



# Künstliche Intelligenz zur Unterstützung von agilen Methoden im SAFe Framework

## Masterarbeit

zur Erlangung des akademischen Grades  
„Master of Science (M.Sc.)“

im Studiengang **Wirtschaftsingenieurwesen**  
der Fakultät für Elektrotechnik und Informatik,  
Fakultät für Maschinenbau und der  
Wirtschaftswissenschaftlichen Fakultät der  
Leibniz Universität Hannover

vorgelegt von

Name: **Gandaev**



Vorname: **Abu-Umar**



**Prüfer:** Prof. Dr. Michael H. Breitner

**Abgabe:** Hannover, 16.04.2025

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung</b>	<b>1</b>
<b>2</b>	<b>Theoretische Grundlagen</b>	<b>3</b>
2.1	Agile Methoden im Überblick . . . . .	3
2.2	Scaled Agile Framework . . . . .	6
2.3	Künstliche Intelligenz . . . . .	10
2.3.1	Definition und historische Entwicklung . . . . .	10
2.3.2	Kategorisierung von KI . . . . .	12
2.3.3	Systematische Darstellung der KI-Landschaft . . . . .	13
<b>3</b>	<b>Methodik</b>	<b>17</b>
3.1	Überblick . . . . .	17
3.2	Der Relevanz-Zyklus . . . . .	21
3.3	Artefaktentwicklung im Design-Zyklus . . . . .	22
3.4	Iterative Evaluation und Validierung . . . . .	22
3.5	Abschluss und Perspektiven . . . . .	23
<b>4</b>	<b>Experteninterviews</b>	<b>26</b>
4.1	Datenerhebung . . . . .	26
4.2	Datenanalyse . . . . .	31
4.3	Datenauswertung . . . . .	35
4.4	Diskussion . . . . .	47
<b>5</b>	<b>Artefakt: Der Agile LLM Compass</b>	<b>58</b>
5.1	Motivation, Herausforderungen und Designziele . . . . .	58
5.2	Design- und Entwicklungsprozess (DSR-Iteration) . . . . .	60
5.2.1	Initialkonzept und Prototyp: Der <i>LLM Selector</i> . . . . .	60
5.2.2	Datenmanagement: Quellen, Aufbereitung und Integration . . . . .	66
5.2.3	Backend-Logik: Verarbeitung und Scoring . . . . .	68
5.2.4	Frontend: Benutzeroberfläche und Interaktion . . . . .	68
5.2.5	Der Scoring-Mechanismus . . . . .	69
5.3	Benutzerinteraktion (User Journey) . . . . .	70
5.4	Der Agile LLM Compass als BPMN-Workflow . . . . .	74
5.5	Evaluation und Limitationen . . . . .	75
5.5.1	Evaluation durch Experten . . . . .	75
5.5.2	Stärken des Artefakts . . . . .	77

5.5.3	Limitationen des Artefakts . . . . .	78
5.5.4	Erweiterungsmöglichkeiten (Future Work) . . . . .	78
5.6	Zusammenfassung . . . . .	79
<b>6</b>	<b>Fazit und Ausblick</b>	<b>81</b>
<b>A</b>	<b>Tabellen und Abbildungen</b>	<b>83</b>
<b>B</b>	<b>Code</b>	<b>87</b>
B.1	Ausführung: run.py . . . . .	87
B.2	Backend: Agile_LLM_Compass.py . . . . .	87
B.3	hallucination_leaderboard_metrics_extraction.py . . . . .	123
B.4	livebench_leaderboard_metrics_extraction.py . . . . .	128
B.5	Frontend: Agile_LLM_Compass_Interface.html . . . . .	131
B.6	README-Datei . . . . .	172
B.7	Abhängigkeiten: requirements.txt . . . . .	175
<b>C</b>	<b>Experteninterviews</b>	<b>176</b>
C.1	Interviewleitfaden . . . . .	176
C.2	Transkripte der Experteninterviews . . . . .	179
C.2.1	Experte E1 . . . . .	179
C.2.2	Experte E2 . . . . .	190
C.2.3	Experte E3 . . . . .	205
C.2.4	Experte E4 . . . . .	217
C.2.5	Experte E5 . . . . .	233
C.2.6	Experte E6 . . . . .	250
C.2.7	Experte E7 . . . . .	261

# Kapitel 1

## Einleitung

Ein Blick auf das britische Universal Credit-Programm zeigt exemplarisch, wie selbst ambitionierte, staatlich geförderte Großprojekte mit agilen Methoden scheitern können. Das mit über 12 Milliarden Pfund veranschlagte Vorhaben zur Vereinheitlichung sozialer Leistungen galt als eines der größten agilen Softwareentwicklungsprojekte Europas – und als eines der sichtbarsten Fiaskos. Gründe lagen u.a. in fehlender technologischer Integration, mangelnder Flexibilität in Entscheidungsprozessen und allem voran einem oberflächlichen Verständnis agiler Prinzipien (Ballard 2013).

Dieses Beispiel verdeutlicht: Agilität allein ist kein Erfolgsgarant. Insbesondere in skalierten Umgebungen wie dem *Scaled Agile Framework (SAFe)* bedarf es intelligenter technologischer Unterstützung, um Prozesse adaptiv, vorausschauend und skalierbar zu gestalten. Künstliche Intelligenz (KI) – insbesondere in Form von *Large Language Models (LLMs)* – bietet genau hier neue Potenziale. Laut dem *State of Agile Report 2024* von Digital.ai (2024) setzen immer mehr Unternehmen KI-basierte Tools in ihren agilen Prozessen ein – Tendenz steigend. Dies unterstreicht die zunehmende Rolle von KI als *Katalysator agiler Skalierung*. Der Vormarsch KI-basierter Systeme eröffnet dabei transformative Möglichkeiten, gleichzeitig aber auch neue Herausforderungen: Wie findet man im rasant wachsenden Ökosystem der KI-Modelle das passende Werkzeug für ein konkretes SAFe-Szenario? Die Vielzahl der Modelle, deren Intransparenz und die hohe Varianz in Funktionalitäten erschweren eine fundierte Auswahl erheblich.

Die Komplexität dieser Auswahlentscheidung, gepaart mit der Intransparenz vieler Modelle, stellt ein relevantes praktisches Problem dar. Eine Fehlentscheidung kann nicht nur zu ineffizienten Prozessen führen, sondern auch agile Kernwerte und Prinzipien gefährden. Diese Masterarbeit adressiert die genannten Herausforderung und folgt der zentralen Forschungsfrage:

*Wie kann KI agile Methoden in SAFe unterstützen?*

Kapitel 2 legt zunächst das notwendige theoretische Fundament. Es behandelt die relevanten Konzepte aus dem Bereich agiler Methoden, mit einem besonderen Fokus auf *SAFe* und dessen Essential-Konfiguration. Ergänzend werden die Grundlagen der *Künstlichen Intelligenz (KI)* erläutert, einschließlich Definitionen, historischer Entwicklung, Kategorisierungen und einer systematischen Darstellung der KI-Landschaft. Dieses Kapitel schafft die begriffliche Basis für das Verständnis der nachfolgenden Analysen und der Artefaktentwicklung. Kapitel 3 widmet sich der Methodik dieser Arbeit. Es wird der gewählte Forschungsansatz des *Design Science Research (DSR)* nach Hevner (2007) detailliert erläutert und die Anwendung der drei

zentralen Zyklen – *Rigor*, *Relevanz* und *Design-Zyklus* – beschrieben. Des Weiteren werden die konkreten Schritte der Artefaktentwicklung, der iterativen Evaluation und Validierung sowie die Reflexion über methodische Limitationen und Perspektiven dargelegt. Daran anknüpfend beschreibt Kapitel 4 die Durchführung und Auswertung der qualitativen Experteninterviews. Das Kapitel erläutert den Prozess der Datenerhebung mittels leitfadengestützter Interviews und die anschließende Datenanalyse mittels qualitativer Inhaltsanalyse. Im Anschluss werden die Ergebnisse daraus im Rahmen der Forschungsfrage systematisch anhand von geeigneter Literatur diskutiert. Das entwickelte Artefakt „*Der Agile LLM Compass*“, wird in Kapitel 5 detailliert vorgestellt. Ausgehend von der Motivation, den spezifischen Herausforderungen und den abgeleiteten Designzielen wird folglich der iterative Design- und Entwicklungsprozess (*DSR-Iteration*) nachgezeichnet - vom Initialkonzept („*LLM Selector*“) bis zum finalen Prototypen. Dabei werden Aspekte wie *Datenmanagement*, *Backend-Logik*, *Frontend-Gestaltung*, der *Scoring-Mechanismus* und die *Benutzerinteraktion (User Journey)* erläutert. Abschließend erfolgt eine kritische Evaluation des Artefakts basierend auf Expertenfeedback, eine Diskussion seiner Stärken und Limitationen sowie ein Ausblick auf mögliche Erweiterungen (*Future Work*).

Kapitel 6 schließt die Arbeit mit einem Fazit und Ausblick. Hier werden die zentralen Ergebnisse der Untersuchung zusammengefasst und die Beantwortung der Forschungsfrage reflektiert. Auch werden die Limitationen der Arbeit diskutiert und Perspektiven für zukünftige Forschungsarbeiten aufgezeigt. Ergänzt wird die Arbeit durch einen Anhang, der weiterführende *Tabellen und Abbildungen (Anhang A)*, den *Code (Anhang B)* sowie Leitfaden und Transkripte zu den *Experteninterviews (Anhang C)* enthält. Diese Gliederung gewährleistet einen nachvollziehbaren Aufbau von den theoretischen und methodischen Grundlagen über die empirische Fundierung und Artefaktentwicklung bis hin zur abschließenden Bewertung und Reflexion der Ergebnisse.

# Kapitel 6

## Fazit und Ausblick

Die vorliegende Masterarbeit thematisierte das zunehmend relevante Spannungsfeld zwischen Künstlicher Intelligenz und etablierten agilen Methoden des Scaled Agile Frameworks. Ausgehend von der Tatsache, dass die Auswahl geeigneter KI-Werkzeuge für SAFe-Kontexte aufgrund der Vielfalt und Intransparenz der Modelle schwierig ist, wurde der zentralen Forschungsfrage nachgegangen:

*Wie kann KI agile Methoden in SAFe unterstützen?*

Die wichtigste Schlussfolgerung lautet, dass die derzeit größten Potenziale eines effektiven KI-Einsatzes in der Anwendung von Large Language Models liegen. Ihre erfolgreiche Nutzung im Rahmen agiler Skalierung geht jedoch über eine rein technologische Integration hinaus. Die Synthese aus Literaturanalyse und insbesondere tiefgehenden Einblicken aus den durchgeführten Experteninterviews machen deutlich: Eine *gefestigte agile Natur*, ein *klares Bekenntnis zu menschlicher Kontrolle* und *durchdachte Governance* sind unabdingbare Voraussetzungen, um KI-Werkzeuge wie LLMs wertstiftend und im Einklang agiler Prinzipien zu integrieren. Die Technologie dient als *Enabler*, nicht jedoch als Ersatz für das agile Fundament. Um die spezifische Herausforderung der Modellauswahl anzugehen, wurde als konkreter Beitrag dieser Arbeit der „**Agile LLM Compass**“ entwickelt und einer Expertenevaluation durchzogen. Das webbasierte Artefakt bietet eine methodisch fundierte Antwort auf das beschriebene Problem. Es verkörpert einen strukturierten, kriterienbasierten Prozess, der Anwendern eine nachvollziehbare Orientierung bei der Vielzahl an verschiedenen LLMs und ihrer Versionen und Ausprägungen ermöglicht. Durch die Gewichtung SAFe-relevanter Kriterien und die Nutzung eines sowohl dynamischen als auch statischen Datenansatzes liefert der Compass personalisierte Empfehlungen und fördert die informierte Entscheidungsfindung.

Die Arbeit beantwortet die Forschungsfrage folglich differenziert: KI kann SAFe unterstützen, indem sie Werkzeuge wie LLMs bereitstellt, die spezifischen Aufgaben von Planung bis Code-Qualität optimieren können. Der effektive Hebel liegt aber immernoch in der vorgelagerten, fundierten Auswahl des passenden Werkzeugs. Der „**Agile LLM Compass**“ demonstriert, wie dieser Auswahlprozess in der Praxis aussehen kann. Es ist jedoch wichtig zu verstehen, dass der Compass ein *Entscheidungsunterstützungssystem* ist. Er liefert demnach fundierte Empfehlungen für die Auswahlphase, nimmt aber *weder die finale Entscheidung ab* noch deckt er die darauffolgende Implementierung oder hochspezifische Nischenanforderungen ab. Er ist in seinem Wesen ein Proof of Concept, der die Machbarkeit und Nützlichkeit eines solchen Ansatzes aufzeigt.

Jedoch unterliegt auch diese Arbeit, wie jede wissenschaftliche Untersuchung, konkreten Limitationen. Diese sind der Fokus auf SAFe/LLMs, eine zu hohe Datenabhängigkeit und die für eine Masterarbeit übliche, geringe Stichprobengröße. Die Qualität der Empfehlungen des „**Agile LLM Compass**“ steht und fällt demnach mit der Güte, Aktualität und Verfügbarkeit der externen Benchmark-Daten. Die Abhängigkeit von Web-Scrapern für dynamische Daten birgt nämlich eine Wartungsherausforderung und potenzielle Fehlerquellen. Der implementierte Matching-Algorithmus zur Verknüpfung von Modellnamen über verschiedene Quellen hinweg ist derzeit noch pragmatischer Natur. Dies liegt unter anderem daran, dass im KI-Kontext im Allgemeinen noch keine einheitliche Namensgebung durchgesetzt wird. Das Scoring-Modell stellt zudem eine notwendige Abstraktion dar, die komplexe Wechselwirkungen zwischen Kriterien nur vereinfacht abbildet. Trotz dieser Limitationen liefert die Arbeit eine wertvolle Basis. Sie bietet nicht nur das Konzept und den Prototyp eines nützlichen Werkzeugs, sondern auch vertiefte Einblicke in die Wahrnehmung und die kritischen Erfolgsfaktoren von KI in skalierten agilen Organisationen. Die abgeleiteten Designprinzipien (Ende Kapitel 4.4) bieten zudem über das Artefakt hinaus Orientierung.

Ein Ausblick bleibt schlussendlich vielversprechend und herausfordernd zugleich. Über die bereits diskutierten Weiterentwicklungsmöglichkeiten des „**Agile LLM Compass**“ (siehe Kapitel 5.5.4) hinaus ergeben sich breitere Forschungsfelder: Wie gestalten sich die Langzeitfolgen der KI-Integration auf agile Kulturen und Prozesse? Welche Governance-Modelle bewähren sich in der Praxis? Wie können die hier gewonnenen Erkenntnisse auf andere Kontexte oder agile skalierte Frameworks übertragen werden? Die rasante Entwicklung der KI erfordert eine kontinuierliche wissenschaftliche Begleitung und kritische Reflexion. Diese Masterarbeit hat einen fundierten Beitrag geleistet, um die Brücke zwischen dem Potenzial der KI und den Anforderungen agiler Skalierung zu bauen - ein insgesamt wichtiger Schritt, um die Zukunft der agilen Arbeit verantwortungsvoll mitzugestalten.