Schwellenwerts liegt, wird die E-Mail automatisch der manuellen Prüfung zugeordnet.

```
JavaScript
const CONFIDENCE_THRESHOLD = 0.80;
const prev = $('Merge Attachment Texts').item().json;
const ollamaResp = $input.item().json;

let ai;

try {
  const raw = ollamaResp.message?.content
    || ollamaResp.choices?.[0]?.message?.content
    || '';

  const jsonMatch = raw.match(/\{[\s\S]*\}/);
  ai = JSON.parse(jsonMatch ? jsonMatch[0] : raw);
} catch(e) {
  ai = {
    category: 'MANUELL',
    confidence: 0,
    project_number_found: false,
    project_number: null,
    manual_review_required: true,
    reason: 'KI-Parse-Fehler: ' + e.message
  };
}

if ((ai.confidence || 0) < CONFIDENCE_THRESHOLD) {
  ai.manual_review_required = true;
}

ai.effective_category = ai.manual_review_required
  ? 'MANUELL'
  : ai.category;

return [{
  json: {
    ...prev,
    ...ai
  }
}];
```

Detailklassifikation am Beispiel Rechnungen

Für bestimmte Kategorien erfolgt eine zweite Analyseebene. Bei Rechnungen wird geprüft, ob eine Projektnummer oder ein anderes Projektkennzeichen in der E-Mail oder im PDF-Anhang enthalten ist. Kann keine sichere Projektnummer identifiziert werden, wird die Nachricht zur manuellen Prüfung weitergeleitet.

```
JavaScript

const d = $input.item.json;
const attachmentText = d.attachmentText || '';

const systemPrompt = `
Du analysierst eine Rechnungs-E-Mail fuer ein GalaBau-Unternehmen.
Suche in Betreff, E-Mail-Text UND Anhangsinhalt nach:
Projektnummern, Auftragsnummern, Baustellenbezeichnungen oder internen Referenzen.

Regeln:
- Erfinde keine Projektnummern.
- Bei mehreren Kandidaten: wahrscheinlichste nennen und begruenden.
- Keine sichere Projektnummer: project_number_found = false,
  manual_review_required = true.

Antworte NUR mit validem JSON:
{
  "confirmed_category": "RECHNUNG",
  "project_number_found": false,
  "project_number": null,
  "alternative_project_numbers": [],
  "document_signals": [],
  "manual_review_required": false,
  "reason": ""
};

const userMessage = `
Betreff: ${d.subject}
Absender: ${d.fromName} <${d.from}>

E-Mail-Text:
${d.body}

${attachmentText ? 'Anhangsinhalt (PDF):\n' + attachmentText : '{(kein Anhang)'}
`;

return [{
  json: {
    ...d,
    requestBody: {
      model: 'llama3.1:8b',
      stream: false,
      options: { temperature: 0 },
      messages: [
        { role: 'system', content: systemPrompt },
        { role: 'user', content: userMessage }
      ]
    }
  }
}];
```

Die dargestellten Ausschnitte verdeutlichen die zentrale Logik des Prototyps. Eingehende E-Mails werden technisch bereinigt, KI-basiert klassifiziert und bei Unsicherheit oder fehlenden Projektinformationen zur manuellen Prüfung weitergeleitet. Dadurch wird die KI nicht als autonom entscheidendes System eingesetzt, sondern als vorbereitende Assistenzfunktion innerhalb eines kontrollierten Workflows.

A4 Mail-Tagebuch-Rohdaten

Anhang A4 enthält die vier ausgefüllten Mail-Tagebuch-Seiten vom 29.04.2026 bis 05.05.2026 mit Bearbeitungszeiten, Aufwandseinschätzungen und dokumentierten Besonderheiten.

Foto des manuell geführten Mail-Tagebuchs

	Tag 1 29.04.	Tag 2 30.04.	Tag 3 04.05.	Tag 4 05.05.
1 Strich = 5 Minuten				
Klassifizierung (rel. Wichtigkeit)	Erhöht	Normal	Normal	Normal
Besonderheiten	Spam-Identifizierung			

A5 Protokolle und Projektanbahnung mit dem Partnerunternehmen

A5.1 Projektanbahnung und Kommunikationsweg

Die Projektanbahnung erfolgte schrittweise über das Verbandsumfeld des GaLaBau. Eine frühe allgemeine Prototypidee zur KI-gestützten E-Mail-Verarbeitung wurde zunächst im Digitalisierungsausschuss des Fachverbands Garten-, Landschafts- und Sportplatzbau Hessen-Thüringen vorgestellt. Im Anschluss wurde über das Verbandsumfeld nach einem geeigneten Partnerunternehmen für eine fallbezogene Erprobung gesucht. Daraufhin meldete sich der Prozessmanager des späteren anonymisierten Partnerunternehmens.

In ersten Gesprächen mit dem Prozessmanager wurde zunächst die allgemeine Idee einer KI-gestützten E-Mail-Kategorisierung eingeordnet. Der Stand war, dass wir über meinen bereits vorhanden P1 als grundlegenden Ansatz zur Kategorisierung eingehender E-Mails nach Nachrichtentypen gesprochen haben. Anfang März 2026 wurde der Ansatz anhand der Organisationsstruktur des Partnerunternehmens weiter konkretisiert. Da das Unternehmen nach Abteilungen und fachlichen Zuständigkeiten strukturiert ist, verschob sich der Fokus von einer allgemeinen Kategorisierung hin zu stärker prozessbezogenen Unterstützungsfunktionen. Daraus entstand P2 als Ansatz zur Erkennung von Antwortnotwendigkeit und zur Unterstützung durch Antwortentwürfe. Da die betriebliche Anschlussfähigkeit von P2 noch nicht eindeutig war, wurde anschließend ein Erstgespräch mit der IT-Abteilung des Partnerunternehmens vereinbart. Dieses Gespräch diente der weiteren Use-Case-Schärfung und wird in Anhang A5.2 dokumentiert.

Diese Projektanbahnung ist keine eigenständige empirische Erhebung, sondern dokumentiert den Kommunikationsweg und die Entstehung des fallbezogenen Use Cases. Sie macht nachvollziehbar, wie aus einer allgemeinen Prototypidee über Gespräche mit Prozessmanagement und IT-Abteilung ein konkreter Untersuchungsfall für die Bachelorarbeit entstand.

A5.2 Ergebnisprotokoll des Erstgesprächs mit der IT-Abteilung

Anlass: Erstes Gespräch mit der IT-Abteilung des Partnerunternehmens zur Prüfung eines KI-gestützten E-Mail-Systems und zur Schärfung eines betrieblich relevanten Use Cases

Datum: 25.03.2026, 11:00-12:00 Uhr

Format: Online-Meeting

Teilnehmende: Verfasser, Prozessmanager und IT-Ansprechpartner des Partnerunternehmens

Datengrundlage: Mündliche Beschreibung bisheriger Prototypideen, technischer Anforderungen und möglicher betrieblicher Einsatzfelder

Dokumentationsform: Anonymisiertes Ergebnisprotokoll. Das Gespräch wurde nicht aufgezeichnet, da firmeninterne E-Mail-Prozesse, vertrauliche Zuständigkeiten und potenziell sensible Kommunikationsabläufe besprochen wurden. Das Protokoll stellt daher kein Transkript dar, sondern eine nachträgliche Verdichtung der im Gespräch erhobenen Projekt- und Prozessinformationen.

Ausgangslage des Gesprächs

Vor dem Gespräch mit der IT-Abteilung bestand bereits eine allgemeine Prototypidee zur KI-gestützten E-Mail-Verarbeitung. P1 zielte auf eine grundlegende Kategorisierung eingehender E-Mails nach Nachrichtentypen. Im weiteren Austausch mit dem Prozessmanager entstand P2 als Ansatz zur Erkennung von Antwortnotwendigkeit und zur Unterstützung durch Antwortentwürfe. Beide Ansätze prüften allgemeine Unterstützungsfunktionen, bildeten jedoch noch keinen stabil abgegrenzten betrieblichen Zielprozess ab.

Zentrale Gesprächsinhalte

- Im Gespräch wurde zunächst erläutert, wie KI-gestützte E-Mail-Verarbeitung im Rahmen der Bachelorarbeit untersucht werden kann.
- Die bisherigen Prototypideen P1 und P2 wurden im Hinblick auf betriebliche Anschlussfähigkeit und Nutzenpotenzial eingeordnet.
- Im Austausch mit der IT-Abteilung wurde der bisherige Ansatz neu fokussiert, da eine reine Antwortunterstützung nicht als vorrangiger Anwendungsfall weiterverfolgt wurde.
- Als neuer Use Case wurde die automatische Weiterleitung eingehender E-Mails aus einem zentralen Info@-Postfach diskutiert.
- Die IT-seitige Ausgangsannahme war, dass eingehende E-Mails im Wesentlichen einem von drei organisatorischen Firmenbereichen zugeordnet werden könnten.
- Daraus entstand P3 als Prototypidee einer KI-gestützten Weiterleitung an drei angenommene Zielbereiche.

Ergebnis des Gesprächs

Das Gespräch markierte den Übergang von allgemeinen E-Mail-Unterstützungsfunktionen zu einem konkreteren betrieblichen Automationsfall. P2 wurde nicht weiter als zentraler Untersuchungsgegenstand verfolgt, da die automatische Weiterleitung des Info@-Postfachs als klarer abgrenzbarer und technisch prüfbarer Use Case erschien. Die Drei-Bereiche-Logik wurde dabei zunächst als plausible Ausgangsannahme übernommen, musste jedoch in einem späteren Prozessmapping mit operativen Anwendenden überprüft werden.

Konsequenzen für die Prototypentwicklung

- P3 wurde als neuer Entwicklungsschritt definiert.
- Der Fokus verschob sich von Antwortgenerierung zu Klassifikation und Routing.
- Die IT-Abteilung brachte die Anforderung ein, technische Machbarkeit und betriebliche Einsetzbarkeit stärker am realen Info@-Prozess auszurichten.
- Die Drei-Zielbereiche-Annahme bildete die Grundlage für das nachfolgende Prozessmapping am 21.04.2026.
- Das spätere Prozessmapping zeigte, dass diese Annahme für den realen Prozess zu stark vereinfacht war und erweitert werden musste.

Methodische Einordnung

Das Erstgespräch mit der IT-Abteilung ist methodisch relevant, weil es die Entstehung des fallbezogenen Use Cases dokumentiert. Es erklärt, warum die Arbeit nicht bei einer allgemeinen E-Mail-Kategorisierung oder Antwortunterstützung stehen blieb, sondern die Weiterleitung eines zentralen Info@-Postfachs als konkreten Automationsfall untersuchte. Das Gespräch ist keine eigenständige Evaluation, sondern Teil der gestaltungsorientierten Prototypentwicklung. Seine Bedeutung liegt darin, die Übergänge zwischen P1, P2 und P3 nachvollziehbar zu machen.

A5.3 Ergebnisprotokoll des Prozessmappings

Anlass: Prozessmapping des zentralen Info@-Postfachs als Grundlage für die fallbezogene Prototypentwicklung

Datum: 21.04.2026, 10:30-11:30 Uhr

Format: Online-Meeting

Teilnehmende: Verfasser, IT-Ansprechpartner, Leitung Empfang, Prozessmanager und Geschäftsführung des Partnerunternehmens

Datengrundlage: Mündliche Prozessbeschreibung durch die operativ verantwortliche Leitung des Empfangs sowie ergänzende Einordnungen durch IT, Prozessmanagement und Geschäftsführung

Dokumentationsform: Anonymisiertes Ergebnisprotokoll. Das Gespräch wurde nicht aufgezeichnet, da firmeninterne E-Mail-Prozesse, vertrauliche Zuständigkeiten und potenziell sensible Kommunikationsabläufe besprochen wurden. Das Protokoll stellt daher kein Transkript dar, sondern eine nachträgliche Verdichtung der im Gespräch erhobenen Prozessinformationen.

Vorgehen des Prozessmappings

- Zu Beginn wurden Ziel, Kontext und Vorgehen der Bachelorarbeit erläutert.
- Anschließend wurde die bis dahin angenommene Automationslogik besprochen, nach der eingehende E-Mails im Wesentlichen einem von drei organisatorischen Firmenbereichen zugeordnet werden sollten.
- Diese Annahme beruhte auf vorgelagerten Gesprächen mit Prozessmanagement und IT-Abteilung und diente als Ausgangspunkt für die Prozessklärung.
- Der Schwerpunkt des Gesprächs lag auf der Beschreibung des tatsächlichen Info@-Bearbeitungsprozesses durch die Leitung des Empfangs.
- In das produktive Postfach wurde im Rahmen des Gesprächs nicht gemeinsam systematisch Einsicht genommen. Die Prozesslogik wurde anhand der mündlichen Beschreibung typischer Eingangsfälle, Zuständigkeiten und Weiterleitungsentscheidungen rekonstruiert.

Zentrale Ergebnisse

- Die ursprüngliche Drei-Bereiche-Logik erwies sich als zu stark vereinfacht.
- Der reale Info@-Prozess umfasst unterschiedliche Eingangstypen, darunter Angebote, Bewerbungen, Rechnungen, Lieferscheine, projektbezogene Nachrichten, administrative E-Mails sowie sonstige oder manuell zu prüfende Fälle.
- Mehrere Kategorien erfordern weitere Unterentscheidungen. Dies betrifft insbesondere Bewerbungsarten, projektbezogene Dokumente, Rückmeldungen auf Preisanfragen und Fälle mit unklarer Zuständigkeit.

- Die Weiterleitung hängt nicht nur vom Inhalt der E-Mail ab, sondern teilweise auch von Projektbezug, Zuständigkeiten, vorhandenen Stammdaten und betrieblichen Erfahrungsregeln.
- Einzelne Zuständigkeiten werden im Alltag nicht vollständig regelbasiert entschieden, sondern situativ durch Rückfragen oder internes Prozesswissen geklärt.
- Daraus ergab sich, dass eine direkte Vollautomation mit drei Zielbereichen fachlich nicht ausreicht.

Konsequenzen für die Prototypentwicklung

- Das Kategorienschema wurde gegenüber der ursprünglichen Annahme erweitert.
- Die spätere Prototypentwicklung berücksichtigte mehrere Hauptkategorien sowie manuelle Prüfpfade.
- Die Idee einer sofortigen produktiven Vollautomation wurde zugunsten einer kontrollierten Labelphase zurückgestellt.
- Die Geocapture-Anbindung wurde als möglicher Baustein für projektbezogene Routingentscheidungen betrachtet, setzte jedoch eindeutige Projektnummern und nutzbare Zuständigkeitsdaten voraus.
- Das Prozessmapping bildete damit eine zentrale Grundlage für die in Kapitel 4 beschriebene Kategorisierungslogik und für die spätere Bewertung der Prozessanschlussfähigkeit in Kapitel 5.

Methodische Einordnung

Das Prozessmapping liefert eine praxisnahe, aber explorative Rekonstruktion des Info@-Bearbeitungsprozesses. Da keine Aufzeichnung, kein vollständiges Transkript und keine systematische Auswertung aller Eingangsmails des Postfachs vorlagen, wird kein Anspruch auf vollständige Organisationsanalyse erhoben. Für die Zielsetzung der Arbeit ist das Ergebnis dennoch tragfähig, weil das Gespräch die zentrale Implementierungsgrenze sichtbar machte: KI-gestützte E-Mail-Automation setzt nicht nur technische Klassifikation voraus, sondern eine zuvor geklärte und formalisierbare Prozesslogik.

A5.4 Protokoll der Labelauswertung

Anlass: Gemeinsame Auswertung der Labelphase des KI-Prototyps für die Info@-Mail-Triage

Datum: 20.05.2026, 14:00-15:30 Uhr

Format: Online-Meeting

Teilnehmende: Verfasser und IT-Ansprechpartner des Partnerunternehmens

Datengrundlage: 57 ausgewertete E-Mails, Spam ausgeschlossen, Labels direkt im betrieblichen Outlook-Postfach gesetzt

Vorbereitung: Der IT-Ansprechpartner hatte vorab Rückmeldungen aus dem Empfang eingeholt, der das Info@-Postfach operativ bearbeitet.

Dokumentationsform: Anonymisiertes Ergebnisprotokoll. Das Gespräch wurde nicht aufgezeichnet, da firmeninterne E-Mail-Prozesse, vertrauliche Zuständigkeiten und potenziell sensible Kommunikationsabläufe besprochen wurden. Das Protokoll stellt daher kein Transkript dar, sondern eine nachträgliche Verdichtung der gemeinsamen Auswertung der Labelphase.

Vorgehen der Auswertung

- Alle 57 E-Mails wurden geprüft, nicht nur falsche oder unklare Fälle.
- Je E-Mail wurden gesetztes Label, tatsächliche Weiterverarbeitung und fachlich richtiges Weiterleitungsziel geprüft.
- Bewertungsmaßstab war nicht allein die semantische Hauptkategorie, sondern das fachlich richtige spätere Weiterleitungsziel, abgebildet über die gesetzten Postfachlabels.
- Die qualitativen Befunde speisen sich aus Rückmeldungen des Empfangs und aus der gemeinsamen E-Mail-für-E-Mail-Auswertung im Gespräch.

Quantitatives Ergebnis

Bewertung	Anzahl	Anteil
Eindeutig richtig zugeordnet	25	43,9 %
Falsch oder uneindeutig zugeordnet	26	45,6 %
Keine passende Kategorie vorhanden	6	10,5 %
Gesamt	57	100,0 %

Die sechs E-Mails ohne passende Kategorie bilden eine eigenständige dritte Gruppe und sind nicht Teil der 26 falsch beziehungsweise uneindeutig zugeordneten E-Mails. Eine nachträgliche Aufteilung der 26 Fälle in eindeutige Modellfehler, fachliche Grenzfälle und Prozessunklarheiten erfolgte nicht, da hierfür eine systematische Mehrpersonenvalidierung erforderlich gewesen wäre.

Qualitative Befunde

- Die Kategorien MANUELL und SONSTIGES waren nicht trennscharf und fingen Nachrichten ab, die fachlich einer eigenen Kategorie bedurft hätten.
- Fehlende Kategorien wurden erkannt. Werbung, Lieferanteninformationen sowie Antwort auf Preisanfrage.
- Prozessklarheit erwies sich als zentraler Engpass. Der Empfang muss teilweise selbst rückfragen, welche Person oder Abteilung zuständig ist. Diese Entscheidung lässt sich nicht ohne Weiteres in eine stabile Routinglogik überführen.
- Die Geocapture-Anbindung war technisch möglich, wurde jedoch durch Stammdatenqualität begrenzt, uneindeutige oder mehrfache Projektnummern und nicht nutzbare Zuständigkeitsverknüpfungen verhinderten verlässliches Routing.
- Unvollständige Eingangsdaten begrenzten die Klassifikation. Fehlende Projektnummern, unklare Betreffzeilen, allgemeine Formulierungen und mehrdeutige Anhänge erschweren die Zuordnung für Mensch und KI gleichermaßen.
- Die Durchlaufzeit wurde als zusätzliche relevante Kennzahl sichtbar. Einzelne E-Mails verursachten nur wenige Minuten aktive Bearbeitungszeit, wurden durch Rückfragen und unklare Zuständigkeiten jedoch erst deutlich später final zugestellt.

Schlussfolgerungen und Konsequenzen

- Der Vollautomationsworkflow wurde nicht produktiv aktiviert, da ein wirtschaftlich positiver Effekt auf Basis der vorhandenen Erhebungsdaten nicht belastbar nachgewiesen werden konnte.
- Für eine Fortführung wäre ein reduzierter Automatisierungsgrad sinnvoll. Fokus auf häufige, stabile und risikoarme Fälle, unklare Fälle bleiben bewusst manuell.
- Das Kategorienschema müsste vor einem Produktivbetrieb weiter differenziert und fachlich validiert werden.

- Das lokal eingesetzte Modell llama3.1:8b war technisch funktionsfähig und ermöglichte eine datenschutzorientierte Verarbeitung, stieß aber bei komplexen Kategorien und langen Anhängen an Grenzen.
- Der Prototyp und die Evaluation wurden im Umfang der Bachelorarbeit an dieser Stelle abgeschlossen. Eine mögliche Fortführung wäre nach Abschluss der Arbeit neu zu bewerten.

A6 Datenschutzdokumentation

Anhang A6 dokumentiert die datenschutzorientierte Einbettung des Prototyps in anonymisierter Form. Auf die Aufnahme vollständiger Originaldokumente wird verzichtet, da diese identifizierende Unternehmens-, Vertrags-, Server- und Kontaktdaten enthalten. Stattdessen werden die im Projekt erstellten beziehungsweise geprüften Datenschutzartefakte, der Datenfluss und die daraus abgeleiteten Schutzmaßnahmen zusammenfassend dargestellt.

A6.1 Datenschutzbezogene Projektdokumente

Dokumentationsbestandteil	Zweck im Projekt	Status	Anonymisierung im Anhang
Datenschutzhinweis beziehungsweise Datenschutzerklärung zum Info@-Prozess	Information über die Verarbeitung eingehender E-Mails im Rahmen der prototypischen Prozessanalyse	erstellt	Unternehmensname, Kontaktdaten und personenbezogene Ansprechpartner wurden nicht übernommen
Auftragsverarbeitungsvertrag mit dem Hosting-Anbieter	Vertragliche Grundlage für die Verarbeitung auf dem eingesetzten VPS beziehungsweise der technischen Infrastruktur des Auftragsverarbeiters	abgeschlossen	Anbietername, Vertragsdaten, Serverkennungen, Adressen und Unterschriften wurden nicht übernommen
Datenfluss- und Verarbeitungsdokumentation	Übersicht darüber, welche Daten zu welchem Zeitpunkt, in welchem System und zu welchem Zweck verarbeitet werden	erstellt	technische Kennungen, IP-Adressen, API-Schlüssel, konkrete Nutzerkonten und interne Zuständigkeiten wurden entfernt

Die Tabelle zeigt, dass Datenschutz nicht erst nach der technischen Entwicklung betrachtet wurde, sondern als eigenes Projektdokumentationsfeld in die prototypische Umsetzung einbezogen war. Für die Bachelorarbeit ist dabei nicht die rechtliche Vollprüfung der Unterlagen entscheidend, sondern der Nachweis, dass Datenflüsse, Auftragsverarbeitung und technische Zugriffspunkte vor einer möglichen produktiven Nutzung bewusst dokumentiert wurden.

A6.2 Anonymisierte Datenflussdokumentation

Prozessschritt	Verarbeitete Daten	Ort der Verarbeitung	Datenschutzorientierte Maßnahme
Eingang im zentralen Info@-Postfach	Absenderdaten, Betreff, E-Mail-Text, Anhänge und Zeitstempel	Microsoft 365 beziehungsweise Outlook des Partnerunternehmens	Verarbeitung bleibt zunächst im bestehenden betrieblichen Postfachprozess
Abruf durch den Workflow	E-Mail-Metadaten, Nachrichtentext und Anhangsinformationen	n8n-Workflow auf der eingesetzten technischen Infrastruktur	Zugriff nur auf die für den Prototyp benötigten Postfachdaten
Vorverarbeitung	bereinigter Nachrichtentext, Absenderangaben, Betreff, Anhangsnamen und technische Metadaten	n8n-Workflow	Reduktion und Strukturierung der Daten vor der KI-Klassifikation
KI-Klassifikation	E-Mail-Text, Betreff, Absenderkontext und bei Bedarf auslesbare Anhangsinhalte	lokal angebundenes Ollama-Modell llama3.1:8b	keine Verarbeitung über externe KI-API. Klassifikation erfolgt innerhalb der projektbezogen kontrollierten Infrastruktur
Detailprüfung bei bestimmten Kategorien	projektrelevante Hinweise, Rechnungs- oder Lieferscheininformationen, mögliche Projektnummern	n8n-Workflow und lokale KI-Verarbeitung	Verarbeitung nur, wenn die Kategorie eine Detailprüfung erfordert
Abfrage des Fachsystems	mögliche Projektnummern oder projektbezogene Identifikatoren	Geocapture-API beziehungsweise Fachsystem des Partnerunternehmens	Beschränkung auf die für Zuordnung und Routing erforderlichen Projektinformationen
Rückgabe in den Mailprozess	KI-Label, manuelle Prüfmarkierung,	Outlook-Postfach des Partnerunternehmens	keine produktive Vollautomation,

	potenzielles Weiterleitungsziel	unklare Fälle bleiben manuell zu prüfen
--	------------------------------------	--

Der Datenfluss verdeutlicht die zentrale Datenschutzlogik des Prototyps. E-Mail-Inhalte werden nicht an externe KI-Dienstleister übertragen, sondern in einem kontrollierten Workflow verarbeitet. Die KI übernimmt dabei keine endgültige Entscheidung, sondern erzeugt vorbereitende Klassifikationsinformationen, die in der Labelphase weiterhin fachlich geprüft wurden.

A6.3 Datenschutzorientierte Maßnahmen

Maßnahme	Umsetzung im Projekt	Bedeutung für die Arbeit
Datenminimierung	Im Workflow werden nur die Informationen verarbeitet, die für Klassifikation, Anhangsanalyse und mögliche Projektzuordnung erforderlich sind.	Die technische Umsetzung folgt dem Ziel, unnötige Datenverarbeitung zu vermeiden.
Lokale KI-Verarbeitung	Die Modellinferenz erfolgt über Ollama mit dem Modell llama3.1:8b und nicht über eine externe KI-API.	Datenschutzbedenken aus Vorerhebung und Projektgesprächen werden durch die Architektur direkt adressiert.
Auftragsverarbeitung	Für die eingesetzte Hosting-Infrastruktur wurde ein Auftragsverarbeitungsvertrag berücksichtigt.	Der VPS-Betrieb wird nicht als beliebige externe Verarbeitung behandelt, sondern projektbezogen dokumentiert.
Zugriffsbeschränkung	Zugangsdaten, API-Schlüssel, technische IDs und Serverdetails werden nicht im Anhang offengelegt.	Der Anhang bleibt prüfbar, ohne sicherheitsrelevante Details preiszugeben.
Geheimhaltung	Die Zusammenarbeit mit dem Partnerunternehmen wurde zusätzlich über eine Geheimhaltungsvereinbarung abgesichert.	Unternehmensinterne Informationen werden im Manuskript und im Anhang anonymisiert dargestellt.
Menschliche Kontrolle	In der Labelphase wurden KI-Ergebnisse nicht produktiv zur	Die Arbeit beschreibt eine Assistenz- und Prüfstruktur,

	automatischen Weiterleitung genutzt, sondern fachlich bewertet.	keine autonom entscheidende Vollautomation.
Dokumentation vor Produktivbetrieb	Datenflüsse, Systemrollen und datenschutzbezogene Unterlagen wurden vor einer möglichen produktiven Nutzung gesammelt.	Datenschutzorientierung wird als Praxistauglichkeitskriterium nachvollziehbar dokumentiert.

Die Datenschutzdokumentation begründet damit die in der Arbeit gewählte vorsichtige Formulierung einer datenschutzorientierten Gestaltung. Eine rechtsgutachterliche Prüfung der DSGVO-Konformität wurde im Rahmen der Bachelorarbeit nicht durchgeführt und wird durch diesen Anhang nicht beansprucht. Der Anhang zeigt vielmehr, welche technischen und organisatorischen Maßnahmen im Projekt berücksichtigt wurden, damit der Prototyp unter realen betrieblichen Bedingungen verantwortbar erprobt werden konnte.