

Datengetriebenes Requirements Engineering für die Integration von KI in Softwaresysteme

Alexander Tewes

Master Informatik

Hochschule Bonn-Rhein-Sieg

Sankt Augustin, Deutschland

alexander.tewes@smail.inf.h-brs.de

Abstract—Angesichts der raschen Entwicklungen und der Besonderheiten von Softwaresystemen, welche Künstliche Intelligenz (KI) nutzen, ist ein angepasstes Requirements Engineering (RE) erforderlich. Die spezifischen Anforderungen von KI-Projekten müssen dabei erkannt und angegangen werden. Hierfür wird eine systematische Überprüfung bestehender Herausforderungen des RE in KI-Projekten durchgeführt. Darauf aufbauend werden neue RE-Ansätze und Empfehlungen präsentiert, die auf die Datensicht von KI-Projekten abzielen. Mithilfe der Analyse bestehender Lösungsansätze, Methoden, Frameworks und Tools soll aufgezeigt werden, inwiefern die Herausforderungen im RE bewältigt werden können. Noch bestehende Lücken im Forschungsstand werden identifiziert und aufgezeigt.

Index Terms—Requirements Engineering, Data, Artificial Intelligence, Machine Learning

I. INTRODUCTION

Requirements Engineering (RE) ist ein zentraler Bestandteil der Softwareentwicklung. Das RE behandelt die systematische Auseinandersetzung mit der Spezifikation und Verwaltung von Anforderungen [1]. RE ist entscheidend, um sicherzustellen, dass entwickelte Systeme den Bedürfnissen und Erwartungen der Stakeholder gerecht werden. Die Einführung von Künstlicher Intelligenz (KI) in die Softwareentwicklung hat neue Herausforderungen und Komplexitäten hervorgebracht. Diese Entwicklung erfordert eine Anpassung der bestehenden RE-Aktivitäten, um die spezifischen Anforderungen von KI-Projekten bewältigen zu können.

Im Vergleich zu traditionellen Softwareprojekten weisen KI-Projekte einzigartige Merkmale auf. Eines der Hauptmerkmale ist die starke Abhängigkeit von Daten, die oft entscheidend für den Projekterfolg ist [2]. KI-Modelle sind häufig als Black-boxes zu betrachten, deren Funktionsweise nicht vollständig transparent ist. Diese Eigenschaften stellen besondere Anforderungen an die Validierung und Verifikation von KI-Systemen [3]. Weiterhin zeichnen sich KI-Systeme durch ihre Fähigkeit zu Lernen und zur stetigen Anpassung aus. Somit wird eine fortlaufende Überwachung und Adaption der Anforderungen notwendig.

Diese Besonderheiten zeigen, dass es vonnöten ist, datengetriebene RE-Methoden für KI-Projekte anzupassen oder zu entwickeln. Ein tiefgehendes Verständnis der spezifischen Herausforderungen von KI-Projekten ist entscheidend, um ein

effektives RE zu gewährleisten. Solch ein angepasstes RE soll sicherstellen, dass die entwickelten KI-Systeme vertrauenswürdig, ethisch und sicher sind [4].

Die Rolle von KI wächst rasant in einer Vielzahl von Anwendungen und Branchen. Die fortschreitende Automatisierung bringt KI in den Vordergrund. Anwendungen reichen von persönlichen Assistenten bis hin zu autonomen Fahrzeugen. Mit diesen Möglichkeiten kommen jedoch auch signifikante Herausforderungen auf, insbesondere in der Entwicklung solcher Systeme. [5]

RE ist in diesem Zusammenhang ein kritischer Faktor. Die korrekte Erfassung und Analyse von Anforderungen entscheidet über den Erfolg eines KI-Projekts. Traditionelle RE-Methoden sind nicht direkt auf KI-Projekte übertragbar, da dessen Eigenschaften sich von traditionellen Softwareprojekten unterscheiden [6]. Das Fehlen angepasster RE-Methoden für KI kann zu unklaren Erwartungen, fehlenden Funktionen und gesellschaftlichen Problemen führen [4].

Die Verwaltung und das Management von Daten haben sich als entscheidende Aspekte in der Entwicklung von KI-Projekten herausgestellt. Dies gilt insbesondere für Projekte, die maschinelles Lernen (ML) einsetzen. Die Komplexität und dynamische Natur von ML-Systemen erfordern einen Paradigmenwechsel in den traditionellen RE-Praktiken [7]. Im Gegensatz zu konventionellen Softwaresystemen leiten ML-Modelle ihre Funktionalität aus den Daten ab, mit denen sie trainiert werden. Dieser Ansatz erfordert eine Erweiterung des RE. Die Eigenschaften der Trainingsdaten, die Interpretierbarkeit des Modells und die Datenverwaltung sind zu berücksichtigen. Die Integration von Big Data in solchen Projekten erhöht zudem maßgeblich die Komplexität. [8]

Die nachfolgenden Abschnitte dieser Arbeit sind wie folgt gegliedert: “Research Method” beschreibt die Methodik, die zur Untersuchung der Herausforderungen und Lösungen im RE für KI-Projekte angewandt wird. Im dritten Abschnitt “Results” werden die daraus entstandenen Forschungsergebnisse dargestellt. Der vierte Teil “Discussion” stellt eine Diskussion der Ergebnisse dar, einschließlich der Implikationen für die Praxis, zukünftige Forschungsrichtungen und Kritikpunkte der Ansätze. Schließlich folgt der fünfte Abschnitt “Conclusion”, der die wichtigsten Ergebnisse der Arbeit aufzeigt und interpretiert.

II. RESEARCH METHOD

In diesem Abschnitt werden zunächst Forschungsfragen formuliert. Danach wird auf die Ziele der Arbeit eingegangen. Im Anschluss wird eine Systematische Literaturrecherche konzipiert und durchgeführt.

A. Research Questions

- **RQ1:** Welche spezifischen Herausforderungen entstehen für das RE bei der Datensicht in KI-Projekten?
- **RQ2:** Wie kann man diese Herausforderungen in der RE-Praxis angehen?
 - **RQ2.1:** Was muss innerhalb der RE-Aktivitäten berücksichtigt werden, um dessen Herausforderungen in KI-Projekten zu bewältigen?
 - **RQ2.2:** Welche Methoden, Frameworks und Tools existieren hierzu?
- **RQ3:** Für welche RE-Aktivitäten fehlen noch Methoden in Bezug auf KI-Projekte?

Diese Fragen bilden das Fundament dieser Arbeit und dienen als Leitlinie für die folgenden Abschnitte. Es wird angestrebt, durch die Beantwortung dieser Fragen eine umfassende Perspektive auf das datengetriebene RE in KI-Projekten zu bieten. Den identifizierten Herausforderungen sollen konkrete Lösungsansätze und Methoden gegenübergestellt werden.

Mittels der Beantwortung der zugrunde liegenden Forschungsfragen sollen folgende Ziele erreicht werden:

- Erkennung der besonderen Anforderungen und Herausforderungen im RE für KI-Projekte im Vergleich zu traditionellen Softwareprojekten.
- Übersicht von datengetriebenen RE-Methoden und -Konzepten, welche auf die Eigenschaften von KI-Projekten zugeschnitten sind.

B. Systematic Literature Review

Durch die systematische Literaturrecherche i.A.a. Kitchenham et al. [9] sollen bereits in der Literatur untersuchte Herausforderungen des RE in KI-Projekten festgestellt werden. Des Weiteren sollen bestehende Lösungsansätze zusammengefasst und noch existierende Lücken im derzeitigen Forschungsstand erkannt werden. Dazu wird zunächst ein Suchprotokoll definiert, welches die Einzelheiten der initialen Suche enthält. Anschließend wird die Relevanz der ausgewählten Papiere bestimmt und in Kategorien gruppiert.

Suchprotokoll

Datenbanken und Suchplattformen:

- IEEE Xplore
- ACM Digital Library

Suchbegriffe:

- “Requirements Engineering”
- “Artificial Intelligence” oder “AI”
- “Machine Learning” oder “ML”
- “Software Engineering”

Suchanfragen:

- “Requirements Engineering” AND (“Artificial Intelligence” OR “AI” OR “Machine Learning” OR “ML”)
- “Software Engineering” AND (“Artificial Intelligence” OR “AI” OR “Machine Learning” OR “ML”)
- Zudem: Nutzen von Snowballing, um weitere relevante Papiere zu identifizieren

Einschlusskriterien:

- Explizite Diskussion von RE-Methoden zur Entwicklung von KI-Systemen.
- Veröffentlichung innerhalb der letzten sechs Jahre.
- Veröffentlichung in Journalen oder Konferenzen.

Ausschlusskriterien:

- Andere Sprachen als Englisch oder Deutsch.
- Übersichtspapiere oder Kurzfassungen.

Durchführung der SLR

Die Durchführung der systematischen Literaturrecherche erfolgt in mehreren Schritten. Im ersten Schritt wird eine initiale Suche basierend auf den definierten Suchbegriffen in den ausgewählten Datenbanken durchgeführt. Daraufhin werden die Suchergebnisse gemäß den festgelegten Einschluss- und Ausschlusskriterien gefiltert.

Jede inkludierte Quelle wird basierend auf ihrer Relevanz für die zentralen Forschungsfragen dieser Arbeit bewertet. Die Bewertungsskala reicht von niedrig (1) bis hoch (3). Das Thema, die Relevanzeinstufung jeder Quelle sowie dessen ausschlaggebender Inhalt für diese Arbeit sind in der Tabelle IV dargestellt.

Quellen, welche die festgelegten Kriterien nicht erfüllen, werden von der weiteren Betrachtung ausgeschlossen. Die Gründe für die Exklusion jeder einzelnen Quelle sind in Tabelle V aufgeführt. Diese Ausschlusskriterien gewährleisten, dass nur die relevantesten Beiträge in die Analyse einbezogen werden.

Um eine strukturierte Übersicht über das Forschungsfeld zu erhalten, wurden die inkludierten Studien in vier Hauptkategorien gruppiert. Tabelle I zeigt die Gruppierung der Papiere und deren Zuordnung zu den entsprechenden Forschungsfragen.

Anschließend erfolgt eine detaillierte Analyse der identifizierten Papiere, um deren Beiträge im Kontext der Forschungsfragen zu bewerten. Dieser Vorgang beinhaltet eine Auswertung der in den Papieren vorgestellten Methoden, Konzepte und Erkenntnisse sowie deren Relevanz für das datengetriebene RE in KI-Projekten. Die Ergebnisse dieser Analyse liefern einen Überblick über den aktuellen Forschungsstand im Bereich RE für KI. Vorhandene Forschungslücken werden aufgezeigt. Dieser Abschnitt wird mittels der zuvor definierten Forschungsfragen strukturiert.

III. RESULTS

RQ1: Welche spezifischen Herausforderungen entstehen für das RE bei der Datensicht in KI-Projekten?

Die Integration von KI in Softwaresysteme bringt einzigartige Herausforderungen für das RE mit sich. Diese Herausforde-

Tabelle I
GRUPPIERUNG VON STUDIEN NACH KATEGORIE

Kategorie	IDs
Herausforderungen in RE für AI (RQ1)	[4], [7], [10]
Fallstudien und Interviews (RQ1, RQ2)	[6], [11]–[13]
Methoden und Frameworks (RQ2)	[15], [16], [18], [20], [21]
Systematische Literaturreviews (RQ3)	[3], [5], [8], [17], [19]

rungen umfassen die sich entwickelnde Natur von Anforderungen, datenzentrierte Probleme, Unsicherheit, Komplexität und ethische Probleme. ML-Systeme haben aufgrund der hohen Datenabhängigkeit Besonderheiten gegenüber traditionellen Softwaresystemen. Die Unvorhersehbarkeit von ML-Modellen ist ein weiterer, herausfordernder Faktor in KI-Projekten. [10]

• **Nichtdeterminismus:**

- ML-Systeme sind potentiell unvorhersehbar und nicht-deterministisch.
- Traditionelle RE-Methoden sind nicht ausreichend, um robuste ML-Systeme zu entwickeln und zu betreiben. [8]

• **Abhängigkeit von Daten:**

- Leistung und Verhalten von KI-Systemen sind abhängig von Qualität, Quantität und Relevanz der Daten.
- RE-Verfahren müssen Strategien für Datenakquise, -vorverarbeitung und -validierung umfassen. [12]

• **Dynamische Anforderungen:**

- KI-Systeme entwickeln sich stetig weiter, insbesondere ML-Systeme.
- Ein flexibles und anpassungsfähiges RE ist nötig. [6]

• **Integrierbarkeit:**

- Integration von KI-Komponenten mit traditionellen Systemen erfordert sorgfältige Berücksichtigung der Systemarchitektur.
- RE muss eine ganzheitliche Sicht auf das System ermöglichen.
- Ziel ist die nahtlose Integration bei minimalen Störungen sicherzustellen. [11]

• **Ethische Bedenken:**

- KI-Projekte haben ethische und gesellschaftliche Auswirkungen.
- Datenschutz, Fairness, Transparenz und Verantwortlichkeit müssen gewährleistet werden. [4]

Um diese Herausforderungen anzugehen, schlagen Belani et al. die RE4AI (Requirements Engineering for Artificial Intelligence)-Taxonomie vor. Diese Taxonomie bietet eine umfängliche Sicht auf die Herausforderungen, die sich dem RE bei der Entwicklung von KI-basierten Systemen stellen. Sie fokussiert sich auf Rollen, Prozesse und Artefakte im Zusammenhang mit RE-Aktivitäten. Die Taxonomie kategorisiert Herausforderungen in die Bereiche Daten, Modelle und Systeme. Die Bedeutung eines angepassten RE-Prozesses wird betont. Insbesondere in Bereichen wie dem Gesundheitswesen, kann die Taxonomie nützlich sein. Dort kann sie

Herausforderungen wie Verschiebungen in der Datenverteilung und unsichere Fehlermodi angehen. Laut Belani et al. steht insbesondere die Datensicht im Fokus der Industrie. Die RE4AI-Taxonomie soll genutzt werden, um im Folgenden die datenbezogenen Probleme zu adressieren. Dabei wird die Abgrenzung von Daten, Modellen und Systemen berücksichtigt. [10]

RQ2.1: Was muss innerhalb der RE-Aktivitäten berücksichtigt werden, um dessen Herausforderungen in KI-Projekten zu bewältigen?

Folgende datenbasierte Herausforderungen entstehen laut der RE4AI-Taxonomie bei den konkreten RE-Aktivitäten [10]:

Tabelle II
ÜBERSICHT DER DATENBASIERTEN HERAUSFORDERUNGEN IN DEN EINZELNEN RE-AKTIVITÄTEN NACH BELANI ET AL. [10]

RE-Aktivität	Herausforderungen
Ermittlung	<ul style="list-style-type: none"> • Verfügbarkeit von (großen) Datensätzen. • Weiterentwicklung des Anforderungsanalytens.
Analyse	<ul style="list-style-type: none"> • Unausgewogene Datensätze, Datensilos. • Rolle: Datenwissenschaftler erforderlich.
Spezifikation	<ul style="list-style-type: none"> • Datenkennzeichnung ist kostspielig und erforderlich. • Rolle: Dateningenieur erforderlich.
Validierung	<ul style="list-style-type: none"> • Kritische Analyse der Trainingsdaten. • Verschiedene Datenabhängigkeiten.
Management	<ul style="list-style-type: none"> • Experimentmanagement. • Keine ausgereifte GORE-ähnliche Methode.
Dokumentation	<ul style="list-style-type: none"> • Visualisierung von Daten und Modellen. • Rolle: Forschungswissenschaftler nützlich.
Allgemein (Alle o. g.)	<ul style="list-style-type: none"> • Datenschutz und Datensicherheit. • Datenabhängigkeiten.

Im Folgenden werden die einzelnen RE-Aktivitäten und dessen Herausforderungen, welche in Tabelle II dargestellt sind untersucht. Dazu werden Lösungsansätze aus der Literatur extrahiert. Diese werden dann den RE-Aktivitäten nach Belani et al. [10] zugeordnet und in Tabelle III dargestellt. Die hierzu benutzten Papiere stammen aus den Kategorien “Fallstudien und Interviews” und “Methoden und Frameworks” aus der Literaturrecherche (siehe Tabelle I). Somit wird eine strukturierte Übersicht der Lösungsansätze aufgezeigt.

RQ2.2: Welche Methoden, Frameworks und Tools existieren hierzu?

Die Lösungsansätze aus Tabelle II des vorherigen Abschnitts enthalten einige Papiere, welche auch konkrete Methoden und Frameworks beschreiben. Diese Papiere tangieren

Tabelle III
ÜBERSICHT ÜBER DIE ERFASSTEN DATENGETRIEBENEN LÖSUNGSANSÄTZE IM RAHMEN DER RE-AKTIVITÄTEN

RE-Aktivität	Lösungsansätze
Ermittlung	<ul style="list-style-type: none"> • Identifizierung relevanter Datenquellen, hochwertiger Daten und systematische Modellierung von Datenanforderungen [6], [15]. • Erfassung und Umsetzung von nicht-funktionalen Anforderungen (NFRs) [11]. • Zielgerichtete Datensammlung und Integration verfügbarer Datensätze [12]. • Ableitung von Qualitätsmerkmalen basierend auf Kundenbedürfnissen und Entwicklung von Rahmenwerken zur Erfassung von Datenanforderungen [13], [16].
Analyse	<ul style="list-style-type: none"> • Definition von Leistungsmaßen zur Erleichterung von Stakeholder-Diskussionen und Behandlung modellspezifischer Bedenken [6]. • Identifikation und Behebung von Ungleichgewichten in Datensätzen und Datensilos [12], [15].
Spezifikation	<ul style="list-style-type: none"> • Spezifikation von Datenanforderungen, Vorhersagekraft, Datenoperationen und Infrastrukturbedenken [6], [18]. • Erstellung von Vorlagen für überprüfbare Datenanforderungen und indessen Berücksichtigung von Datenkennzeichnung [11], [16]. • Abdeckung von Datenkennzeichnung und Feature-Engineering mithilfe von Software-Tools [12]. • Spezifikation von Datenmerkmalen in Datensätzen [13], [15].
Validierung	<ul style="list-style-type: none"> • Kontinuierliche Analyse von Betriebsdaten, regelmäßiges Neutraining von ML-Modellen und strukturierter Ansatz zur Definition und Analyse von Elementen der ML-Systeme [6], [15]. • Risikoeinschätzung der Datenqualität für Features von ML-Systemen und mithilfe dessen Priorisierung der Datenmerkmale [20]. • Einbeziehung der Analyse von Trainingsdaten, Verwaltung von Datenabhängigkeiten und evidenzbasierte Bewertung der Datenqualität [11], [13], [16]. • Integration von Modelltraining, Bewertung und Verwaltung von Datenabhängigkeiten in bestehenden Softwareentwicklungsprozess [12].
Management	<ul style="list-style-type: none"> • Kontinuierliche Überwachung und Anpassung von ML-Systemen, Abstimmung zwischen Geschäfts- und ML-Zielen [6], [18]. • Management von ML-Experimenten und dadurch Einblicke in NFRs [11]. • Integration von ML in Software-Engineering-Praktiken [12]. • Strukturierte Methoden zur Verwaltung von Datenanforderungen und klare Qualitätsmerkmale für ML-Systeme [13], [15].
Dokumentation	<ul style="list-style-type: none"> • Klare Dokumentation von Datenaufbereitungsschritten, Modellerklärungen und Schwerpunkt auf Vorhersagevisualisierung [6]. • Unterstützung der Dokumentation durch Beitrag von Forschungswissenschaftlern [11]. • Entwicklung von integrierten Entwicklungsumgebungen und Dashboards zur Visualisierung von ML-Modellen [12]. • Visualisierung ausgewählter Datenmerkmale und ihrer Hierarchie, Bedarf an Dokumentationstechniken [13], [15].
Allgemein (Alle o. g.)	<ul style="list-style-type: none"> • Schwerpunkt auf rechtlichen und regulatorischen Einschränkungen, Datensicherheit, Datenschutz und Ethik in der Datenverarbeitung [6], [12], [18]. • Spezifikation von Dateneigenschaften einschließlich Datenschutz und Sicherheit, Integration dieser Aspekte in die Entwicklung von ML-Systemen [13].

alle RE-Aktivitäten, wobei sie sich auf einzelne Aspekte fokussieren. Hierbei zeigen nur vier von acht referenzierten Papieren Methoden oder Frameworks auf.

Um die Frage RQ2.2 zu beantworten, werden im Folgenden bestehende Ansätze aus der Forschung aufgezeigt. Dazu werden diese zuerst zusammenfassend dargestellt. Im Anschluss wird eine Übersicht über die Relevanz der Methoden für die einzelnen RE-Aktivitäten gegeben. Zudem werden Gemeinsamkeiten der Methoden aufgezeigt.

Multi-layered Collaborative Framework [16]

Die Forschung von Dey und Lee stellt ein neuartiges mehrschichtiges kollaboratives Rahmenwerk vor. Dieses soll die spezifischen Anforderungen an ML-Projekte angehen. Es entstand aus der Herausforderung, ML-Komponenten in sicherheitskritische Systeme zu integrieren. Die Autoren berufen sich auf das für KI angepasste Twin Peaks Modell (siehe Abbildung 1) mit der Aussage, dass Daten maßgeblich die Umsetzung von KI-Systemen steuern.

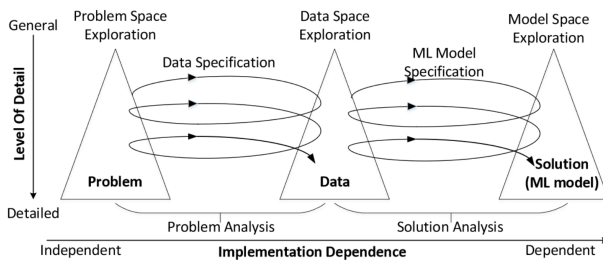


Abbildung 1. Twin Peaks Modell für KI [16]

Das Rahmenwerk ist in drei Schichten gegliedert: die Problemschicht, die Datenschicht und die Evidenzschicht. Die Schichten bauen aufeinander auf und behandeln verschiedene Aspekte der Ermittlung von Anforderungen:

- **Problemschicht:** Diese Schicht konzentriert sich auf das Verständnis des breiten Kontexts und spezifischer Probleme, die das ML-System zu lösen versucht. Dabei werden drei Dimensionen betrachtet: den operativen Bereich, hochrangige Anforderungen und die damit verbundene Risiken. So wird der Grundstein für datenbezogene Entscheidungen gelegt. Als Output produziert die Problemschicht u.a. high-level Requirements, Konzepte der aufgabenspezifischen Domänen für das ML und Risikofaktoren auf Domänen- sowie Systemlevel im Rahmen der sogenannten "Problem Space Exploration Summary".
- **Datenschicht:** In dieser Schicht werden folgende Schritte zum Umgang mit Datenunsicherheit in ML-Projekten dargestellt:
 - **Zielgerichtete Ermittlung und Analyse von Datenanforderungen:** Das Ziel ist das Sicherstellen von Datensicherheit. Dazu wird eine Dekomposition von domänen- und datenspezifischen Risikofaktoren angefertigt. Von dieser wird dann eine Zieldefinition erstellt, woraus sich konkrete Aufgaben ableiten. Somit sollen

klare Richtlinien für Datenqualität, Datenkennzeichnung und korrekte Repräsentationen der Domänen und Risiken geschaffen werden.

- **Spezifikation der Datenanforderungen:** Basierend auf den zuvor ermittelten Aufgaben werden Datenanforderungen spezifiziert. Eine Vorlage für die Anforderungsspezifikation umfasst den Typ, die Kennung, die Beschreibung, erforderliche Beweise und Akzeptanzkriterien für diese Beweise.

Diese Schritte sollen sicherstellen, dass die in ML-Modellen verwendeten Daten zuverlässig und repräsentativ sind. Als Output produziert diese Schicht die Datenanforderungsspezifikation.

- **Evidenzschicht:** Die letzte Schicht beinhaltet die Konsolidierung und Bewertung der in den o.g. Schichten gesammelten Beweise. Der Schwerpunkt liegt hier auf der Bewertung der Beweise im Hinblick auf die Datenanforderungen. Die Bewertung wird durch die Perspektiven von ML- und Systemexperten unterstützt. Das Ziel der Schicht ist die Validierung der Trainingsdaten und deren Ausrichtung an die Anforderungen des Systems. Der Output dieser Schicht umfasst den "Data Uncertainty Evaluation Sheet", welcher die konkreten Datenanforderungen sowie die Beweise und Argumente dafür umfasst.

Perspective-Based Specification [18]

Villamizar et al. schlagen einen perspektivenbasierten Ansatz zur Spezifikation von ML-Systemen vor. Dieser Ansatz basiert auf dem Verständnis, dass ML-Systeme komplex und facettenreich sind. Selten sind die Spezifikationen bei der Entwicklung solcher Systeme präzise genug. Der Ansatz baut auf fünf Schlüsselperspektiven auf: Ziele, Benutzererfahrung (UX), Infrastruktur, Modell und Daten. Villamizar et al. haben diese aus einer Literaturrecherche und aus Praxiserfahrungen abgeleitet:

- **Zielsetzung:** Konzentriert sich auf die Endziele des ML-Systems. Diese Perspektive ist entscheidend für die Ausrichtung der Funktionalität des Systems auf Geschäftsziele.
- **Benutzererfahrung (UX):** Stellt sicher, dass das System intuitiv und benutzerfreundlich ist. Hier wird betrachtet, wie das System mit seinen Benutzern interagiert.
- **Infrastruktur:** Bezieht sich auf die grundlegenden Hardware- und Softwareanforderungen des ML-Systems. Dieses Feld ist entscheidend für die Umsetzung der technischen Grundlage des Systems zur Entwicklungs- und Laufzeit.
- **Modell:** Konzentriert sich auf die ML-Algorithmen und ihre Leistung. Diese Perspektive betrachtet die Effektivität und Effizienz der eingesetzten Modelle.
- **Daten:** Beinhaltet die Qualität, Beschaffung und Verwaltung der verwendeten Daten. Diese sind für das Training der ML-Modelle wichtig. So soll die Zuverlässigkeit und Repräsentativität der Daten sichergestellt werden. Ethik und Privatsphäre sind dabei zu berücksichtigen.

Jede Perspektive enthält Aufgaben, die von verschiedenen Akteuren wie Geschäftsinhabern, Designern, Softwareentwicklern und Datenwissenschaftlern angegangen werden. Dieser kollaborative Ansatz soll eine ganzheitliche Sicht auf das System gewährleisten. Somit werden die für ML entscheidende Aspekte systematisch angegangen.

Risk-Based Data Validation Approach (RBDVA) [20]

Im Kontext von ML-basierten Anwendungssystemen präsentieren Foidl und Felderer den sogenannten risikobasierten Datenvalidierungsansatz (Risk-Based Data Validation Approach (RBDVA)). Diese Methode ist eine Reaktion auf die Herausforderungen der Datenvalidierung in ML-basierten Softwaresystemen, bei der das enorme Volumen und die Komplexität der Daten herkömmliche Validierungsmethoden ungeeignet machen.

- **Risikoeinschätzung in Datenmerkmalen:** Zentral für diesen Ansatz ist die Schätzung des Risikos schlechter Datenqualität für einzelne Features eines ML-Systems. Dieses Risiko wird als Funktion zweier Hauptfaktoren definiert: der Wahrscheinlichkeit, dass das Feature von niedriger Datenqualität ist sowie der Auswirkung dieser niedrigen Qualität auf die Leistung des ML-Modells. Durch die Quantifizierung dieser Risiken ermöglicht der RBDVA eine effektive Validierungsstrategie.
- **Priorisierung und Strenge der Datenvalidierung:** Der RBDVA schlägt eine strukturierte Methode vor, um Merkmale basierend auf ihrem bewerteten Risikoniveau zu priorisieren. Es erfolgt eine Einteilung in niedriges, mittleres oder hohes Risiko. Diese Priorisierung unterstützt die Softwareentwickler bei der Implementierung von Datenvalidierungsmaßnahmen. Zusätzlich schlägt der Ansatz vor, die Intensität der Datenvalidierungsmaßnahmen basierend auf dem Risikoniveau zu variieren, um so einen ausgewogenen und ressourceneffizienten Validierungsprozess zu gewährleisten.
- **Unterstützende Tools für die Datenvalidierung:** Die Autoren verweisen auf praktische Werkzeuge und Bibliotheken, die die Datenvalidierung in auf ML basierenden Systemen erleichtern. Dazu gehören:
 - **engarde:** Ein Python-Paket für defensive Datenanalyse.
 - **voluptuous:** Eine Python-Bibliothek für Datenvalidierung.
 - **great expectations:** Eine Python-Bibliothek für die Validierung, Dokumentierung und Profilierung von Daten.Es wird jedoch angemerkt, dass dem Feld immer noch ausgereifte Datenvalidierungswerkzeuge und etablierte Best Practices im Vergleich zu traditionellen Softwareprojekten fehlen.

Zusammenfassend führt Foidl und Felderers RBDVA einen neuen und pragmatischen Ansatz zur Datenvalidierung in ML-Systemen ein. Durch die Berücksichtigung von Risiken und der Relevanz von Features ermöglicht diese Methode einen effizienteren Validierungsprozess. Die Einbeziehung unterstüt-

zender Werkzeuge und Bibliotheken unterstreicht die Praktikabilität dieses Ansatzes. So bietet der RBDVA eine Grundlage für weitere Forschung und Entwicklung.

Two-Layer Data Modeling Method [15]

In ihrem Papier schlagen Shao und Wang einen Ansatz zur Modellierung von Datenanforderungen für ML-Systeme vor. Dabei wird folgende Struktur beschrieben:

- **Zweischichtige Struktur:** Die Methode umfasst zwei integrierte Schichten:
 - **Die untere Schicht - Lernkontext (Learning Context):** Diese Schicht bildet die Grundlage des Modells. Sie umfasst das Metadatenmodell. Dieses beschreibt die Elemente des ML-Systems, seine Umgebung und der Beziehungen zwischen den Elementen. So wird sichergestellt, dass alle relevanten Anforderungen an das ML-System im Datenmodellierungsprozess berücksichtigt werden.
 - **Die obere Schicht - Eigenschaftsbasierte Spezifikationen (Property-based Specifications):** Aufbauend auf dem Lernkontext beinhaltet diese Schicht eine Reihe von eigenschaftsbasierten Spezifikationen. Sie nutzt die Informationen aus der unteren Schicht, um spezifische Datenanforderungen zu definieren. Das umfasst die Spezifikation der benötigten Datentypen, ihrer Eigenschaften sowie deren Verarbeitung und Nutzung im System.
- **Tool für die Datenmodellierung:** Zur Erleichterung der Anwendung dieser Methode entwickelten Shao und Wang ein unterstützendes Tool mit folgenden Bestandteilen:
 - **Interaktive Schnittstelle:** Die Schnittstelle ist dafür verantwortlich, die Benutzer durch den Datenmodellierungsprozess zu führen. Es hilft dabei, Merkmale und Eigenschaften des vom Benutzer benötigten Datensatzes zu erfassen und das aktuelle Datenmodell anzuzeigen. Es umfasst Module zur Merkmalsauswahl, zur Visualisierung ausgewählter Merkmale und zu Spezifikationen.
 - **Wissensdatenbank:** Diese Komponente speichert die vorab erstellten Informationen zur Modellierung von Datenanforderungen hierarchisch mittels der JavaScript Object Notation (JSON). Sie erhält Eingaben von der interaktiven Schnittstelle bezüglich der Merkmalsauswahl. Die Wissensdatenbank bietet anschließende Unterstützung zur Modellierung, um ein kohärentes und umfassendes Datenmodell zu gewährleisten.
- **Fallstudie - Autonomes Fahrsystem:** Die Autoren demonstrieren die Anwendung ihrer Methodik anhand einer Fallstudie im Bereich der autonomen Fahrsysteme. Die Fallstudie veranschaulicht, wie die Methodik und das Werkzeug Benutzer bei der Erstellung spezifischer Datenmodellen anleiten, was zu detaillierten Datenanforderungsspezifikationen führt. Genannte Aspekte sind die Notwendigkeit beschrifteter Daten, Datenaktualisierungszeiträume und die Einbeziehung verschiedener Datenty-

pen (Video, Binär, XML, Text) mit zuvor spezifizierten Eigenschaften.

Shao und Wangs Methode zur Modellierung von Datenanforderungen bietet einen strukturierten Ansatz zur Definition von Datenmodellen im Kontext von ML-Systemen. Unterstützt durch ein hierfür entwickeltes Werkzeug erleichtert diese Methode das Verständnis von Datenanforderungen. Die Methode hat zum Ziel, ML-Systeme auf einer soliden Grundlage von gut spezifizierten und relevanten Daten aufzubauen.

Goal-Centralized Metamodel Based Requirements Integration [21]

Tun et al. stellen einen Ansatz vor, welcher die Integration von Zielen und NFAs in ML-Systeme ermöglichen soll. Dazu wird ein Metamodell geschaffen, das zu den Artefakten der Entwicklung konsistent ist. Das Metamodell berücksichtigt die verschiedenen Anforderungen an das System. Weiterhin wird eine zielzentrierte Herangehensweise gewählt. Das Metamodell wird mithilfe der Unified Modeling Language (UML) modelliert. Hierzu muss eine Erweiterung in Form eines UML Profils erstellt werden. Im Papier werden folgende Metamodelle im Rahmen einer Case Study zum autonomen Fahren erfasst:

- **AI Project Canvas Model:** Stellt die kritischen Komponenten eines ML-Systems dar. Diese Komponenten bieten die Methoden, um Ziele des Projekts zu evaluieren. Es unterstützt so die Entscheidungsfindung.
- **ML Project Canvas Model:** Beleuchtet die Komponenten der Anforderungen für ML-Systeme. Diese Komponenten bieten die Funktion, Methoden und Metriken aus den gegebenen Datenquellen zu extrahieren.
- **Goal Model:** Stellt funktionale und nicht-funktionale Ziele dar. Das Goal Model zeigt die Beziehungen der verschiedenen Elemente des ML-Modells.
- **SysML Model:** Kombiniert RE und Traceability von SysML mit dem Gesamtziel. Das SysML Model umfasst die Spezifikationen des Systems und dessen Funktionalität, welches durch die ML-Modelle gegeben ist.
- **Safety Case Model:** Beschreibt das Sicherheitsziel zur Integration der Modelle, die auf der Goal Structuring Notation (GSN) [30] basieren.
- **STAMP Model:** Das Systems-Theoretic Accident Model and Processes (STAMP) Model stellt im Kontext des Metamodells dar, wie Gefahren und Unfälle analysiert werden können, um die Ziele des Systems zu erreichen.

Die genannten Modelle werden in das UML-Gesamtmodell integriert. Dazu werden die einzelnen Modelle gruppiert und farbig markiert. Im Zentrum des UML-Modells steht das Ziel der Anwendung. Die Beziehungen zwischen den einzelnen Elementen werden via der UML dargestellt. Dabei werden unter anderem Kardinalitäten und Kompositionen verwendet. Insgesamt soll durch diesen Ansatz die Konsistenz und Traceability im Rahmen des RE für ML-Systeme gesteigert werden.

Einordnung der Methoden

Bei der Betrachtung der vorgestellten Methoden muss beachtet werden, dass die einzelnen Ansätze verschiedene Ziele verfolgen. Insbesondere der Spezialisierungsgrad der einzelnen Methoden ist unterschiedlich. Das Multi-layered Collaborative Framework nach Dey und Lee [16] sowie die Perspective-based Specification nach Villamizar et al. [18] sind allgemein auf datenintensive KI- bzw. ML-Projekte anwendbar. Hingegen ist der RBDVA von Foidl und Felderer fokussiert auf die Validierung von Daten, Shao und Wangs Two-Layer Data Modeling Method [15] auf die Datenmodellierung. Der Ansatz der Goal-Centralized Metamodel Based Requirements Integration schafft ein konkretes UML-Modell, um die einzelnen Modellierungsperspektiven in ML-basierten Projekten zu integrieren und eine entsprechende Gesamtsicht zu schaffen [21].

Unterstützung der RE-Aktivitäten mittels der Methoden

Die im Rahmen der RQ2.1 behandelten Herausforderungen der RE-Aktivitäten werden durch die vorgestellten Methoden angegangen. In der Abbildung 2 wird das RE und der Entwicklungsprozess im Rahmen von KI-Projekten dargestellt. Hierbei werden die behandelten Methoden den einzelnen RE- und Entwicklungsaktivitäten zugeordnet. Die Zuordnung ist als eine Unterstützung der einzelnen Aktivitäten zu sehen und ersetzt diese nicht.

Das Multi-Layered Collaborative Framework bietet die breiteste Unterstützung für die RE-Aktivitäten von den untersuchten Ansätzen. Durch die Problem- und Datenschicht wird die Ermittlung und Analyse unterstützt. Die Outputs der einzelnen Schichten, die "Produced Traceable Artifacts", sind der Spezifikation und Dokumentation zuzuordnen. In der Evidenzschicht werden die Artefakte der vorigen Schichten validiert. Das Management wird mittels des strukturierten Vorgehens und durch die Erstellung von verfolgbaren Dokumenten unterstützt. [16]

Die Perspective-Based Specification und die Goal-Centralized Metamodel Based Integration unterstützen die Analyse, Spezifikation und Dokumentation. Dazu zeigt die Perspective-Based Specification Templates auf, welche die Herausforderungen der einzelnen Perspektiven angeht. Hingegen gibt die Goal-Centralized Metamodel Based Integration die Modellierung mittels UML vor, welche die Ziele und NFAs in einem einzelnen Artefakt mit den bestehenden Modellen zusammenführt. [18], [21]

Die Two-Layer Data Modeling Method unterstützt nicht nur das RE, sondern auch die Entwicklung im KI-Rahmen. Es werden Datenanforderungen spezifiziert und in der Knowledge Base dokumentiert. Des Weiteren werden auch Label im Datenmodell festgelegt, womit das Data Labeling unterstützt wird. Der RBDVA ist der Validierung im RE zuzuordnen, da er sich ausschließlich auf die Datenvalidierung fokussiert. [15], [18]

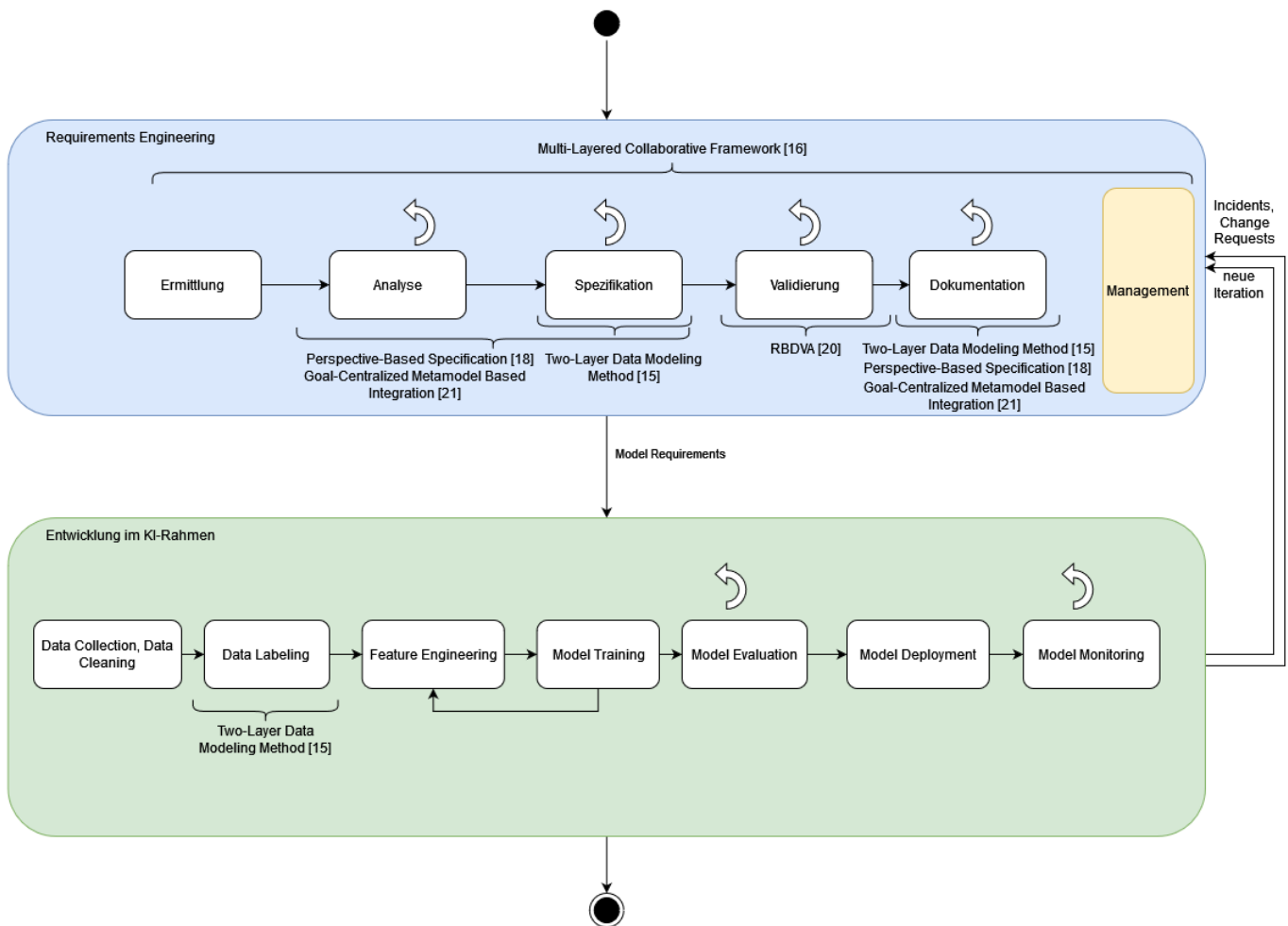


Abbildung 2. RE und Entwicklungsprozess im KI-Kontext. Zuordnung der Ansätze zu den RE-Aktivitäten. Gesamte Darstellung i.A.a. Glinz et al. [1]. Entwicklung im KI-Rahmen (grün) i.A.a. Amershi et al., dabei stellen die großen Pfeile Rücksprünge an vorige Aktivitäten dar [12].

Gemeinsamkeiten der Methoden

Die Übersicht der vorgestellten Methoden, Frameworks und Tools zeigt, dass diese unterschiedliche Herangehensweisen und Schwerpunkte in der Bewältigung der Herausforderungen von RE in KI-Projekten haben. Trotz der Unterschiede gibt es folgende Gemeinsamkeiten zwischen den Methoden:

- **Schichtung und Modularität:** Sowohl das Multi-layered Collaborative Framework als auch die Two-Layer Data Modeling Method sind in Schichten aufgeteilt. Dies ermöglicht eine systematische Behandlung verschiedener Aspekte von Datenanforderungen und deren Validierung. Durch diese Strukturierung werden die komplexen Gegebenheiten in kleinere Teile zerlegt. Somit können spezifische Anforderungen einfacher aufgeteilt werden. Zudem können einzelne Stakeholder gezielter eingesetzt werden, indem sie nur in relevanten Teilen des Projekts eingesetzt werden. [15], [16]
- **Nutzen von verschiedenen Perspektiven:** Die Goal-Centralized Metamodel Based Requirements Integration und die Perspective-Based Specification betrachten das

System aus verschiedenen Perspektiven. So versuchen diese eine ganzheitliche Sicht auf das zu entwickelnde System zu ermöglichen. Dadurch können die Anforderungen verschiedener Stakeholder und Interessensgruppen berücksichtigt werden. [18], [21]

- **Risikomanagement durch Datenqualität:** Der RBDVA betont Risikobewertung und -management in Bezug auf Daten in ML-Projekten. Dieses Prinzip findet sich auch im Multi-layered Collaborative Framework und der Perspective-Based Specification wieder. Der Fokus liegt auf der Minimierung von Risiken, die aus schlechter Datenqualität resultieren. [18], [20]
- **Kollaborative und interdisziplinäre Ansätze:** Der Perspective-Based Specification Ansatz betont die Notwendigkeit einer Zusammenarbeit zwischen verschiedenen Stakeholdern und Disziplinen. Dieser Aspekt wird auch im Multi-layered Collaborative Framework berücksichtigt. Hier wird die effektive Umsetzung von ML-Projekten durch eine enge Zusammenarbeit zwischen Datenwissenschaftlern, Softwareentwicklern, Domänenexperten und Endnutzern erwähnt. [16], [18]

- **Flexibilität und Anpassungsfähigkeit:** Trotz ihrer spezifischen Schwerpunkte bieten die Methoden ein gewisses Maß an Flexibilität. Somit sind die Methoden auf verschiedene Domänen und Anforderungen anwendbar. Dies ist besonders wichtig im dynamischen und vielfältigen Kontext von KI.

Die dargelegten Ansätze bieten einige Methoden, um die Herausforderungen bei der Entwicklung von KI-basierten Softwaresystemen zu bewältigen. Dabei behandeln diese Lösungsansätze, welche auch von anderen Papieren erfasst wurden (siehe Tabelle III).

Um die Praktikabilität der Methoden zu überprüfen, müssen diese in konkreten Projekten eingesetzt und dessen Nutzen evaluiert werden. Trotz der Gemeinsamkeiten hängen die Methoden vom Projektkontext und den Anforderungen ab. In welchem Kontext einzelne Methoden angewandt werden sollte, ist noch nicht klar. Dey und Lee's Methode bedient hier jedoch die spezifischen Anforderungen von sicherheitskritischen ML-Systemen [16]. Zwei weitere Methoden wurden im Kontext des autonomen Fahrens mittels entsprechenden Case Studies geprüft [15], [21].

RQ3: Für welche RE-Aktivitäten fehlen noch Methoden in Bezug auf KI-Projekte?

Basierend auf der Analyse der SLRs (siehe Tabelle I) wurden folgende Lücken in den RE-Aktivitäten für KI-Projekte identifiziert:

- **Integration von KI in Softwaresysteme:** Die Integration von KI-Komponenten in Softwaresysteme bringt erhebliche Herausforderungen mit sich, vor allem in Bezug auf Datenabhängigkeiten. Dies hängt mit der Notwendigkeit zusammen, angemessene Anforderungsspezifikationen für ML-Systeme zu entwickeln. Das wird häufig durch die Intransparenz von ML-Modellen und daraus resultierenden Inkonsistenzen bei Eingaben und Ausgaben erschwert. Darüber hinaus ist die Anpassung traditioneller Modellierungsmethoden für KI-Systeme wesentlich. [8], [19]
- **NFRs, Daten- und Ethikanforderungen:** Anforderungen im Rahmen der Datensammlung und ethische Bedenken erfordern innovative Ansätze im RE, insbesondere im Hinblick auf gesetzliche und soziale Richtlinien. Außerdem sind NFRs wie Transparenz, Vertrauen, Datenschutz, Sicherheit, Zuverlässigkeit und Sicherheit in KI-Systemen wichtig und erfordern spezielle RE-Ansätze. [19]
- **Zusammenarbeit zwischen Fachleuten:** Die Kollaboration zwischen Softwareentwicklern und Datenwissenschaftlern ist ein ausschlaggebender Faktor in KI-Projekten. Eine Einbindung von Datenwissenschaftlern bei der Identifizierung relevanter Daten ist ebenfalls von Bedeutung. Diese Aspekte müssen neuartige RE-Methoden angehen. [19]
- **Anpassungen von RE-Aktivitäten:** Die Abhängigkeiten der Anforderungen von verfügbaren Daten in ML-Systemen erfordern Anpassungen in den RE-Aktivitäten.

Es besteht eine Forschungslücke bei der Überwachung von ML-Systemen. Zudem fehlen validierte Techniken für die Spezifikation von ML-Systemen. [3], [8], [17]

Alle identifizierten Lücken weisen auf die Notwendigkeit eines RE hin, welches für KI-Projekte angepasst ist. Die zuvor vorgestellten Methoden und Lösungsansätze gehen einen Schritt in diese Richtung, jedoch bedarf es einer stetigen Weiterentwicklung besonders aufgrund der rasanten Evolution von KI-Systemen. Weiterhin fehlen noch konkrete Erfahrungen in der Praxis, welche die Wirksamkeit bestehender Methoden bestätigen könnten.

IV. DISCUSSION

In Abschnitt II wurde die Literaturrecherche nach Kitchenham et al. durchgeführt [9]. Dabei wurden Suchanfragen definiert, mit denen die entsprechenden Papiere erfasst wurden. Der Umfang der Literaturrecherche war somit durch die Suchbegriffe beschränkt. Insbesondere bei der Identifikation des Forschungsstands könnten veränderte Suchanfragen mehr Ergebnisse bieten, welche für die Forschungsfragen relevante Ansätze liefern. Das ist unter anderem bei der Beantwortung der RQ2.2 wichtig. Hier hätte man in der Literaturrecherche konkrete Suchbegriffe nutzen können. Konkrete Begriffe wären "Data", "Data modeling" und ("Data" AND "Method"). Insgesamt hätte eine Anpassung eine breitere Abdeckung der vorhandenen Methodik bewirken können.

Die Beantwortung der Forschungsfragen in Abschnitt III hat gezeigt, dass enorme Herausforderungen in KI-Projekten vorhanden sind. Insbesondere wird die Wichtigkeit der Datenperspektive in ML-Systemen betont, da diese die grundlegende Funktionalität eines solchen Systems ausmacht. Es existieren bereits Lösungsansätze, konkrete Frameworks und Methoden, welche diese Herausforderungen angehen. Diese sind jedoch noch sehr jung und wurden erst innerhalb der letzten vier Jahre entwickelt. Aufgrund der Zunahme der Relevanz von KI in Forschung und Industrie wächst auch der Bedarf an solchen neuartigen Ansätzen.

Die in RQ2.2 vorgestellten Ansätze sind noch nicht praxiserprobt und betonen weiteren Forschungsbedarf. Drei Papiere haben eine Case Study zur Methode durchgeführt: [15], [16], [21]. Die beiden anderen Papiere haben keine Case Study durchgeführt [18], [20]. Eine weitere Frage die sich hierzu stellt, ist: welche Methoden sind in welchen Projekten einzusetzen? Alle drei Case Studies der Papiere wurden im Bereich des autonomen Fahrens durchgeführt. Da dieser Bereich sicherheitskritische Aspekte behandelt, wurden diese in der Methodik berücksichtigt. Andere Bereiche bedürfen ebenfalls eines entsprechenden Fokus. Hier ist jedoch die Flexibilität der Methoden zu erwähnen, welche eine Anpassung an die Anforderungen ermöglicht.

Das Forschungsfeld der KI und die Anzahl der Anwendungen, welche KI-Technologien einbinden, wächst. Um das Potential der Technologie nachhaltig ausnutzen zu können, besteht ein Bedarf an verantwortungsvollen Ansätzen. Diese müssen ethische, rechtliche und insbesondere datenschutztechnische Bedenken berücksichtigen, da es massiver Datenmen-

gen bedarf, um KI-Modelle zu trainieren [4], [12]. Nur eine der fünf in RQ2.2 dargelegten Methoden berücksichtigt diese Bedenken [18]. Hier besteht ein Defizit, da Privatsphäre und Datenschutz in der Praxis aufgrund rechtlicher Rahmenbedingungen erfüllt sein muss.

Eine Herausforderung der Methoden ist die Notwendigkeit von Fachexperten. Dabei wird vorausgesetzt, dass Domänenexperten, ML-Experten, RE-Experten und Datenwissenschaftler vorhanden sind [16]. Diese müssen zusammenarbeiten und gemeinsame Ergebnisse liefern. Da jeder Fachexperte eine andere Sichtweise auf das Projekt hat, stellt sich die Herausforderung, diese zusammen zu bringen. Villamizar et al. versuchen diese Lücke durch die Betrachtung aus den fünf verschiedenen Perspektiven Zielsetzung, UX, Infrastruktur, Modell und Daten zu schließen [18]. Des Weiteren werden die Geschäftsziele und die Nutzersicht berücksichtigt. Bei der Entwicklung einer KI-basierten Software dürfen diese Aspekte nicht in den Hintergrund geraten, indem sich z.B. nur auf die Implementierung eines soliden ML-Modells konzentriert wird [6].

V. CONCLUSION

Die Analyse der Herausforderungen, deren Lösungsansätze und der bestehenden Methoden des RE für KI-Projekte zeigt, dass KI neue, komplexe Anforderungen an das RE stellt. Es erfordert spezialisierte Ansätze, die über das klassische RE in traditionellen Softwareprojekten hinausgeht. Die Erkenntnisse aus der Forschung bieten Einblicke in die Entwicklung und Notwendigkeit solcher angepassten RE-Methoden.

Die dynamische und sich weiterentwickelnde Natur von KI-Systemen, speziell im Bereich des ML, erfordert ein flexibles und anpassungsfähiges RE. Somit kann auch die Berücksichtigung der Interessen von Stakeholdern im Laufe der Lebenszeit solcher Softwaresysteme gewährt werden. Daten spielen hierbei die zentrale Rolle. So muss die Datenqualität, -quantität und -verarbeitung beachtet werden, um eine solide Grundlage für KI-Projekte zu gewährleisten. Das angepasste RE muss in der Lage sein, mit diesen datenzentrierten Herausforderungen umzugehen. Um die Risiken in den Projekten zu minimieren, muss die Sicherheit und Integrität der Daten sichergestellt werden.

Ein Kernaspekt ist die Berücksichtigung von ethischen, rechtlichen und sozialen Herausforderungen von KI-Projekten. Daraus ergibt sich eine ganzheitliche Betrachtung des RE, das diese Aspekte einbezieht. Eine Integration von KI-Komponenten in traditionelle Anwendungssysteme erfordert eine Abwägung von Systemarchitektur, Datenabhängigkeiten und Datenschutz.

Bei einer sich schnell entwickelnden Technologie besteht ein ständiger Forschungsbedarf. Die identifizierten Forschungslücken weisen darauf hin, dass trotz der aufgezeigten Fortschritte im RE für KI-Projekte noch Bedarf an weitergehender Forschung und Entwicklung besteht. Insbesondere die Integration von KI in bestehende Softwarelandschaften stellt eine Herausforderung dar. Auch die Zusammenarbeit zwischen Fachexperten muss im RE berücksichtigt werden.

Die vorgestellten Ansätze sind noch jung und müssen sich in der Praxis beweisen. Es bedarf eines nachhaltigen, holistischen RE-Ansatzes, um KI- und ML-Systeme erfolgreich entwickeln zu können.

RE für KI stellt ein dynamisches und interdisziplinäres Forschungsfeld dar. Die Entwicklung und praktische Nutzung von spezialisierten, nachhaltigen RE-Methoden wird für den Erfolg von KI-Projekten entscheidend sein. Dabei muss jedoch auch verantwortlich gehandelt werden. Nur so können auch die ethischen, rechtlichen und sozialen Bedenken berücksichtigt werden, um nachhaltig Nutzen aus der Technologie zu ziehen.

REFERENCES

- [1] M. Glinz, H. van Loenhoud, S. Staal, and S. Bühne, 'Handbook for the CPRE Foundation Level according to the IREB Standard', International Requirements Engineering Board, 2020.
- [2] M. W. Shen, 'Trust in AI: Interpretability is not necessary or sufficient, while black-box interaction is necessary and sufficient', 2022, doi: 10.48550/ARXIV.2202.05302.
- [3] N. Yoshioka, J. H. Husen, H. T. Tun, Z. Chen, H. Washizaki, and Y. Fukazawa, 'Landscape of Requirements Engineering for Machine Learning-based AI Systems', in 2021 28th Asia-Pacific Software Engineering Conference Workshops (APSEC Workshops), Taipei, Taiwan: IEEE, Dec. 2021, pp. 5–8. doi: 10.1109/APSECW53869.2021.00011.
- [4] W. Maalej, Y. D. Pham, and L. Chazette, 'Tailoring Requirements Engineering for Responsible AI', *Computer*, vol. 56, no. 4, pp. 18–27, Apr. 2023, doi: 10.1109/MC.2023.3243182.10.1109/APSECW53869.2021.00011.
- [5] Z. Pei, L. Liu, C. Wang, and J. Wang, 'Requirements Engineering for Machine Learning: A Review and Reflection', in 2022 IEEE 30th International Requirements Engineering Conference Workshops (REW), Melbourne, Australia: IEEE, Aug. 2022, pp. 166–175. doi: 10.1109/REW56159.2022.00039.
- [6] A. Vogelsang and M. Borg, 'Requirements Engineering for Machine Learning: Perspectives from Data Scientists', in 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW), Jeju Island, Korea (South): IEEE, Sep. 2019, pp. 245–251. doi: 10.1109/REW.2019.00050.
- [7] F. Khomh, B. Adams, J. Cheng, M. Fokaefs, and G. Antoniol, 'Software Engineering for Machine-Learning Applications: The Road Ahead', *IEEE Softw.*, vol. 35, no. 5, pp. 81–84, Sep. 2018, doi: 10.1109/MS.2018.3571224.
- [8] A. Gjorgjevikj, K. Mishev, L. Antovski, and D. Trajanov, 'Requirements Engineering in Machine Learning Projects', *IEEE Access*, vol. 11, pp. 72186–72208, 2023, doi: 10.1109/ACCESS.2023.3294840.
- [9] B. Kitchenham, O. Pearl Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, 'Systematic literature reviews in software engineering – A systematic literature review', *Information and Software Technology*, vol. 51, no. 1, pp. 7–15, Jan. 2009, doi: 10.1016/j.infsof.2008.09.009.
- [10] H. Belani, M. Vukovic, and Z. Car, 'Requirements Engineering Challenges in Building AI-Based Complex Systems', in 2019 IEEE 27th International Requirements Engineering Conference Workshops (REW), Jeju Island, Korea (South): IEEE, Sep. 2019, pp. 252–255. doi: 10.1109/REW.2019.00051.
- [11] K. M. Habibullah and J. Horkoff, 'Non-functional Requirements for Machine Learning: Understanding Current Use and Challenges in Industry', in 2021 IEEE 29th International Requirements Engineering Conference (RE), Notre Dame, IN, USA: IEEE, Sep. 2021, pp. 13–23. doi: 10.1109/RE51729.2021.00009.
- [12] S. Amershi et al., 'Software Engineering for Machine Learning: A Case Study', in 2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP), Montreal, QC, Canada: IEEE, May 2019, pp. 291–300. doi: 10.1109/ICSE-SEIP.2019.00042.
- [13] K. Nakamichi et al., 'Requirements-Driven Method to Determine Quality Characteristics and Measurements for Machine Learning Software and Its Evaluation', in 2020 IEEE 28th International Requirements Engineering Conference (RE), Zurich, Switzerland: IEEE, Aug. 2020, pp. 260–270. doi: 10.1109/RE48521.2020.00036.

- [14] H. H. Altarturi, K.-Y. Ng, M. I. H. Ninggal, A. S. A. Nazri, and A. A. Ghani, 'A requirement engineering model for big data software', in 2017 IEEE Conference on Big Data and Analytics (ICBDA), Kuching: IEEE, Nov. 2017, pp. 111–117. doi: 10.1109/ICBDA.2017.8284116.
- [15] W. Shao and X. Wang, 'A Data Modeling Method for Machine Learning Systems', in 2022 International Conference on Data and Software Engineering (ICoDSE), Denpasar, Indonesia: IEEE, Nov. 2022, pp. 1–5. doi: 10.1109/ICoDSE56892.2022.9972115.
- [16] S. Dey and S.-W. Lee, 'A Multi-layered Collaborative Framework for Evidence-driven Data Requirements Engineering for Machine Learning-based Safety-critical Systems', in Proceedings of the 38th ACM/SIGAPP Symposium on Applied Computing, Tallinn Estonia: ACM, Mar. 2023, pp. 1404–1413. doi: 10.1145/3555776.3577647.
- [17] H. Villamizar, T. Escovedo, and M. Kalinowski, 'Requirements Engineering for Machine Learning: A Systematic Mapping Study', in 2021 47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Palermo, Italy: IEEE, Sep. 2021, pp. 29–36. doi: 10.1109/SEAA53835.2021.00013.
- [18] H. Villamizar, M. Kalinowski, and H. Lopes, 'Towards Perspective-Based Specification of Machine Learning-Enabled Systems', in 2022 48th Euromicro Conference on Software Engineering and Advanced Applications (SEAA), Gran Canaria, Spain: IEEE, Aug. 2022, pp. 112–115. doi: 10.1109/SEAA56994.2022.00025.
- [19] K. Ahmad, M. Bano, M. Abdelrazek, C. Arora, and J. Grundy, 'What's up with Requirements Engineering for Artificial Intelligence Systems?', in 2021 IEEE 29th International Requirements Engineering Conference (RE), Notre Dame, IN, USA: IEEE, Sep. 2021, pp. 1–12. doi: 10.1109/RE51729.2021.00008.
- [20] H. Foidl and M. Felderer, 'Risk-based data validation in machine learning-based software systems', in Proceedings of the 3rd ACM SIGSOFT International Workshop on Machine Learning Techniques for Software Quality Evaluation, Tallinn Estonia: ACM, Aug. 2019, pp. 13–18. doi: 10.1145/3340482.3342743.
- [21] H. T. Tun, J. H. Husen, N. Yoshioka, H. Washizaki, and Y. Fukazawa, 'Goal-Centralized Metamodel Based Requirements Integration for Machine Learning Systems', in 2021 28th Asia-Pacific Software Engineering Conference Workshops (APSEC Workshops), Taipei, Taiwan: IEEE, Dec. 2021, pp. 13–16. doi: 10.1109/APSECW53869.2021.00013.
- [22] U.-E. Habiba, J. Bogner, and S. Wagner, 'Can Requirements Engineering Support Explainable Artificial Intelligence? Towards a User-Centric Approach for Explainability Requirements', in 2022 IEEE 30th International Requirements Engineering Conference Workshops (REW), Melbourne, Australia: IEEE, Aug. 2022, pp. 162–165. doi: 10.1109/REW56159.2022.00038.
- [23] S. Dey, 'Evidence-driven Data Requirements Engineering and Data Uncertainty Assessment of Machine Learning-based Safety-critical Systems', in 2022 IEEE 30th International Requirements Engineering Conference (RE), Melbourne, Australia: IEEE, Aug. 2022, pp. 219–224. doi: 10.1109/RE54965.2022.00027.
- [24] F. Ishikawa and Y. Matsuno, 'Evidence-driven Requirements Engineering for Uncertainty of Machine Learning-based Systems', in 2020 IEEE 28th International Requirements Engineering Conference (RE), Zurich, Switzerland: IEEE, Aug. 2020, pp. 346–351. doi: 10.1109/RE48521.2020.00046.
- [25] K. Ahmad, 'Human-centric Requirements Engineering for Artificial Intelligence Software Systems', in 2021 IEEE 29th International Requirements Engineering Conference (RE), Notre Dame, IN, USA: IEEE, Sep. 2021, pp. 468–473. doi: 10.1109/RE51729.2021.00070.
- [26] Umm-E-Habiba, 'Requirements Engineering for Explainable AI', in 2023 IEEE 31st International Requirements Engineering Conference (RE), Hannover, Germany: IEEE, Sep. 2023, pp. 376–380. doi: 10.1109/RE57278.2023.00058.
- [27] F. Dalpiaz and N. Niu, 'Requirements Engineering in the Days of Artificial Intelligence', IEEE Softw., vol. 37, no. 4, pp. 7–10, Jul. 2020, doi: 10.1109/MS.2020.2986047.
- [28] M. N. Al Islam, Y. Ma, P. Alarcon, N. Chawla, and J. Cleland-Huang, 'RESAM: Requirements Elicitation and Specification for Deep-Learning Anomaly Models with Applications to UAV Flight Controllers', in 2022 IEEE 30th International Requirements Engineering Conference (RE), Melbourne, Australia: IEEE, Aug. 2022, pp. 153–165. doi: 10.1109/RE54965.2022.00020.
- [29] K. Ronanki, 'Towards an AI-centric Requirements Engineering Framework for Trustworthy AI', in 2023 IEEE/ACM 45th International Conference on Software Engineering: Companion Proceedings (ICSE-Companion), Melbourne, Australia: IEEE, May 2023, pp. 278–280. doi: 10.1109/ICSE-Companion58688.2023.00075.
- [30] A. Groza and N. Marc, 'Consistency checking of safety arguments in the Goal Structuring Notation standard', in 2014 IEEE 10th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj Napoca, Romania: IEEE, Sep. 2014, pp. 59–66. doi: 10.1109/ICCP.2014.6936981.

APPENDIX

Tabelle IV
INKLUDIERTE QUELLEN

ID	Thema	Ausschlaggebender Inhalt	Relevanz
[3]	RE Landscape for ML-based AI Systems	Systematic Literature Review zu RE und ML. Fokus auf Big-Data.	2
[4]	Tailoring RE for Responsible AI	Anpassung von RE für Responsible AI.	1
[5]	RE for ML: A Review	Systematic Literature Review zu RE Herausforderungen in AI-Projekten.	2
[6]	Data Scientists' Perspective on RE for ML	Interview Studie über RE-Aktivitäten in ML-Projekten aus Data Science Sicht.	3
[7]	Key Challenges in SE for AI	Schlüsselherausforderungen in SE für AI.	2
[8]	Overview: RE for AI Challenges	Umfassende Übersicht zu RE für AI Herausforderungen.	3
[10]	RE Challenges in AI Systems	Herleitung einer Taxonomie: RE-Aktivitäten und Herausforderungen in AI-Projekten.	2
[11]	NFR in AI Projects	Qualitative Interviewstudie zu NFR in AI-Projekten.	2
[12]	SE for ML: A Case Study	Fallstudie bei Microsoft zu SE für ML.	3
[13]	Quality Evaluation for ML Software	RE-driven Evaluation der Qualität von ML-Systemen.	1
[15]	Data Modeling for ML Systems	Entwicklung einer Data Requirements Modeling Methode für ML-Systeme.	3
[16]	Collaborative Framework for ML-based Systems	Entwicklung eines kollaborativen Frameworks für sicherheitskritische ML-Systeme.	3
[17]	Systematic Mapping: RE for ML	Systematic Mapping Study zu RE für AI.	3
[18]	Specifying ML-Enabled Systems	Perspective-Based Ansatz für Spezifikation von ML-Systemen	3
[19]	RE for AI Systems: A SLR	Systematic Literature Review zu RE für AI.	2
[20]	Risk-based data validation for ML	Risikobasierter Ansatz zur Priorisierung von Merkmalen bei der Datenvalidierung in ML-basierten Softwaresystemen.	3
[21]	Metamodel Based Requirements Integration	Zielzentrierter Ansatz, um ein umfassendes Metamodell für ML-Projekte zu schaffen.	3

Tabelle V
EXKLUDIERTER QUELLEN

ID	Thema	Grund für Exklusion
[22]	RE Support for XAI	Keine datengetriebene Methode, Fokus auf Nutzerzentrierung.
[23]	Data Requirements Engineering	Vorarbeit zu [16] ohne die in späteren Studien durchgeführte Verfeinerung.
[24]	Requirements Engineering for ML Systems	Kein Fokus auf Daten, Entwicklung von GORE-MLOps Ansatz.
[25]	Human-centric RE for AI Systems	Keine Case Study, kein Fokus auf Methoden/Daten.
[26]	RE for Explainable AI	Fokus auf Explainable AI, Vorarbeit zu [22].
[27]	RE in the Age of AI	Kein Mehrwert gegenüber neueren Studien. (>2020)
[28]	RE for UAV Anomaly Models	Zu spezifischer Fokus auf Anwendungsgebiet.
[29]	AI-centric RE Framework	Kurze Arbeit, die nur auf ethische Richtlinien eingeht.