

Bachelorarbeit am Institut für Informatik der Freien Universität Berlin

Human-Centered Computing (HCC)

Hybrid Intelligence – Bestimmung der Sweet-Spot-Konfiguration für den Zuschnitt von Bildmedien

Karsten Sonnack

Betreuerin: Frau Prof. Dr. Claudia Müller-Birn

Erstgutachterin: Frau Prof. Dr. Claudia Müller-Birn

Zweitgutachter: Herr Prof. Dr. Lutz Prechelt

Berlin, 11. Juli 2023

Zusammenfassung

Die Entwicklung von Künstlicher Intelligenz (KI) war in den vergangenen Jahren maßgeblich von einem technikzentrierten Designansatz geprägt. Der neuartige, sich davon abhebende Ansatz der „Hybrid Intelligence“ fokussiert eine integrative Herangehensweise, die sowohl die menschliche als auch die algorithmisch-technische Perspektive integriert. Die vorliegende Arbeit untersucht ein praktisches Anwendungsbeispiel der Realisierung von „Hybrid Intelligence“. In einem realen betrieblichen Anwendungsfeld soll im Kontext des grafischen Bildzuschnitts ein Empfehlungssystem entwickelt werden, welches Anwender:innen durch KI-Technologie algorithmisch generierte Bildzuschnitte vorschlägt. Dabei wird eine Form der Interaktion und eine Tiefe der KI-Integration gesucht, die sowohl menschliche als auch algorithmische Fähigkeiten berücksichtigt und optimal kombiniert. Ziel ist die Erschaffung eines interaktiven, intelligenten Systems, welches leistungsfähiger ist als eine KI-Technologie oder der Mensch alleine es wäre. Zur Analyse, Konkretisierung und Evaluation dieses Automatisierungsproblems orientiert sich die Arbeit am siebenstufigen Prozess nach Mackeprang et al. für den Aufbau interaktiver intelligenter Systeme [29]. Am Ende dieses Prozesses steht die Identifizierung der sogenannten Sweet-Spot-Konfiguration. Diese beschreibt die optimale Kombination der Zusammenarbeit von Mensch und KI. Hierzu werden zunächst alle möglichen Konfigurationen definiert, die sich jeweils in Interaktionsform und Tiefe der KI-Integration unterscheiden. Im nächsten Schritt werden diese Konfigurationen in funktionalen Programmcode überführt und ihre Leistungsfähigkeit in Usertests bewertet. Die Evaluation erfolgt auf Basis der benötigten Zeit je Bildzuschnitt, der Qualität der algorithmisch generierten Bildzuschnitte, der von den Anwender:innen subjektiv wahrgenommenen Arbeitsbelastung sowie des Nutzens für Anwender:innen im Umgang mit dem System. Dabei zeichnet sich die Konfiguration mit einem mittleren Automatisierungsgrad als die performanteste aller getesteten Varianten ab. Diese entspricht einem Automatisierungslevel 5 nach Sheridan et al. [35]. Neben einer Zeitersparnis von 62% im Vergleich zur Ausgangskonfiguration, die einem komplett manuellen Bildzuschnitt durch die Anwender:innen entspricht, ermöglicht sie zudem eine um 60% geringere wahrgenommene Arbeitsbelastung. Gleichzeitig stellte sich heraus, dass die Benutzer:innenfreundlichkeit optimiert werden konnte und der Korrekturaufwand bei den algorithmisch generierten Bildzuschnitten für die Nutzer:innen am geringsten ausfiel. Dieses Ergebnis zeigt, dass das gewählte methodische Vorgehen geeignet ist, um im Zuge der Implementierung KI-gestützter Automatisierung das synergetische Potenzial hybrider Intelligenz zu optimieren.

Summary

Lately, the development of artificial intelligence (AI) has been dominated by a technology-centric design approach. The novel approach of "Hybrid Intelligence", which stands out from this, focuses on an integrative method that integrates both the human and the algorithmic-technical perspective. The present work examines a practical application example of the realization of "Hybrid Intelligence". In the context of graphical image cropping, a recommendation system has been developed in a real operational application field, which suggests algorithmically generated image cropping to users using AI technology. The aim is to find a form of interaction and a degree/level of AI integration that takes into account and optimally combines both human and algorithmic skills. The goal is to create an interactive, intelligent system that is more efficient than either AI technology or humans alone would be. To analyze, concretize, and evaluate this automation problem, the work is guided by Mackeprang et al.'s seven-step process for building interactive intelligent systems [29]. At the end of this process, the identification of the so-called sweet spot configuration will take place, which describes the optimal combination of human and AI collaboration. For this purpose, first, all possible configurations are defined, each differing in the form of interaction and depth of AI integration. In the next step, these configurations are transferred into functional program code and their performance is evaluated in user tests. The evaluation is based on the time required per image section, the quality of the algorithmically generated image sections, the workload subjectively perceived by the users, and the benefits for users in dealing with the system. The configuration with a medium level of automation was identified as the best performing of all the tested variants. This corresponds to an automation level of 5 according to Sheridan et al. [35]. In addition to a time saving of 62% compared to the initial configuration, which corresponds to a completely manual image cropping by the users, it also enables a 60% lower perceived workload. At the same time, an optimized user-friendliness as well as the lowest correction effort for the algorithmically generated image cropping were shown. The results prove that the selected methodological approach is suitable for optimizing the synergetic potential of hybrid intelligence while implementing AI-supported automation.

Inhaltsverzeichnis

1. Einleitung	1
1.1. Motivation	1
1.2. Problemstellung	2
1.3. Forschungsziele und Forschungsfrage	3
2. Stand der Forschung	5
2.1. Qualifikationsentwicklungen	5
2.1.1. Deskillung	5
2.1.2. Upskilling	6
2.1.3. Gefahr der Dequalifizierung im Zeitalter von künstlicher Intelligenz	6
2.2. Hybrid Intelligenz	8
3. Methodik	10
3.1. Human-Computer Configuration Design	10
4. Ausgangszustand	14
4.1. Systemarchitektur	14
4.2. Webanwendung für Bildzuschnitte	16
5. Analyse	18
5.1. Identifikation des „point of friction“	18
5.2. Den „point of friction“ im Arbeitsablauf lokalisieren	19
5.3. Leistungsmetriken zur Bewertung von Konfigurationen	20
5.4. Grenzkonfigurationen festlegen	24
5.5. Zwischenkonfigurationen definieren	25
5.6. Ausschluss nicht geeigneter Zwischenkonfigurationen	28
6. Gestaltung	29
6.1. Level-1-Konfiguration (LoA1)	29
6.2. Level-2-Konfiguration (LoA2)	29
6.3. Level-3-Konfiguration (LoA3)	30
6.4. Level-5-Konfiguration (LoA5)	31
7. Implementierung	32
8. Evaluation	33
8.1. Aufbau und Durchführung der User-Tests	33
8.2. Ergebnisse	34
8.2.1. Zuschnittszeit	34
8.2.2. Wahrgenommene Arbeitsbelastung	34
8.2.3. Eingesetzter Aufwand zur Finalisierung von Bildzuschnitten	35
8.2.4. ResQue Fragebogen	38
8.3. Diskussion der Ergebnisse	39

8.4. Ableitung der Sweet-Spot-Konfiguration	42
8.5. Sweet-Spot-Konfiguration und der Aspekt der Dequalifizierung	42
9. Fazit	44
A. Anhang	45
A.1. Fragen des NASA TLX Fragebogens	45
A.2. Fragen des angepassten ResQue Fragebogens	45
A.3. Ergebnisse der Usertests	46
A.4. GIT Repository	46
Literaturverzeichnis	47

Abbildungsverzeichnis

1.	Interaktionsformen intelligenter Systeme [34]	9
2.	Prozess nach Mackeprang et al. [29]	12
3.	Systemarchitektur	14
4.	Kennwerte eines Bildzuschnitts	15
5.	Webanwendung für den Bildzuschnitt - Bildauswahl	16
6.	Webanwendung für den Bildzuschnitt - Formatauswahl	17
7.	Kontextabhängige Motivwahl	18
8.	Arbeitsablauf als Flussdiagramm	20
9.	Verschiebung des Zuschnittfensters	22
10.	Formel zur Berechnung der Verschiebung des Zuschnittfensters	22
11.	Vergrößerung bzw. Verkleinerung des Zuschnittfensters	23
12.	Formel zur Berechnung der Vergrößerung bzw. Verkleinerung des Zuschnittfensters	23
13.	Level-2-Konfiguration – AoI-Auswahl	30
14.	Level-2-Konfiguration – Bildauswahl	30
15.	Level-3-Konfiguration	31
16.	Level-5-Konfiguration	31

Tabellenverzeichnis

1.	10 Levels of Automation nach Sheridan et al. [35]	11
2.	Aufgabenzuweisung der LoA1-Konfiguration	24
3.	Aufgabenzuweisung der LoA5-Konfiguration	25
4.	Aufgabenzuweisung der LoA2-Konfiguration	26
5.	Aufgabenzuweisung der LoA3-Konfiguration	27
6.	Aufgabenzuweisung der LoA4-Konfiguration	27
7.	Evaluationsergebnisse der Zeiterfassung	34
8.	Evaluationsergebnisse des Nasa Raw TLX Fragebogens	35
9.	Eingesetzter Aufwand zur Finalisierung von Bildzuschnitten - Vergleich der Prototypen hinsichtlich der Verschiebung des Bildmittelpunktes	36
10.	Eingesetzter Aufwand zur Finalisierung von Bildzuschnitten - Vergleich der Prototypen hinsichtlich der Zuschnittsvergrößerung bzw. -verkleinerung	37
11.	Ergebnisse des ResQue-Fragebogens	39

1. Einleitung

Seit ChatGPT im November 2022 das Bewusstsein der breiten Öffentlichkeit erlangte, dominiert das Thema Künstliche Intelligenz (KI) und intelligente Systeme die öffentliche Debatte. Gleichwohl sind interaktive, intelligente Systeme – also Systeme, die Formen von künstlicher Intelligenz einbinden – längst fester Bestandteil unseres Alltags. Man begegnet ihnen täglich in Form von Empfehlungsdiensten auf Streaming-Plattformen, in der Content-Aufbereitung von Social-Media-Plattformen oder auch in der Interaktion mit virtuellen Assistenten und Chatbots. Die zunehmende Verbreitung dieser Systeme hat weitreichende Konsequenzen für die Interaktion zwischen Mensch und Computer und verändert das Rollenverständnis beider Seiten grundlegend.

1.1. Motivation

Die Unterstützung und Optimierung menschlicher Entscheidungsprozesse durch die Zusammenarbeit zwischen Mensch und intelligenten Systemen ist bereits ein gängiges Paradigma [45, 11]. Dies spiegelt sich auch im wissenschaftlichen Diskurs wider. Dabei ist es hauptsächlich die inzwischen erreichte Leistungsfähigkeit intelligenter Systeme, die es ermöglicht, höhere Effizienz und Genauigkeit in Tätigkeiten zu realisieren, die zuvor manuell durch den Mensch ausgeführt wurden. Im Bereich der Medizin wird seit einigen Jahren der Einsatz intelligenter Systeme für eine Verbesserung der Diagnosequalität intensiv genutzt und untersucht [32, 12, 6, 25]. Vor allem im Bereich der bildgebenden medizinischen Verfahren zeigen Studien die Wirksamkeit dieser Technologien [19, 42]. Die Entwicklungen sind dabei primär durch einen technikzentrierten Designansatz geprägt [37, 36, 46, 44]. Dabei steht vorrangig die technische, am Algorithmus orientierte Perspektive auf eine Problemstellung im Fokus, die menschliche Komponente wird nur wenig berücksichtigt [26, 22]. Nun agieren intelligente Systeme aber in den seltensten Fällen vollständig isoliert und autonom, ohne Interaktion mit dem Mensch. Vielmehr handeln beide zusammen als sozio-technisches System. Deshalb sollten bei der Entwicklung Bedürfnisse und Kompetenzen beider Interaktionspartner berücksichtigt und idealerweise miteinander kombiniert werden. Neben einer potenziellen Optimierung der Ergebnisse lassen sich dadurch auch Spannungsfelder vermeiden, welche sich auf Interaktion und Nutzungsverhalten ungünstig auswirken könnten.

Ein bekanntes Beispiel für dieses Spannungsfeld ist das autonome Fahren. Durch den Einsatz intelligenter Systeme in Fahrzeugen verschiebt sich das Rollenverständnis von Fahrer:innen. Autonome Fahrzeuge ermöglichen Fahrer:innen im Normalbetrieb den Übergang von einer aktiven, Fahrzeug führenden Rolle hin zu einer beobachtenden Rolle [18]. Indem das Fahrzeug eigenständig Entscheidungen treffen und Handlungen einleiten kann, werden Fahrer:innen kognitiv entlastet und ihr Fahrkomfort gesteigert. In Situationen, in denen das autonome Fahrzeug keine eigenständige Entscheidung treffen kann, fordert es Fahrer:innen auf, die Kontrolle über das Fahrzeug und die Situation zu übernehmen. Genau dieser Übergang stellt im Kontext des autonomen Fahrens ein

relevantes Sicherheitsrisiko dar [16]. Für das Führen eines Fahrzeugs sind eine Reihe wahrnehmungsbezogener und motorischer Fähigkeiten notwendig [40]. Sie werden durch praktische Erfahrung erlernt und kontinuierlich verinnerlicht. Werden diese Fähigkeiten allerdings nicht fortlaufend abgerufen und angewandt, besteht das Risiko, dass sie sich verschlechtern oder vollständig verloren gehen. Kommt es nun zu einer Situation, die komplex und damit potenziell gefährlich ist, fordert das intelligente System Fahrer:innen programmgetreu dazu auf, die Kontrolle über das Fahrzeug zu übernehmen. Die Fahrer:innen sind dann allerdings mit einer Situation konfrontiert, in der sie selbst aufgrund erodierter Fähigkeiten nicht mehr adäquat handeln können [16, 48].

Das Beispiel beschreibt das Risiko eines schrittweisen Verlusts von Fähigkeiten, weshalb man in diesem Kontext auch von Deskillung oder Dequalifizierung spricht. Diese Begriffe bezeichnen den graduellen Verlust beruflicher oder auch alltäglicher Fähigkeiten aufgrund technologischer oder arbeitspraktischer Veränderungen, was in Folge von Automatisierung als häufiges Phänomen auftritt [34]. Gerade im Kontext künstlicher Intelligenz ist anzunehmen, dass die Relevanz dieses Phänomens noch weiter zunehmen wird. Das gewählte Beispiel des autonomen Fahrens steht an dieser Stelle nur exemplarisch für eine Vielzahl von Anwendungsfällen, in denen intelligente Systeme einerseits Komfort-, Effizienz- oder auch Qualitätsvorteile bieten, andererseits aber das Risiko in sich bergen, dass wertvolle Fähigkeiten verloren gehen, die sonst als Teil des alltäglichen Lebens erhalten blieben [38, 37]. Besonders kritisch sind dabei alle Prozesse und Tätigkeiten, bei denen keine vollständige Zuverlässigkeit bei der Entscheidungsfindung intelligenter Systeme gegeben ist und es notwendig wird, dass der Mensch manuell in Abläufe eingreift. Umso wichtiger ist es, bei der Entwicklung intelligenter Systeme die Interaktion zwischen Mensch und Computer zu berücksichtigen.

Das unter dem Begriff „Hybrid Intelligence“ bekannte Interaktionsverständnis beschreibt genau einen solchen alternativen Ansatz, welcher über die reine Substitution menschlicher Fähigkeiten hinausgeht [5, 28, 15]. Grundgedanke dabei ist, eine partnerschaftliche Kollaboration zu schaffen, in der die komplementären Fähigkeiten von Mensch und intelligenten Systemen kombiniert eingesetzt werden [15]. Ziel ist nicht, den Menschen schrittweise zu ersetzen, sondern vielmehr einzigartige menschliche Fähigkeiten wie Intuition, Flexibilität und Kreativität für die Problemlösung zu nutzen [15].

Die vorliegende Arbeit versteht sich als Beispiel einer praktischen Realisierung von Hybrid Intelligence. Sie befasst sich mit der Implementierung KI-basierter Algorithmen zur Aufbereitung grafischen Bildmaterials in den bereits existierenden Arbeitsablauf eines Betreibers von Healthcare-Anwendungen.

1.2. Problemstellung

Gegenstand dieser vorliegenden Arbeit ist eine reale Problemstellung der Significo GmbH im Kontext des automatisierten Bildzuschnitts. Die Significo GmbH ist ein Software-Unternehmen, welches digitale Healthcare Produkte konzipiert, entwickelt und betreibt.

Neben nativer mobiler Apps werden auch browserbasierte Webanwendungen entwickelt. Dabei setzt das Unternehmen stark auf die Einbindung von Bildmedien. Die Verwaltung dieser Medien erfolgt durch ein sogenanntes Digital Asset Management (DAM) System, welches die Bilddaten über fest definierte Schnittstellen direkt an die jeweiligen Apps und Webanwendungen ausliefert. Dementsprechend muss eine Vielzahl an Formaten und potenziellen Endgeräten bedient werden. Dies stellt Significo vor zunehmende Herausforderungen bei der Bereitstellung von Bildmedien. Egal in welcher App, für welche Webanwendung, auf welchem Endgerät oder in welchem inhaltlichen Kontext Bilder eingebunden werden – sie müssen immer passend und korrekt dargestellt werden. Um in all diesen Fällen eine maximale Kompatibilität gewährleisten zu können, werden Bildinhalte bei der Significo GmbH grundsätzlich in sieben vordefinierte Formate zugeschnitten, die dann je nach Endgerät und Anwendung passend abgerufen werden können. Diese Zuschnitte wurden bisher routinemäßig manuell von Contentmanager:innen der Significo GmbH mithilfe einer hausinternen Webanwendung erstellt. Gleichzeitig sind bereits intelligente Lösungen auf dem Markt, die in der Lage wären, bei genau solch einer Bildaufbereitung zu unterstützen. Diese werden im Arbeitsablauf aktuell jedoch noch nicht genutzt. Vor diesem Hintergrund wurde im Rahmen der vorliegenden Arbeit ein Automatisierungsprojekt bei der Significo GmbH aufgesetzt, welches darauf abzielt, den Prozess des Bildzuschnitts zu beschleunigen, um Personalressourcen zu schonen und gleichzeitig sicherzustellen, dass die Anforderungen an die Zuschnittsqualität erfüllt werden.

Eine reine Automatisierung dieser Abläufe ohne Interaktion mit den Anwender:innen konnte bereits zum Projektstart ausgeschlossen werden, birgt sie doch große Risiken. So könnten beispielsweise unpassende Bildausschnitte ohne menschliche Kontrolle direkt an die Apps und Webanwendungen der Significo GmbH ausgeliefert werden. Viel sinnvoller erschien ein Lösungsansatz, der auf die kollaborative Zusammenarbeit zwischen Mensch und Computer setzt. Ziel war es, den Grad an Automatisierung zu finden, der die Kompetenzen und Fähigkeiten beider Interaktionspartner in optimaler Weise miteinander kombiniert – also eine „Hybrid Intelligence“ zu schaffen.

1.3. Forschungsziele und Forschungsfrage

Diese Arbeit versteht sich als ein Beitrag zur Gestaltung von interaktiven, intelligenten Systemen, in denen Mensch und Computer kollaborativ zusammenarbeiten. Das konkrete Ziel dieser Arbeit ist die systematische Optimierung einer Webanwendung für den Zuschnitt von Bildmedien. Bezogen auf die oben beschriebene Problemstellung sollen mögliche Potenziale für eine Automatisierung durch künstliche Intelligenz analysiert, in konkrete Aufgabenzuweisungen für Mensch und Computer überführt und verschiedene Automatisierungsgrade vergleichend untersucht und bewertet werden. Dazu werden entsprechende Prototypen der Webanwendung entwickelt, die sich in Automatisierungsgrad und Interaktionsform unterscheiden. Durch die Erhebung empirischer Kennzahlen soll die Leistungsfähigkeit dieser Prototypen hinsichtlich ihrer Optimierungsziele verglichen

werden. Neben objektiven Metriken (der benötigten Zeit pro Zuschnitt) werden auch subjektive Bewertungsmaße (Arbeitsbelastung, Zufriedenheit der Nutzer:innen, Zuschnittsqualität) erhoben. Im Rahmen der Evaluation werden diese Ergebnisse abschließend bewertet und in eine Handlungsempfehlung überführt. Dabei soll die Frage beantwortet werden, wie eine Aufgabenzuweisung für den Zuschnitt von Bildmedien zwischen Mensch und Computer definiert sein sollte, um einen Grad der Automatisierung zu realisieren, der die Potenziale einer kollaborierenden Arbeitsweise zwischen Mensch und intelligentem System optimal ausschöpft. Der Frage, der innerhalb dieser Arbeit nachgegangen werden soll, lautet:

„Wie muss die Aufgabenzuweisung für den Zuschnitt von Bildmedien zwischen Mensch und intelligentem System definiert sein, um einen Grad der Automatisierung zu realisieren, der die Potenziale einer kollaborierenden Arbeitsweise optimal ausschöpft?“

2. Stand der Forschung

Wie bereits in Kapitel 1.1 erwähnt, ist Deskilling eines der kontrovers diskutierten Spannungsfelder bei der Verwendung intelligenter Systeme. Insbesondere dann, wenn Entscheidungen vom System nicht immer selbstständig getroffen werden können bzw. dürfen oder aber die Ergebnisse nicht zuverlässig genug sind und einer weiteren qualifizierten Beurteilung durch Nutzer:innen bedürfen. Das folgende Kapitel gibt Einblicke in die Ursprünge des Begriffs der Dequalifizierung und zeigt Parallelen zu aktuell stattfindenden Transformationen in der Arbeitswelt durch den Einsatz künstlicher Intelligenz auf. Darüber hinaus wird der Ansatz von Hybrid Intelligence als ein möglicher Lösungsweg vorgestellt, welcher die Risiken von Deskilling nicht ignoriert und die Folgen dieses Effektes potenziell eingrenzt.

2.1. Qualifikationsentwicklungen

Betrachtet man die Entwicklungen der vergangenen dreihundert Jahre innerhalb der westlichen kapitalistischen Gesellschaft, so sind diese geprägt von einer gestiegenen Alphabetisierung, einem gestiegenen allgemeinen Bildungsniveau und einem stark gewachsenen Anteil an Personen, die in das Erwerbsleben integriert sind [41]. Sowohl Bildung als auch das zur Verfügung stehende Qualifikationsniveau, welches in Arbeitsprozessen genutzt werden kann, sind in dieser Zeitspanne unbestreitbar gestiegen [41]. Diese Tatsache ist aber nicht gleichbedeutend mit der Annahme, dass dieses Qualifikationspotenzial auch wirklich ausgeschöpft wird. Vielmehr gibt es hinsichtlich der Konsequenzen eines gesteigerten Automatisierungsgrades zwei sehr gegensätzliche Sichtweisen bezüglich der Qualifikationsentwicklung: die des Deskilling und die des Upskilling.

2.1.1. Deskilling

Der Begriff Deskilling entstand aus Beobachtungen während der Zeit der Industrialisierung, als mit der kapitalistischen Produktionsweise durch Arbeitsteilung der konzeptionelle und fachliche Teil einer Aufgabe von der eigentlichen Ausführung der Aufgabe getrennt wurde [39]. Auf diese Weise konnten hoch qualifizierte Arbeitskräfte durch die schnell wachsende Gruppe günstigerer und geringer qualifizierter Fabrikarbeiter:innen ersetzt werden [9]. Braverman [9] beschreibt Deskilling als einen Prozess der Verringerung des Qualifikationsniveaus von Arbeitnehmenden. Beispielhaft wird von Vertreter:innen der Deskilling-Theorie das Weberhandwerk und seine Veränderung während der industriellen Revolution aufgeführt [41, 34, 10]. Der mechanische Webstuhl und die damaligen organisatorischen Fortschritte ebneten den Weg für eine Substitution qualifizierter Handwerker:innen durch ungelernte Fabrikarbeiter:innen [10]. Ermöglicht wurde dies durch die Trennung von konzeptionellen Tätigkeiten und der eigentlichen Ausführung im Herstellungsprozess. Ein zentrales Produktionssystem sicherte und integrierte das Wissen von Handwerker:innen und abstrahierte dieses in Form von allgemeingültigen

Vorgaben und standardisierten Arbeitsabläufen. Auf diese Weise konnten auch minderqualifizierte Fabrikarbeiter:innen Stoffe produzieren, indem sie diesen Vorgaben folgten. Ohne das industrielle Produktionssystem selbst genügte die Qualifikation der Fabrikarbeiter:innen jedoch nicht, um eigenständig Stoffe herzustellen – ein frühes Beispiel für Deskillung [34]. Auch Karl Marx gilt als einer der frühen Vertreter:innen dieser These. Er vertrat die Sichtweise, dass Industrialisierung und Mechanisierung eine starke Tendenz haben, zur Dequalifizierung beizutragen [24]. Dabei ist für ihn weniger die Technologie selbst die Ursache als vielmehr der systemische Rahmen und die Art und Weise, wie Technologie eingesetzt und genutzt wird [24]. Im Sinne von Marx ist es also das kapitalistische Produktionssystem, welches darauf abzielt oder als Nebeneffekt billigend in Kauf nimmt, die Position der Arbeitnehmer:innen zu schwächen und dadurch Effekte von Dequalifizierung begünstigt [24]. Dabei lassen sich nach Heisig [41] die drei nachfolgenden Hauptmechanismen ableiten:

1. Beseitigung des Bedarfs an qualifizierten Arbeitnehmer:innen durch Vereinfachung und Automatisierung einzelner Aufgaben.
2. Erleichterung der Arbeit.
3. Herabstufung von Arbeitsplätzen oder Tätigkeiten von einer qualifizierten teureren Arbeitskraft zu einer weniger qualifizierten günstigeren Arbeitskraft, um die Lohnkosten als zentralen Kostentreiber zu minimieren.

2.1.2. Upskilling

Die Theorie des Upskilling vertritt eine positivere Perspektive auf die Qualifikationsentwicklung als Folge von Automatisierung und Technisierung. Der Begriff des Upskilling beschreibt dabei den Prozess, bei dem Arbeitnehmer:innen neue Fähigkeiten und Kenntnisse erwerben, um den Anforderungen des sich ständig verändernden Arbeitsmarktes gerecht zu werden. Wie am Beispiel des Weberhandwerks bereits erläutert, führten in der Vergangenheit technische Innovationen zur Auflösung alter Berufe. Die Vertreter:innen der Upskilling-Theorie argumentieren, dass eine direkte Folge davon die Schaffung neuer Arbeitsplätze sei, die wiederum ganz neue Fähigkeiten erfordern. Im Beispiel der industrialisierten Stoffherstellung waren für den Bau, den Betrieb und die Wartung eines mechanischen Webstuhls neue und höhere Qualifikationen notwendig, aus denen sich wiederum bis dato unbekannte Berufe und Verantwortungen ableiteten [41, 7].

2.1.3. Gefahr der Dequalifizierung im Zeitalter von künstlicher Intelligenz

Sowohl für Upskilling als auch Deskillung gibt es wissenschaftliche Belege, sodass sich bis heute beide Sichtweisen auf die Veränderungen im Qualifikationsanspruch legitimieren und begründen lassen [41]. Dementsprechend sollte bei der Gestaltung neuer Systeme in jedem Fall das Risiko des Deskillung berücksichtigt und bewertet werden. Das ist

insbesondere dann wichtig, wenn intelligente Systeme in bestehende Arbeitsabläufe eingebunden werden sollen. So sehen Rafner et al. [34] Analogien zwischen dem Wandel der menschlichen Arbeitskraft in der Zeit der industriellen Revolution und dem sich aktuell verschiebenden Rollenverständnis zwischen Mensch und Computer durch die Verbreitung intelligenter Systeme. Sie identifizieren in deren zunehmender Nutzung die reale Gefahr, dass sich bestehende Berufsgruppen auflösen oder einzelne Tätigkeitsstrukturen innerhalb dieser Berufsgruppen von dequalifizierenden Veränderungen betroffen sein werden. Die Vertreter:innen der Upskilling-Theorie hingegen fokussieren in diesem Zusammenhang die Potenziale von Requalifizierung, da für die Entwicklung und den Betrieb solcher Systeme neue Kompetenzen notwendig sind. Aus ihrer Sicht kommt es nicht zu einer Dequalifizierung, sondern zu einer Verschiebung der Fähigkeiten.

Unabhängig davon, welche Sichtweise man nun in der Deskillung-Upskilling-Debatte einnehmen möchte, bergen diese anstehenden Verschiebungen im Arbeitsmarkt ein großes Konfliktpotenzial. So kamen Brynjolfsson und McAfee bereits 2011 zu der Schlussfolgerung, dass Wissensarbeit sich viel leichter zerlegen und automatisieren lässt als viele manuelle Tätigkeiten [17]. In großen Teilen unserer Gesellschaft ist Wissensarbeit bereits der zentrale Teil unserer Lebensrealität und bietet damit ein großes Potenzial für Veränderungen. Die weltweite Forschungsgemeinde untersucht daher seit einiger Zeit die Effekte von Deskillung durch die Nutzung intelligenter Systeme [38, 39, 40]. So identifizieren Sinagra et al. [38] im Kontext der KI-gestützten Darmkrebsdiagnose, dass Deskillung und ein übermäßiges Vertrauen in die Fähigkeiten des intelligenten Systems zu den Hauptgefahren im Umgang mit den untersuchten Systemen zählen. Sie sind der Meinung, dass junge Mediziner:innen darin geschult werden sollten, diese Systeme besser zu verstehen, um diese anschließend sinnvoll und kritisch einsetzen zu können. Sutton et al. [39] sprechen in diesem Zusammenhang auch von einem Technologie-Dominanz-Effekt, der zu einer schlechteren Entscheidungsfindung führt. Dabei beziehen sie sich unter anderem auf eine Studie, in der Ärzt:innen intelligente Systeme bei der Diagnose einsetzen sollten. Die Studie ergab, dass Testpersonen ihren eigenen Diagnosen weniger Vertrauen schenkten als den maschinell erzeugten. Zu beobachten war dies insbesondere in konkreten Fällen, in denen das intelligente System eine falsche Diagnose generierte, während die Ärzt:innen eigentlich die richtige Diagnose stellten [20, 21]. Es kam in den beschriebenen Fällen also zu einem Vertrauensverlust in die eigenen Fähigkeiten. Weiterhin deutet diese Studie darauf hin, dass Anwender:innen, die wenig praktische Berufserfahrung mitbrachten, im Umgang mit den untersuchten intelligenten Systemen auch schlecht dazulernten. Zusammengefasst ist aus diesen Ergebnissen abzuleiten, dass erfahrene Anwender:innen durch den Einsatz intelligenter Systeme zum einen bestehende Fähigkeiten verlieren, da sie intelligenten Systemen übermäßiges Vertrauen schenken, und zum anderen Wissen und Fähigkeiten weniger gut an unerfahrene Anwender:innen weitergegeben werden. Beide Effekte begünstigen eine Dequalifizierung innerhalb der Berufsgruppe.

Die Dequalifizierung durch den Einsatz intelligenter Systeme und ein übermäßiges Vertrauen in deren Ergebnisse erweist sich also als ein reales Problem. Rafner et al. [34] benennen drei Ansätze, um dem Effekt des Deskillung entgegenzuwirken. Sie empfehlen,

durch entsprechende Ausbildung den Erwerb grundlegender Fähigkeiten zu fördern und diese weiterhin als festen Bestandteil der Lehre zu erhalten, unabhängig vom aktuellen technischen Fortschritt. Zudem sollten Fachleute weiterhin eine aktive Rolle übernehmen und Empfehlungen der KI im Nachgang prüfen. Zuletzt betonen sie die Relevanz der Beziehungsausrichtung. Diese sollte auf Zusammenarbeit ausgerichtet sein und nicht auf Wettbewerb oder Substitution. So begegnen sich Mensch und intelligentes System auf Augenhöhe und arbeiten partnerschaftlich zusammen.

Eine vielversprechende Herangehensweise zur Auflösung des beschriebenen Problems liefert der Ansatz, der als „Hybrid Intelligence“ bezeichnet wird. Durch die Kombination von menschlichen Fähigkeiten und algorithmenbasierten Potenzialen intelligenter Systeme soll eine neue Form der Interaktion geschaffen werden, die nicht auf Wettbewerb abzielt. Nachfolgend wird ein Einblick in die Herkunft dieses Ansatzes sowie der dahinterstehenden Konzepte gegeben.

2.2. Hybrid Intelligence

Die Überlegungen, die zur Formulierung der „Hybrid Intelligence“ führten, resultieren aus Beobachtungen und Erfahrungen bei der Nutzung intelligenter Systeme, wie sie beispielhaft in den vorausgehenden Abschnitten beschrieben wurden. Eine formale Einordnung von Systemen mit hybrider Intelligenz wird von Rafner et al. [34] in einer Art Landkarte der Interaktionsformen intelligenter Systeme vorgenommen (siehe Abbildung 1).

Ganz grundlegend gibt es zwei Interaktionsformen zwischen Mensch und intelligenten Systemen. Zum einen gibt es Systeme, die Entscheidungen treffen, ohne dabei Anwender:innen einzubeziehen. Dabei spricht man von sogenannten Human-Out-Of-The-Loop Systemen (HOOTL). Zum anderen gibt es Systeme, die eigenständig Aufgaben und Entscheidungen treffen können, dabei aber immer Anwender:innen miteinbeziehen und Möglichkeiten der Intervention bieten. Diese werden als Human-In-The-Loop Systeme (HITL) kategorisiert. Eine Untergruppe dieser HITL-Systeme sind die sogenannten Human-On-The-Loop-Systeme (HOTL). Diese führen alle notwendigen Aufgaben und Entscheidungen eigenständig durch, sodass Anwender:innen lediglich das finale Ergebnis bestätigen müssen. Systeme mit einer hybriden Intelligenz ordnen sich als Untergruppe von HITL-Systemen ein und weisen Überschneidungen mit HOTL-Systemen auf [15, 5]. Abbildung 1 visualisiert diese Zusammenhänge.

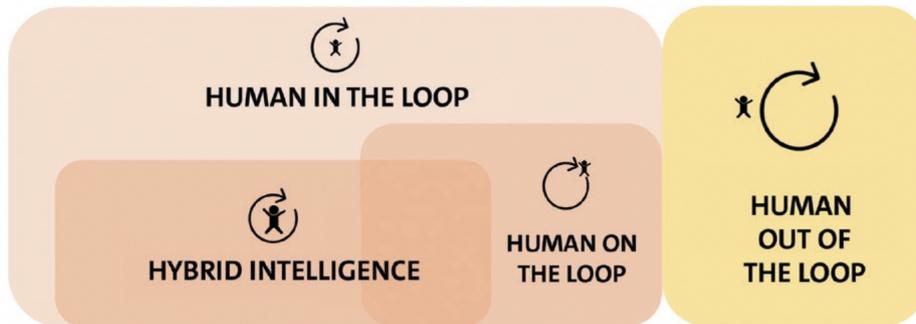


Abbildung 1: Interaktionsformen intelligenter Systeme [34]

Dellermann [15] definiert den Begriff Hybrid Intelligence als die Fähigkeit, komplexe Problemstellungen durch die Kombination von menschlicher und künstlicher Intelligenz zu lösen und dabei bessere Ergebnisse zu erzielen als die, die jede der Parteien separat hätte erreichen können. Hinzu kommt die Option sowohl für die menschliche als auch für die technische Systemkomponente, sich durch gegenseitiges Lernen kontinuierlich zu verbessern [15]. Dabei bezieht er sich auf die Komplementarität von Mensch und KI. So verfügt der Mensch über Sinneswahrnehmungen und emotionale Intelligenz, ist flexibel und kreativ [8, 15], während die KI sich durch Mustererkennung und die Berechnung von Wahrscheinlichkeiten auszeichnet [15]. Der Grundgedanke von Hybrid Intelligence ist die Kombination dieser sich ergänzenden Fähigkeiten, um ein sozio-technisches System zu schaffen, welches die derzeitigen Grenzen intelligenter Systeme überwinden kann [15]. Dabei konzentriert sich dieser Ansatz weder nur auf die menschliche Intelligenz noch auf die reine Automatisierung durch intelligente Systeme. Vielmehr sollen komplexere Probleme gelöst werden, indem eine bewusste Zuteilung der Teilprobleme zum menschlichen und algorithmischen Agenten stattfindet, je nach Fähigkeit und Kompetenz [15].

3. Methodik

Um das synergetische Potenzial hybrider Intelligenz optimal zu nutzen, werden beim Aufbau von interaktiven intelligenten Systemen Methoden benötigt, die im Entwicklungsprozess dabei unterstützen, belastbare Entscheidungen bezüglich des Automatisierungsgrads zu treffen. Designer:innen und Entwickler:innen sehen sich immer häufiger mit der Herausforderung konfrontiert, wie Optionen für die Gestaltung interaktiver, intelligenter Systeme in Hinblick auf den Automatisierungsgrad systematisch definiert und abgegrenzt werden können. Die wissenschaftliche Literatur dazu ist sehr begrenzt und konkrete Methoden sind wenig entwickelt. Einer der wenigen methodischen Ansätze, die sich in diesem Kontext als potenziell hilfreich erweisen, leitet sich aus der Arbeit von Mackeprang et al. [29] zum Thema „Human-Computer Configuration Design“ ab.

3.1. Human-Computer Configuration Design

Eine der historischen Grundlagen für den von Mackeprang et al. erarbeiteten Prozess liefern Sheridan et al. [35] mit der Definition der „ten levels of automation“ (LoA) – einer Skala zur Kategorisierung verschiedener Automatisierungsstufen, die ursprünglich im Kontext der Robotik formuliert wurde. Systeme mit einem „Level of Automation 1“ (LoA1) umfassen keine Form der Automatisierung; alle Tätigkeiten werden manuell durch den Menschen durchgeführt. Systeme mit einem „Level of Automation 10“ (LoA10) hingegen arbeiten voll autonom; alle Tätigkeiten werden durch den Computer ausgeführt und es gibt keinerlei Interaktion mit dem Menschen. Dazwischen spannen sich acht weitere Automatisierungsstufen mit jeweils ganz eigenen charakteristischen Eigenschaften (siehe Tabelle 1). Als Erweiterung dieses Ansatzes gilt das von Parasuraman et al. [31] entwickelte Rahmenwerk zur Untersuchung von Automatisierungsdesigns. Es ergänzt vier funktionale Aufgabentypen, die Ziel von Automatisierungsprozessen sein können. Diese Aufgabentypen sind: Informationsbeschaffung, Informationsanalyse, Entscheidungs- und Handlungsauswahl sowie Handlungsumsetzung. In einem iterativen Prozess sollen alle Aufgabentypen hinsichtlich eines optimalen LoA untersucht werden.

LoA	Definition
10	Der Computer entscheidet alles, handelt autonom und ignoriert den Menschen.
9	Der Computer informiert den Menschen nur, wenn er, der Computer, sich dazu entschließt.
8	Der Computer informiert den Menschen nur, wenn er gefragt wird.
7	Der Computer führt automatisch aus und informiert dann notwendigerweise den Menschen
6	Der Computer gibt dem Menschen begrenzte Zeit, um vor der automatischen Ausführung ein Veto einzulegen
5	Der Computer führt den Vorschlag aus, wenn der Mensch zustimmt
4	Der Computer generiert einen Vorschlag und eine Alternative
3	Der Computer schränkt die Auswahl ein auf einige wenige Alternativen
2	Der Computer bietet eine ganze Reihe von Entscheidungs- und Handlungsalternativen an
1	Der Computer bietet keine Unterstützung, der Mensch muss alle Entscheidungen und Handlungen selbst treffen und ausführen

Tabelle 1: 10 Levels of Automation nach Sheridan et al. [35]

Der Prozess von Mackeprang et al. baut nun weiter auf diesen Überlegungen auf. Durch ein strukturiertes Vorgehen werden alle für die Optimierung von Automatisierungslösungen notwendigen Entscheidungen abgearbeitet. Dazu wurde ein siebenstufiger Ablauf definiert, der eine systematische Analyse des Problemraums ermöglicht. Abbildung 2 zeigt diesen Ablauf.

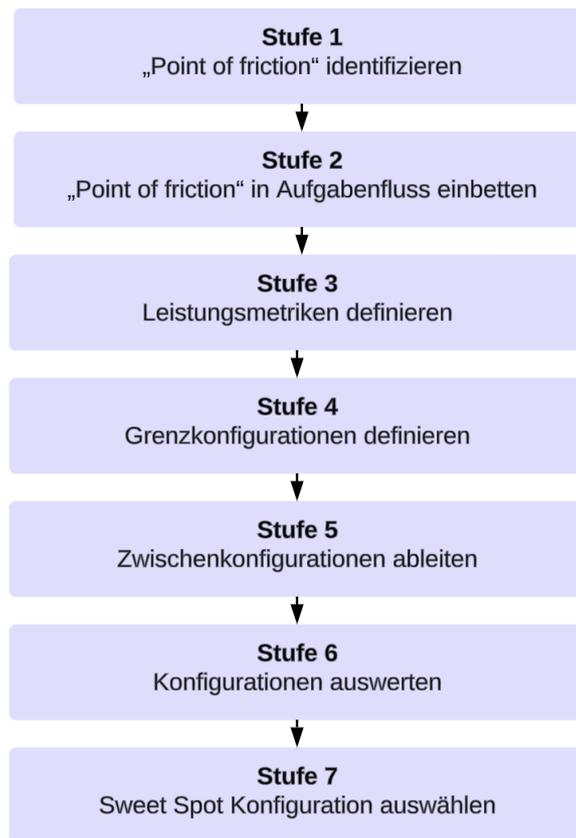


Abbildung 2: Prozess nach Mackeprang et al. [29]

Mackeprang et al. führen in ihrem Stadienmodell zur Konfiguration interaktiver intelligenter Systeme den Begriff des „point of friction“ ein, mit dem der Bereich benannt wird, an dem die Ziele des Systems mit den Zielen des Menschen kollidieren. Dieser muss nach dem Stadienmodell von Mackeprang et al. zunächst identifiziert, die widersprüchlichen Zielsetzungen dabei erkannt und in Folge alle berücksichtigt werden. Um diesen „point of friction“, der als „Aspekt der Interessenkollision“ übersetzbar wäre, weiter zu definieren, empfehlen sie die Übersetzung des Problems in eine modellhafte Darstellung. Dazu eignet sich beispielsweise die Abbildung des Prozesses in einem Fluss-Diagramm, welches die untersuchte Aufgabe in weitere Teilaufgaben untergliedert und deren Zusammenhänge aufzeigt.

Um bewerten zu können, wann die optimale Balance zwischen den gegensätzlichen Zielsetzungen erreicht ist, müssen in einem nächsten Stadium Metriken zu deren Messung definiert werden. Diese sollten nach Mackeprang et al. in der Lage sein, erstens die menschliche, zweitens die algorithmische sowie drittens die integrative menschlich-algorithmische Sichtweise einer Bewertung zuführbar zu machen. Insbesondere der integrativen Sichtweise kommt dabei ein hoher Stellenwert zu, da hybride Arbeitsabläufe, in denen Mensch

und Computer partnerschaftlich zusammenarbeiten, eine qualitative Veränderung in Bezug darauf bewirken, wie Menschen ein technisches System wahrnehmen [14].

Im nächsten Schritt folgt die Definition der sogenannten Grenzkonfigurationen. Diese beschreiben den jeweils niedrigsten und höchsten denkbaren Automatisierungsgrad für das untersuchte Problem. Dabei bedienen sich Mackeprang et al. der zuvor erläuterten zehn LoAs nach Sheridan et al. [35]. Danach ergibt sich eine Konfiguration durch einen festen LoA und eine Aufgabenzuweisung, die alle abgeleiteten Teilaufgaben entweder dem Menschen oder dem Computer eindeutig zuordnet. Mackeprang et al. [29] gehen davon aus, dass sich die optimale Konfiguration, zwischen diesen beiden Grenzkonfigurationen befinden muss. Sie benennen diese optimale Konfiguration als „Sweet-Spot-Konfiguration“. Schließlich werden im weiteren Verlauf des Modells die dazwischenliegenden LoAs, auch Zwischenkonfigurationen genannt, ebenfalls in Konfigurationen mit eigenen Aufgabenzuweisungen übersetzt [29].

Nachdem das zu untersuchende Problem vollständig formal beschrieben wurde, können die entwickelten Konfigurationen in funktionale Prototypen überführt, getestet und deren Leistungsfähigkeit basierend auf den entwickelten Leistungsmetriken erfasst werden. Die leistungsfähigste aller Konfiguration erweist sich abschließend als die gesuchte Sweet-Spot-Konfiguration.

4. Ausgangszustand

Bevor die Untersuchung des vorliegenden Automatisierungsproblems beginnt, wird im folgenden Abschnitt eine Übersicht über den Ausgangszustand gegeben. Dazu wird die Einbettung der Webanwendung für den Bildzuschnitt in die Systemarchitektur der Significo GmbH erläutert sowie deren Aufbau und Grundfunktionalitäten beschrieben.

4.1. Systemarchitektur

Innerhalb der Significo GmbH wird die Speicherung, die Verwaltung sowie der Zuschnitt und die Auslieferung von Bildmedien über drei miteinander verbundene Systemkomponenten realisiert. Im Zentrum steht das Digital Asset Management (DAM) System, welches für die Speicherung und Verwaltung der Bildmedien verantwortlich ist. Das Erstellen von Bildzuschnitten erfolgt über eine hauseigene Webanwendung, während die Auslieferung der Zuschnitte an die Nutzer-Anwendungen über einen DAM-Service stattfindet. Abbildung 3 visualisiert diese Systemkomponenten und deren Zusammenhänge schematisch.

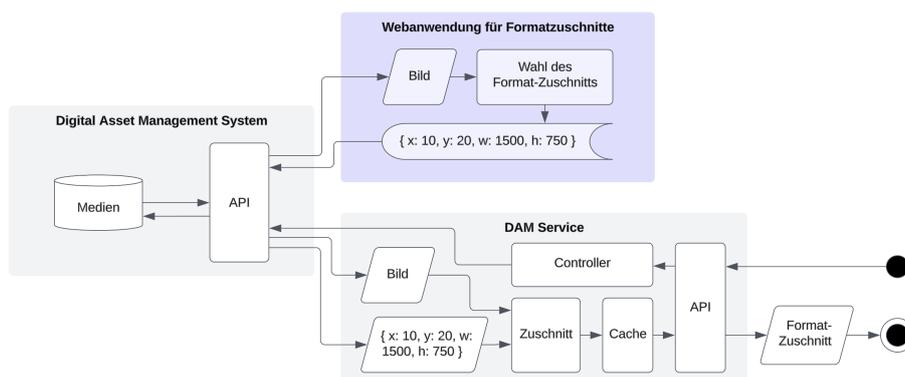


Abbildung 3: Systemarchitektur

Das DAM-System ist für die zentrale Speicherung und Verwaltung von Bildmedien zuständig. Es bietet Contentmanager:innen verschiedene Möglichkeiten, Bildmedien inhaltlich zu verwalten. Bildmedien werden dabei sowohl in höchstmöglicher Auflösung als auch im originalen Bildformat gespeichert. Über systemeigene Schnittstellen (engl.: Application Programming Interface, kurz: API) besteht die Möglichkeit, dass andere Systeme und Anwendungen auf die gespeicherten Bildmedien zugreifen und diese bei Bedarf auch anpassen können.

Der Bildzuschnitt selbst erfolgt wiederum über eine eigenständige Webanwendung. Diese ist über die DAM-API mit dem DAM-System verbunden, wodurch Nutzer:innen Zugriff auf alle Bildmedien haben, die zuvor im DAM-System gespeichert wurden. Im Rahmen

des Zuschnitts müssen für jedes Bild sieben vordefinierte Bildzuschnitte gewählt werden. Ein Bildzuschnitt wird als rechteckiger Ausschnitt eines Bildes definiert, welcher durch eine Kombination aus vier Kennwerten beschrieben werden kann. Abbildung 4 verdeutlicht diesen Zusammenhang visuell. Das rote Rechteck zeigt den gewählten Bildausschnitt. Dieser wird eindeutig spezifiziert durch die Kombination aus x-Koordinate und y-Koordinate der oberen linken Ecke, sowie Breite und Höhe des Rechtecks. Ein Zuschnitt ist also kein eigenständiges, physisches Bild, welches separat erzeugt, gespeichert und verwaltet werden muss, sondern lediglich ein Datensatz aus den zuvor beschriebenen Kennwerten. Die Angabe der Kennwerte erfolgt dabei in Pixeln und bezieht sich auf die Auflösung des Originalbildes, so wie es im DAM-System abliegt. Die Kennwerte der verschiedenen Bildzuschnitte werden von der Webanwendung an das DAM-System übermittelt und innerhalb der Bildreferenz in einem eigens dafür definierten Meta-Datenfeld gespeichert.

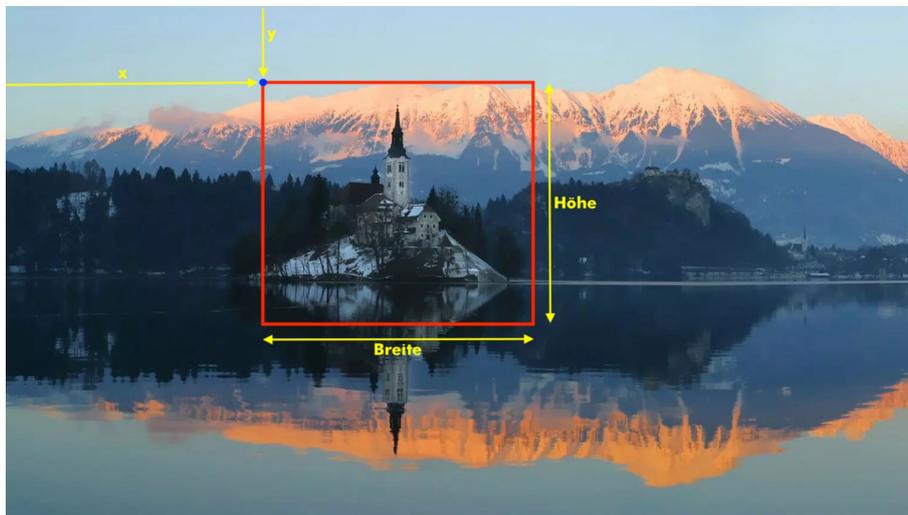


Abbildung 4: Kennwerte eines Bildzuschnitts

Die Generierung und Auslieferung des eigentlichen Zuschnitts erfolgt über einen speziell dafür entwickelten DAM-Service. Dieser ist ebenfalls über die DAM-API mit dem DAM-System verbunden und hat somit Zugriff auf die Bildmedien und deren Metadaten. Wird ein Bild von einer nativen App oder einer Webanwendung der Significo GmbH in einem speziellen Format beim DAM-Service angefragt, so nimmt dieser die Anfrage entgegen und fordert das entsprechende Bild sowie die notwendigen Zuschnittsinformationen vom DAM-System an. Da der Bildzuschnitt nicht als eigenständiges Bild im DAM-System abliegt, wird dieser vom DAM-Service erstellt und im Anschluss an die anfragende Resource ausgeliefert. Um schnellere Zugriffszeiten und eine Entlastung des DAM-Services zu erreichen, werden diese Zuschnitte in einem serviceeigenen Cache temporär gespeichert und müssen somit nicht bei jeder Anfrage neu erstellt werden. Die Bilddatei des jeweiligen Zuschnitts existiert also so lange der DAM-Service in Betrieb ist. Wird dieser heruntergefahren, gehen die Informationen im Cache und damit auch die tatsächliche

Bilddatei des Zuschnitts verloren. Bei einer erneuten Anfrage muss der Zuschnitt wieder neu generiert werden.

4.2. Webanwendung für Bildzuschnitte

Nachfolgenden betrachten wir Aufbau und Grundfunktionalität der Webanwendung für den Bildzuschnitt. Diese untergliedert sich in zwei zentrale Funktionsbereiche, die Bildauswahl und den Zuschnitt in das gewünschte Format.

Anhand der Abbildung 5 werden die einzelnen Bearbeitungsschritte beschrieben. Bei Aufruf der Webanwendung werden Anwender:innen zunächst alle Bildmedien angezeigt, welche im DAM-System gespeichert sind. Über eine Filtersektion am linken Rand besteht die Möglichkeit, die Bildmedien einzugrenzen. So können beispielsweise Bildmedien identifiziert werden, für die noch keine Zuschnittsinformationen definiert wurden oder deren Zuschnittsinformationen unvollständig sind. Andere Filteroptionen beziehen sich z.B. auf den Bildtitel, DAM spezifische IDs oder das Uploaddatum. Anwender:innen wählen im nächsten Schritt nun die Bilder aus, deren Format-Zuschnitte neu vergeben oder geändert werden sollen. Die Auswahl erfolgt über eine Markierung der entsprechenden Tabellenzeile. Sobald alle Bilder für die weitere Bearbeitung ausgewählt wurden, kann über den Button „Weiter“ zum nächsten Bearbeitungsschritt navigiert werden.

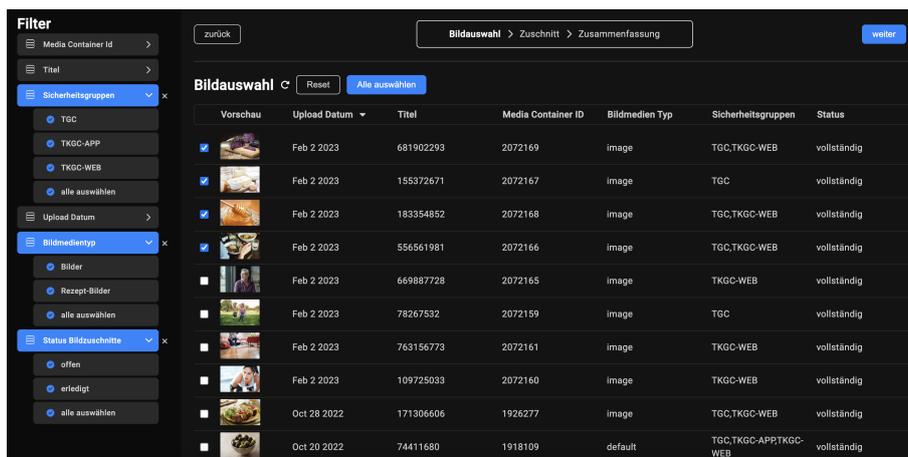


Abbildung 5: Webanwendung für den Bildzuschnitt - Bildauswahl

Der weitere Prozess des Bildzuschnitts kann in Abbildung 6 nachverfolgt werden. Diese Ansicht untergliedert sich von oben nach unten in drei Hauptbereiche. Die zuvor ausgewählten Bilder werden im Bereich „Bildauswahl“ in einer Karussell-Darstellung angezeigt. Der mittlere Bereich „Formatausschnitte“ umfasst die sieben vordefinierten Formate sowie die dafür bereits ausgewählten Bildzuschnitte. Der untere Bereich mit „Formatausschnitte wählen“ und „Vorschau“ realisiert die eigentliche Zuschnittsfunktion. Durch die manuelle Größenveränderung und das Verschieben des Zuschnittrechtecks

innerhalb des Bereichs „Formatausschnitt wählen“ kann der Bildzuschnitt manuell gewählt werden. Über „Ausschnitt wählen“ wird die vorherige Auswahl mit dem neuen Ausschnitt, der in der Vorschau zu sehen ist, überschrieben. Durch die Schaltflächen „Speichern“ oder „Speichern und weiter“ werden die vorgenommenen Änderungen an das DAM-System gesendet und in dem dafür vorgesehen Metadatenfeld gespeichert.

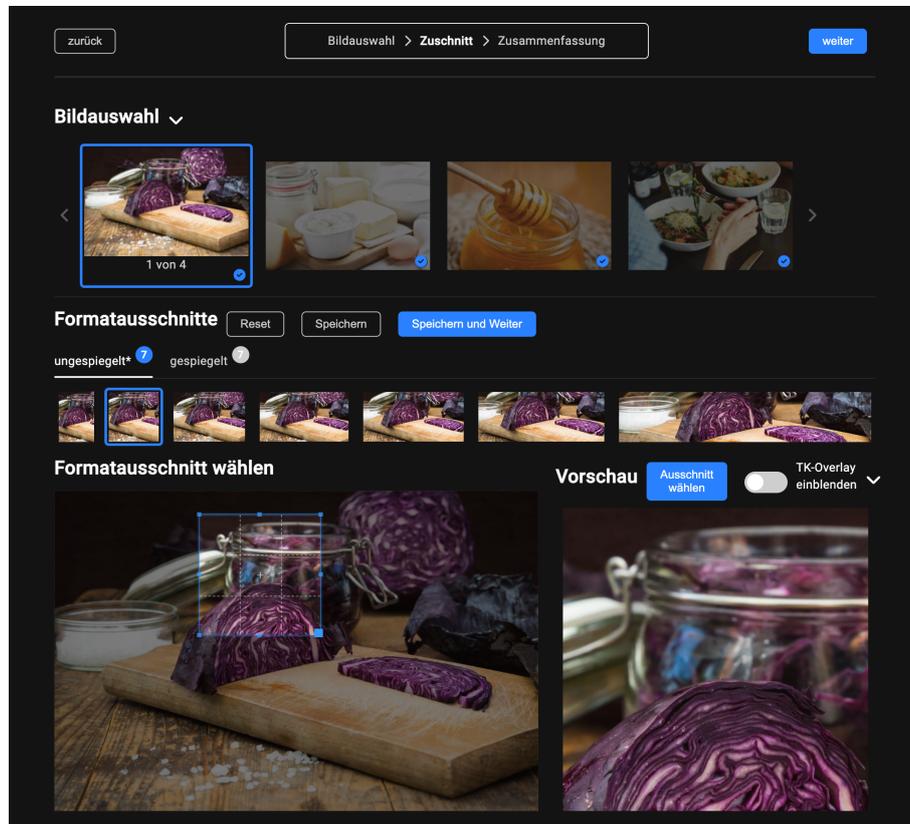


Abbildung 6: Webanwendung für den Bildzuschnitt - Formatauswahl

5. Analyse

Nachfolgend wird die Automatisierungsaufgabe den Stufen von Mackeprang et al. [29] folgend systematisch analysiert und die zur Lösung der Problematik entscheidenden Festlegungen getroffen.

5.1. Identifikation des „point of friction“

Die korrekte Auswahl eines Bildzuschnitts ist maßgeblich abhängig von dem inhaltlichen Kontext, in dem dieser verwendet werden soll. So können in einem Bild, wie in Abbildung 7 illustriert, verschiedene sogenannte „areas of interest“ (dt.: Bereiche von Interesse) definiert werden, die alle für sich ein Kommunikationsziel bzw. Interessensbereich bedienen und in unterschiedlichen inhaltlichen Kontexten Sinn ergeben. Dabei beschreibt ein „area of interest“ (AoI) einen Bildbereich mit spezifischer Aussagekraft. Ein Hersteller von Outdoor-Equipment würde möglicherweise einen Bildausschnitt mit dem Zelt in Form eines Produktfotos verwenden, Reisebüros würden eventuell das Seepanorama bevorzugen, während die Person auf dem Felsen für Stressreduktionskampagnen im Healthcare Kontext passen könnte.



Abbildung 7: Kontextabhängige Motivwahl

Veraltete Zuschnittsalgorithmen, wie wir sie schon seit vielen Jahren kennen, generierten Format-Zuschnitte ohne jeglichen Bezug zu Bildinhalten. Die sehr einfachen Algorithmen basierten auf festen Logiken. Dabei wurden Zuschnitte ausgehend von einer festgelegten Position, z.B. der Bildmitte, generiert. Das bereits etwas fortgeschrittenere Softwarepaket „ImageMagick“, das erstmalig 1990 auf den Markt kam, verfügte dann bereits über erste automatische Zuschnittsfunktionen [1]. Ermöglicht wurde das durch eine Analyse der Bildpixel [1], wodurch erstmals auch Bildinhalte miteinbezogen werden konnten. Der Umfang, mit dem Bildinhalte analysiert wurden, war allerdings nicht mit modernen Zuschnittsalgorithmen vergleichbar. Moderne Bildbearbeitungsanwendungen kombinieren verschiedene Ansätze der Bildanalyse und nutzen multiple Algorithmen, um die

Zuschnittsqualität zu verbessern. Mithilfe dieser Algorithmen werden hervorstechende Bereiche eines Bildes auf Basis von Pixel-Helligkeit und Pixel-Farben identifiziert, sowie Methoden zur Gesichtserkennung, Objekterkennung oder der Erkennung von Texten eingesetzt [4, 30]. Jedoch sind diese Algorithmen bislang nicht in der Lage, die identifizierten AoIs mit inhaltlichen Kontexten, in welchem die Bildzuschnitte verwendet werden sollen, zuverlässig und sinnvoll zu verknüpfen. So erkennen moderne Algorithmen die Person in Abbildung 7 auf dem Felsen mittels Gesichtserkennung und können diese Information beim Zuschnitt auch berücksichtigen. Für die Einschätzung, ob diese Person für den inhaltlichen Kontext, in dem der Zuschnitt verwendet werden soll, relevant ist, liefern auch moderne Smart-Cropping-Algorithmen keine zuverlässige Lösung. Die ästhetische Bildkomposition ist für moderne Algorithmen auch weiterhin eine große Herausforderung, wengleich erste vielversprechende Lösungsansätze, die Relevanz dieser Problematik zukünftig abschwächen könnten, aktuell entwickelt werden [47, 43, 27].

Es sind also vor allem diese zwei Aspekte, Kontextbezug und Ästhetik, bei denen moderne Algorithmen weiterhin an ihre Grenzen stoßen und lediglich ein menschlicher Beitrag die Qualität der Zuschnitte verbessern kann. Dem gegenüber steht das Bedürfnis von Nutzer:innen nach einem unkomplizierten, komfortablen und vor allem effizienten Prozess für den Zuschnitt von Bildmedien, bei dem nicht jeder Schritt manuell durchgeführt werden muss. Das Interesse stößt allerdings an dem Punkt an seine Grenzen, an dem die Unterstützung durch algorithmenbasierte Automatisierung zu unzuverlässigen Ergebnissen führt. Genau dieser Zielkonflikt der beiden Interaktionspartner:innen ist der „point of friction“ des in dieser Arbeit vorliegenden Automatisierungsproblems. Ziel ist es, einen Automatisierungsgrad zu entwickeln, der eine Balance zwischen den Kompetenzen und Interessen beider Interaktionspartner:innen herstellt und deren Zusammenarbeit optimiert.

5.2. Den „point of friction“ im Arbeitsablauf lokalisieren

Der Prozess des Bildzuschnitts setzt sich aus zwei Hauptaufgaben zusammen – der Bestimmung des AoI und der Bestimmung der Bildkomposition. Die Bildkomposition beschreibt dabei die Art und Weise, wie der AoI im ausgewählten Bildausschnitt positioniert ist, sowie den flächenmäßigen Anteil des AoIs am gesamten Bildzuschnitt. Abbildung 8 zeigt ein Flussdiagramm für diese beiden Hauptaufgaben.

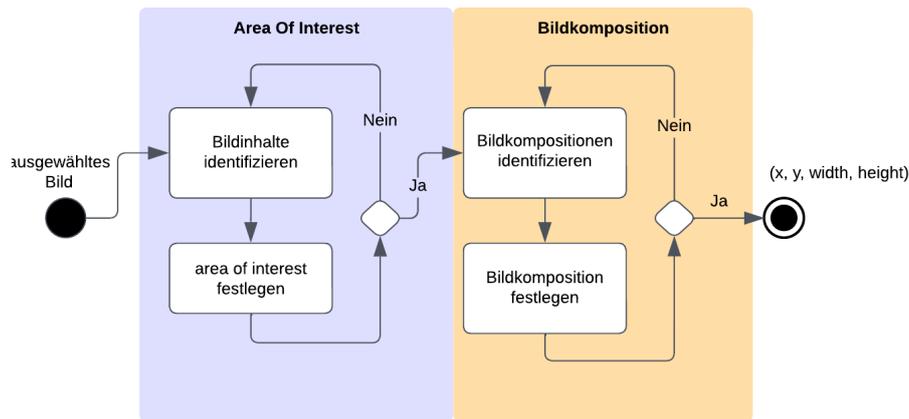


Abbildung 8: Arbeitsablauf als Flussdiagramm

Der Prozess des Bildzuschnitts startet nachdem ein Bild für den Zuschnitt ausgewählt wurde. Wie bereits angeführt, kann dabei jedes Bild je nach Motiv mehrere AoIs beinhalten. So können mehrere Szenerien, Gegenstände oder Personen potenziell von Bedeutung sein und müssen im ersten Schritt identifiziert werden. Im Anschluss erfolgt die Bestimmung des AoI, also der Bereich eines Bildes, der für den inhaltlichen Kontext, in dem der Zuschnitt verwendet werden soll, am passendsten erscheint. Ist der AoI festgelegt, wird im nächsten Schritt der Bildausschnitt festgelegt, welcher den AoI am treffendsten im Bildzuschnitt positioniert. Dazu wird der Bereich um den AoI herum analysiert und mögliche Bildkompositionen identifiziert. Anschließend wird eine dieser Bildkompositionen final ausgewählt. Als Endergebnis des Prozesses ergeben sich Kennwerte, welche den gewählten Bildzuschnitt eindeutig beschreiben.

5.3. Leistungsmetriken zur Bewertung von Konfigurationen

Dem Prozess nach Mackeprang et al. [29] folgend werden im nächsten Schritt die Leistungsmetriken definiert. Diese sollen sowohl die unterschiedlichen Zielsetzungen der Interaktionspartner:innen als auch eine integrative Perspektive abbilden (siehe Abschnitt 3).

Das Ziel aus Anwender:innenperspektive ist, deren Aufwand so gering wie möglich und den Prozess so komfortabel wie möglich zu gestalten. Auf diese Weise soll der für den manuellen Bildzuschnitt aufgewandte Zeitaufwand der Anwender:innen gesenkt werden, sodass diese Personalressourcen für andere Tätigkeiten frei werden. Die potenzielle Zeiteinsparung kann nun auf unterschiedliche Art und Weise erfasst werden. So kann einerseits die Zeit gemessen werden, die notwendig ist, um einen Bildzuschnitt zu erstellen. Andererseits können Nutzer:innen aufgefordert werden, ihren subjektiv erlebten Arbeitsaufwand zu beurteilen. Beide Maße werden in der vorliegenden Arbeit zur Untersuchung der verschiedenen Prototypen genutzt. Der subjektiv erlebte Aufwand der Nutzer:innen wird über den Nasa-Raw-Task-Load-Index (RAW-TLX) erfasst. Dabei handelt es sich

um einen weitverbreiteten Fragebogen, welcher die subjektiv wahrgenommene Arbeitsbelastung erfasst [13]. Die Belastung wird dabei anhand sechs verschiedener Dimensionen bewertet. [3]. Die Dimensionen sind: geistige Beanspruchung, körperliche Beanspruchung, erlebter Zeitdruck, Leistung, Grad der Anstrengung und Grad der Frustration [3].

Das Ziel aus algorithmischer Perspektive ist, möglichst passende Bildzuschnitte zu generieren. Aus diesem Grund muss ihre Qualität erfasst bzw. beurteilt werden. Eine Beurteilung der Zuschnittsqualität ist nicht einfach, da keine standardisierten Messgrößen zur Verfügung stehen. Eine Methode zur Lösung des Problems wäre der systematische Paarvergleich aller Ergebnisse der algorithmusgenerierten Bildzuschnitte (also die durch die Anwender:innen noch nicht weiter optimierten Zuschnitte) mit jeweils definierten Referenzzuschnitten im Sinne von „Goldstandards“ durch ausgewählte Expert:innen. Als Goldstandard stünde dazu der manuelle Zuschnitt zur Verfügung, der ausschließlich von menschlichen Fachpersonen generiert wird. Unabhängige Expert:innen vergleichen dann für jedes Format von jedem bearbeiteten Bild paarweise diesen „Goldstandard“ mit den algorithmusgenerierten Zuschnitten hinsichtlich der Qualität des Zuschnitts. Aus diesen binären Urteilen (besser vs. schlechter) von Expert:innen ließen sich mathematische Kennwerte berechnen, die Grade der Qualitätsabweichung vom Goldstandard abbilden. Ein entsprechendes Vorgehen erschien allerdings mit gewissen Problemen behaftet. Ein Grundproblem entsteht auf Seiten der Definition des Goldstandards. Wählt man dazu die von Anwender:innen rein manuell zugeschnittenen Ausschnitte aus, spielt natürlich das subjektive Qualitätserleben der bearbeitenden Person eine zentrale Rolle. Würde dasselbe Bild von mehreren Anwender:innen unabhängig bearbeitet, ergäben sich sehr wahrscheinlich zumindest leicht oder auch stärker voneinander abweichende Ergebnisse. Der Goldstandard hätte damit auch eine Streuung. Je größer die Zahl der bearbeiteten Bilder und die Zahl der bearbeitenden Personen, desto mehr würde sich diese Streuung in der Gesamtbewertung relativieren. Aus durchführungspraktischen Gründen musste die Zahl bearbeiteter Bilder und die Zahl bearbeitender Personen allerdings stark begrenzt bleiben, worunter sich die Qualität des so definierten Goldstandards verringert hätte. Ein weiteres Problem resultiert daraus, dass für die Expert:innenbeurteilung korrekterweise eine von den Bearbeiter:innen unabhängige Personengruppe hätte herangezogen werden müssen. Auch dies wäre in dem geplanten Untersuchungskontext aufgrund der Personalsituation nur schwierig oder gar nicht zu realisieren gewesen. Entsprechend erschien letztlich der Einsatz der Methode des Paarvergleichs als praktisch nicht umsetzbar.

Stattdessen wurde ein alternativer Ansatz für die Beurteilung der Zuschnittsqualität gewählt. Dieser basiert auf der realitätsnahen Annahme, dass zuschneidende Personen in der Rolle von Praxisexpert:innen die algorithmisch generierten Zuschnitte immer dann manuell anpassen, wenn sie nicht den Qualitätsanforderungen dieser Expert:innen gerecht werden. Vor diesem Hintergrund lässt sich ein abgeleitetes Maß für die Zuschnittsqualität generieren, das den Aufwand im Rahmen einer ggf. erforderlichen Nachbearbeitung der algorithmusbasierten Vorschläge abbildet. Dabei wird die Zuschnittqualität nicht direkt gemessen. Stattdessen wird der zusätzliche Bearbeitungsaufwand, der notwendig ist, um die subjektiv wahrgenommene Qualität herzustellen, als Qualitätsmaß

definiert. Dieser erbrachte Aufwand basiert auf zwei Basisoperationen, die beide sehr einfach messbar sind. Einerseits ist es die Verschiebung des Zuschnittfensters, andererseits ist es die Vergrößerung beziehungsweise Verkleinerung des Zuschnittfensters. Dabei lässt sich die Verschiebung über die Distanz zwischen dem Bildmittelpunkt des vorgeschlagenen Zuschnitts und dem Bildmittelpunkt des final gewählten Zuschnitts messen, was der dafür erforderlichen Diagonalebewegung des Cursors entspricht. Ein Maß für die Vergrößerung bzw. Verkleinerung ergibt sich ebenfalls aus der Diagonalebewegung des Cursors, die erforderlich ist, um die Größe des Zuschnittfensters zu verändern. Beide Maße können aus den Zuschnittsdaten errechnet und in der Einheit „Pixelanzahl“ angegeben werden. Die Abbildungen 9 und 10 zeigen nochmals die zugrundeliegenden Zusammenhänge für die Verschiebung des Rechtecks, während Abbildung 11 und 12 die Zusammenhänge für die Vergrößerung bzw. Verkleinerung visuell verdeutlichen.

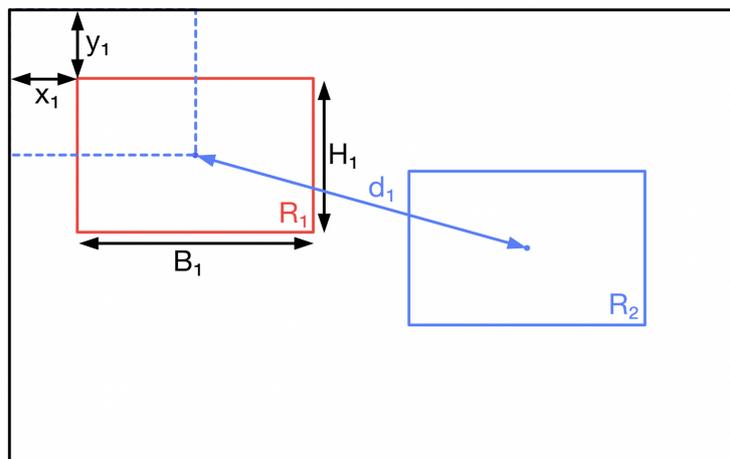


Abbildung 9: Verschiebung des Zuschnittfensters

$$d_1 = \sqrt{\left(\left(x_2 + \frac{B_2}{2}\right) - \left(x_1 + \frac{B_1}{2}\right)\right)^2 + \left(\left(y_2 + \frac{H_2}{2}\right) - \left(y_1 + \frac{H_1}{2}\right)\right)^2}$$

Abbildung 10: Formel zur Berechnung der Verschiebung des Zuschnittfensters

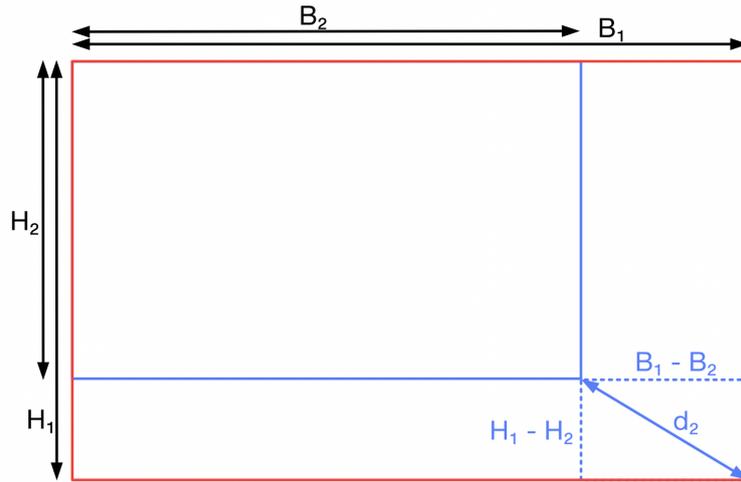


Abbildung 11: Vergrößerung bzw. Verkleinerung des Zuschnittfensters

$$d_2 = \sqrt{(H_1 - H_2)^2 + (B_1 - B_2)^2}$$

Abbildung 12: Formel zur Berechnung der Vergrößerung bzw. Verkleinerung des Zuschnittfensters

Da Bilder mit jeweils unterschiedlichen Auflösungen zugeschnitten werden, müssen die so errechneten pixelbezogenen Werte hinsichtlich der Auflösung des zugrundeliegenden Bildes noch standardisiert werden. Dazu werden die ermittelten Veränderungswerte in Relation zur Anzahl der Pixel der Bilddiagonalen des Originalbildes gesetzt, da diese der jeweils maximal möglichen Veränderung entspricht, mit der Bildmittelpunkt bzw. Zuschnittfenster verschoben werden können.

Wie bereits in Abschnitt 3 erläutert, argumentieren Mackeprang et al. [29] in ihrer Arbeit, dass zusätzlich zu menschlichen und algorithmischen Sichtweisen auch eine integrative, menschlich-algorithmische Sichtweise einbezogen werden muss. So kann die Delegation von Aufgaben, die zuvor von Menschen ausgeführt wurden, an technologische Akteur:innen zu bedeutenden Verschiebungen in sozialen Praktiken und Verantwortlichkeiten führen [23]. Auch die eingangs ausführlich erläuterten Dequalifizierungseffekte spielen hier eine zentrale Rolle. Indem interaktive, intelligente Systeme Aufgaben unterstützen oder selbst übernehmen, ändern sie die Art und Weise, wie Anwender:innen sich selbst im Zusammenspiel mit dem Gesamtsystem erleben. Es wird also ein Maß benötigt, das dabei hilft abzubilden, wie Menschen interaktive, intelligente Systeme wahrnehmen. Im Kontext von Empfehlungssystemen eignet sich dafür besonders gut der ResQue Fragebogen [33]. Dieser wurde speziell entwickelt, um die Qualität der Benutzer:innenerfahrung im Umgang mit Empfehlungssystemen zu messen. Der Fragebogen umfasst dabei fünfzehn Dimensionen und stellt pro Dimension bis zu vier Fragen bereit.

Der Fragebogen ist so entwickelt, dass je nach Anwendungsfall der Umfang und der Fokus durch die Wahl der Dimensionen und Fragen auf die vorliegenden Gegebenheiten adaptiert werden können. Da es sich bei dem zu untersuchenden Automatisierungsproblem um ein Empfehlungssystem im klassischen Sinne handelt, wurde dieser Fragebogen für die Messung der integrativen menschlich-algorithmischen Sichtweise gewählt.

5.4. Grenzkonfigurationen festlegen

Im nächsten Schritt sollen nun die Grenzkonfigurationen definiert werden, welche die jeweils äußeren Grenzen des für das vorliegende Automatisierungsproblem relevanten LoA-Spektrums abbilden. Das gesamte LoA-Spektrum erstreckt sich von einer vollständig manuellen Realisierung des Aufgabenablaufs auf der einen bis hin zu einer vollautomatischen Bearbeitung auf der anderen Seite. Mackeprang et al. [29] argumentieren, dass der bereits identifizierte „point of friction“, die maximal möglichen Grenzkonfigurationen als Sweet-Spot-Konfiguration ausschließt, da diese von vornherein verhindern würden, die Ziele eine:r der Interaktionspartner:innen zu realisieren. Somit muss sich die Sweet-Spot-Konfiguration zwischen diesen beiden Grenzkonfigurationen befinden.

Im zu untersuchenden Fall wird die untere Grenzkonfiguration des LoA-Spektrums durch den LoA1 beschrieben. Der Zuschnitt von Bildmedien wird dabei vollständig manuell auf Grundlage menschlicher Arbeit vollzogen; er entspricht dem Ausgangszustand der Anwendung. Tabelle 2 zeigt die Aufgabenzuweisung dieser Konfiguration, in der alle Teilaufgaben dem Menschen zugeordnet sind.

	
Bild auswählen	
Bildinhalte identifizieren	
Area of Interest festlegen	
Bildkompositionen identifizieren	
Bildkomposition festlegen	
Bildkomposition bestätigen	

Tabelle 2: Aufgabenzuweisung der LoA1-Konfiguration

Die Definitionen von Sheridan et al. [35] und Parasuraman [31] et al. spezifizieren für jeden LoA individuell, wie Mensch und Computer in Bezug auf die Entscheidungsfindung miteinander interagieren. Allerdings gibt es keine Aussage darüber, wie Ergebnisse und Fehlentscheidungen im Nachhinein geändert werden können. Ab einem LoA6 übernimmt der Computer bereits eine zentrale Rolle in der Entscheidungsfindung und überlässt Anwender:innen lediglich eine zeitlich limitierte Möglichkeit des Widerspruchs. Es bestünde

im hier untersuchten Anwendungsfall das erhöhte Risiko, dass algorithmisch generierte, unpassende Bildzuschnitte gespeichert und an Nutzer:innen der Produkte der Significo GmbH ausgespielt werden. Unternehmerische und arbeitsorganisatorische Vorgaben des Unternehmens fordern jedoch einen sicheren Prozess zur Wahrung der Zuschnittsqualität, weshalb Interventionsmöglichkeiten für den Menschen zwingend notwendig sind. Aus diesem Grund wurden alle Automatisierungsgrade von Level 6 bis Level 10 aus der weiteren Untersuchung ausgeschlossen. Lediglich Automatisierungen der Level 1 bis 5 werden wegen der dort noch bestehenden Interventionsmöglichkeiten weiter berücksichtigt. Folglich ergibt sich als obere Grenzkonfiguration eine Level-5-Automatisierung.

Die Aufgabenzuweisung des LoA5 bildet bereits eine stark algorithmisch geprägte Sichtweise ab. Der Computer übernimmt alle Teilaufgaben im Rahmen des Bildzuschnitts, von der Bestimmung des AoI bis zur Festlegung der Bildkomposition. Der Mensch nimmt dabei eine beobachtende, kontrollierende Rolle ein und handelt lediglich bei unpassend gewählten Zuschnitten.

	
Bild auswählen	
	Bildinhalte identifizieren
	Area of Interest festlegen
	Bildkompositionen identifizieren
	Bildkomposition festlegen
Manuelle Anpassung vornehmen, falls erforderlich	
Bildkomposition bestätigen	

Tabelle 3: Aufgabenzuweisung der LoA5-Konfiguration

5.5. Zwischenkonfigurationen definieren

Ausgehend von den zuvor definierten Grenzkonfigurationen können nun weitere Zwischenkonfigurationen abgeleitet werden, indem die Definitionen der dazwischenliegenden LoAs herangezogen und in Konfigurationen für den Aufgabenablauf überführt werden. In diesem Abschnitt werden nun diese Konfigurationen genauer spezifiziert und deren Aufgabenzuweisung näher erläutert.

Der Bildzuschnitt beginnt, nachdem ein Bild von den Anwender:innen ausgewählt wurde, weshalb die Bildauswahl den Ausgangspunkt für alle nachfolgend definierten Konfigurationen darstellt. Tabelle 4 zeigt die Aufgabenzuweisung der Level-2-Konfiguration.

	
Bild auswählen	
Bildinhalte identifizieren	
Area of Interest festlegen	
Bildkompositionen identifizieren	
Bildkomposition festlegen	
Bildkomposition bestätigen	

Tabelle 4: Aufgabenzuweisung der LoA2-Konfiguration

Auf den ersten Blick unterscheidet sich diese nicht von der Level-1-Grenzkonfiguration. Der Mensch übernimmt weiterhin alle notwendigen Teilaufgaben im Prozess des Bildzuschnitts. Dennoch unterscheiden sich die beiden Konfigurationen in der Interaktionsform. Die Level-2-Konfiguration basiert auf einem Zuschnittsalgorithmus, dessen Ergebnisse durch manuelle Eingaben beeinflusst werden können. So kann neben einem manuell gewählten AoI auch ein Skalierungsfaktor übergeben werden. Dieser steuert die Größe des Bereichs um den AoI, der weiterhin für den Bildzuschnitt berücksichtigt werden soll. Auf diese Weise können algorithmisch generierte Vorschläge berücksichtigt werden, die maßgeblich von den zuvor getroffenen menschlichen Entscheidungen abhängen.

Die Aufgabenzuweisung der Level-3-Automatisierung (siehe Tabelle 5) ist durch eine algorithmisch gewählten AoI sowie durch eine algorithmisch generierte Bildkompositionen gekennzeichnet. Der Computer übernimmt hier bereits eine zentrale Rolle im Prozess des Bildzuschnitts. Anwender:innen erhalten verschiedene Vorschläge für mögliche Bildkompositionen vom System, wählen die passendste aus, passen sie bei Bedarf noch an und bestätigen das Endergebnis.

	
Bild auswählen	
	Bildinhalte identifizieren
	Area of Interest festlegen
	Bildkompositionen identifizieren
	Vorschläge für verschiedene Bildkompositionen generieren
Eine der vorgeschlagenen Bildkompositionen auswählen	
Manuelle Anpassung vornehmen, falls erforderlich	
Bildkomposition bestätigen	

Tabelle 5: Aufgabenzuweisung der LoA3-Konfiguration

Die Aufgabenzuweisung der Level-4-Konfiguration ähnelt stark der Level-3-Konfiguration. Der entscheidende Unterschied ist hierbei, dass bereits einer der algorithmisch generierten Vorschläge durch den Computer empfohlen und vorausgewählt wird. Anwender:innen haben allerdings weiterhin die Möglichkeit der Intervention und können einen alternativen Vorschlag wählen oder auch manuelle Anpassungen vornehmen (siehe Tabelle 6).

	
Bild auswählen	
	Bildinhalte identifizieren
	Area of Interest festlegen
	Bildkompositionen identifizieren
	Vorschläge für verschiedene Bildkompositionen generieren
	Eine Vorauswahl auf eine Bildkomposition treffen
Manuelle Anpassung vornehmen, falls erforderlich oder alternativen Vorschlag wählen	
Bildkomposition bestätigen	

Tabelle 6: Aufgabenzuweisung der LoA4-Konfiguration

5.6. Ausschluss nicht geeigneter Zwischenkonfigurationen

Für die Umsetzung der algorithmisch generierten Bildzuschnitte werden spezielle Funktionen des DAM-Systems genutzt. Für alle Medien, die im DAM-System verwaltet werden, besteht somit die Möglichkeit automatische Zuschnitte zu generieren. Dabei sind im DAM-System die Algorithmen, die für den automatischen Zuschnitt notwendig sind, nicht als solche selbst implementiert, sondern es greift auf bereits existierende Microsoft-Services zurück. Microsoft bietet über die Azure Plattform bereits seit einigen Jahren Möglichkeiten für den automatischen Zuschnitt von Bildmedien an [2]. Das DAM-System stellt diese Microsoft-Services für alle Bildmedien im DAM-System bereit. Die Abhängigkeit von den Microsoft-Services zieht jedoch Einschränkungen bezüglich der Funktionalität mit sich, welche die Implementierung einer der Zwischenkonfigurationen maßgeblich beeinflusst und in Teilen limitiert.

So ist es zwar möglich, Zuschnitte auf Basis eines algorithmisch festgelegten AoI zu generieren und dabei den Bereich, der um eine AoI herum sichtbar ist, über manuell vorgegebene Skalierungsfaktoren zu begrenzen. Allerdings ist der um einen AoI herum sichtbar ist. Allerdings ist der algorithmische Agent nicht in der Lage, den verschiedenen Bildkompositionsvarianten ein entsprechendes Konfidenzniveau zuzuordnen, d.h. diese bezüglich inhaltlichem Kontext und Ästhetik zu bewerten und zu gewichten. Die im vorangegangenen Kapitel beschriebene Level-4-Konfiguration erfordert jedoch eine algorithmisch generierte Empfehlung der passendsten Bildkomposition. Eine solche Auswahl kann aber nur dann vorgenommen werden, wenn Konfidenzniveaus der generierten Bildkompositionsvarianten vorliegen. Alternativ wäre eine Entscheidung basierend auf fest definierten Regeln und Logiken denkbar, würde dann hingegen nicht mehr den Anforderungen einer intelligenten Empfehlung entsprechen. Diese Einschränkung verhindert somit eine Implementierung der zuvor definierten Level-4-Konfiguration, weshalb diese aus dem Untersuchungsumfang ausgeschlossen wurde.

6. Gestaltung

Die Überführung der in Kapitel 5 erarbeiteten Konfigurationen in visuelle Prototypen basiert auf der Anwendung Figma, einem Tool für die Entwicklung von Prototypen in den Bereichen UX- und UI-Design. Für ein besseres Verständnis der einzelnen Konfigurationen und deren individuellen Unterschieden wird nachfolgend jeder Prototyp einzeln vorgestellt und dessen Interaktionsdesign erläutert. Bei unzureichender Qualität der algorithmisch generierten Vorschläge ermöglicht jeder der vorgestellten Prototypen weiterhin den manuellen Eingriff durch Anwender:innen. Auf diese Weise wird die Einhaltung unternehmensspezifischer Vorgaben zur Sicherstellung der Zuschnittsqualität durchgehend gewährleistet.

6.1. Level-1-Konfiguration (LoA1)

Die Level-1-Konfiguration entspricht, dem Ausgangszustand des zu untersuchenden Automatisierungsproblems. Eine detaillierte Beschreibung erfolgte bereits in Kapitel 4, weshalb an dieser Stelle darauf verzichtet wird.

6.2. Level-2-Konfiguration (LoA2)

Die Aufgabenzuweisung der Level-2-Konfiguration sieht weiterhin vor, dass der Mensch jeden der notwendigen Teilschritte für den Zuschnitt von Bildmedien selbst ausführt. Unterstützt wird er dabei durch einen Zuschnittsalgorithmus, der auf Basis eines manuell gewählten AoI sowie eines manuell gewählten Skalierungsfaktors Vorschläge für mögliche Bildzuschnitte generieren kann. Zu Beginn des Zuschnitts müssen Nutzer:innen daher den AoI manuell über das Auswahlrechteck auswählen (siehe Abbildung 13). Im Anschluss wird automatisiert ein Vorschlag für einen Bildzuschnitt mit einem Skalierungsfaktor von 0,5 generiert und im rechten Bereich des Fensters unter „Vorschläge“ angezeigt (siehe Abbildung 14). Über das Slider-Element haben Nutzer:innen die Möglichkeit, den Skalierungsfaktor im Intervall zwischen 0 und 1 manuell zu verändern. Der algorithmus-generierte Vorschlag für den Bildzuschnitt aktualisiert sich entsprechend des gewählten Parameters. Dabei wird der Vorschlag nicht automatisch in die Auswahl des Zuschnittfensters übernommen. Anwender:innen müssen diese Übernahme über einen Klick auf den generierten Vorschlag bewusst anstoßen. Der Computer nimmt also ausschließlich eine vorschlagende Rolle ein, während der Mensch weiterhin die volle Handlungs- und Entscheidungsträger:innenschaft behält.

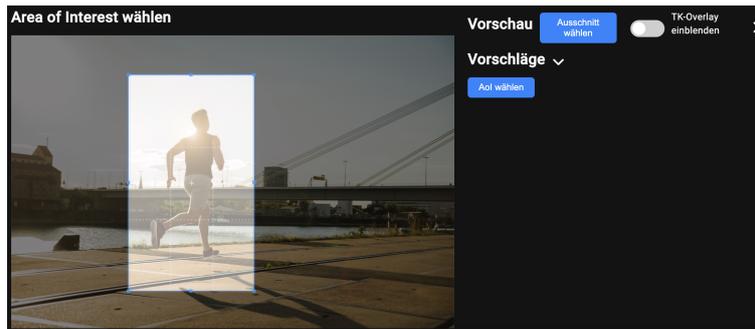


Abbildung 13: Level-2-Konfiguration – AoI-Auswahl

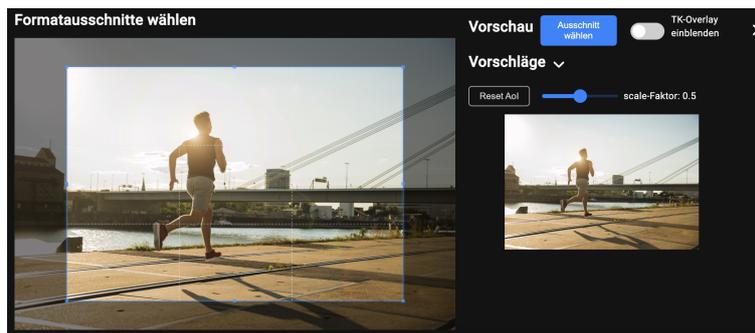


Abbildung 14: Level-2-Konfiguration – Bildauswahl

6.3. Level-3-Konfiguration (LoA3)

Die Level-3-Konfiguration gewährt dem Computer bereits mehr Autonomie und überlässt diesem die Wahl des AoI. Zudem generiert der Computer bereits einige Zuschnittsvorschläge aus denen Anwender:innen anschließend den passendsten Vorschlag auswählen. Das Interaktionsdesign sieht daher keine manuelle Wahl des AoI mehr vor. Nach der Auswahl eines Bildes werden Zuschnittsvorschläge algorithmisch generiert und auf der rechten Seite des Fensters unter „Vorschläge“ angezeigt (siehe Abbildung 15). Die Anzahl der generierten Vorschläge kann dabei variieren und ist abhängig von dem zuzuschneidenden Bild. Werden mehrere Vorschläge generiert, können diese im Bereich „Vorschläge“ durch Scrolling eingesehen werden. Die Level-3-Konfiguration trifft keine Vorauswahl auf einen der generierten Zuschnittsvorschläge. Initial ist daher auch keiner von diesen vorausgewählt. Anwender:innen müssen einen Vorschlag bewusst auswählen, damit dieser in das Zuschnittsfenster übernommen wird. Der Computer übernimmt durch die automatische Auswahl des AoI bereits mehr Entscheidungen und Handlungen, agiert jedoch weiterhin nur in einer vorschlagenden Rolle. Die Anwender:innen hingegen geben Entscheidungen ab und müssen weniger Teilaufgaben übernehmen. Durch die Möglichkeit bei unpassenden Vorschlägen einen Zuschnitt manuell vorzunehmen, behält der Mensch aber weiterhin die volle Handlungs- und Entscheidungsträger:innenschaft.

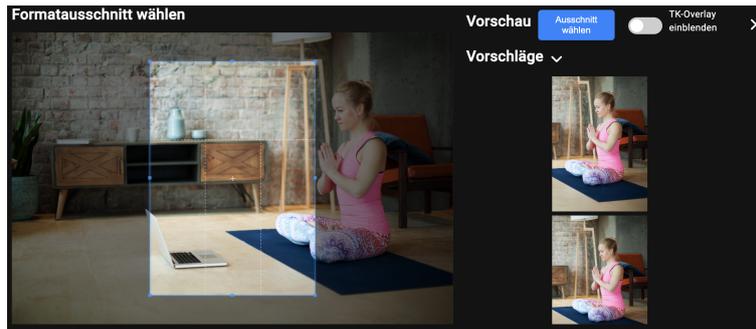


Abbildung 15: Level-3-Konfiguration

6.4. Level-5-Konfiguration (LoA5)

Die Gestaltung der Level-5-Konfiguration ermöglicht einen deutlich höheren Grad an Autonomie des unterstützenden Systems, während die Rolle der Anwender:innen sich in eine beobachtende Richtung mit kontrollierendem Charakter verlagert. Nachdem Anwender:innen die Bilder für den Zuschnitt ausgewählt haben, werden vollautomatisch Zuschnitte generiert und vorausgewählt (siehe Abbildung 16). Die autonom generierten Zuschnitte müssen nur noch geprüft und bestätigt bzw. bei Bedarf manuell korrigiert werden. Dazu kann durch die Auswahl des Bearbeiten-Symbols die Anwendung temporär in einen manuellen Modus überführt werden, sodass die Anwender:innen Korrekturen an den Bildzuschnitten vornehmen können.

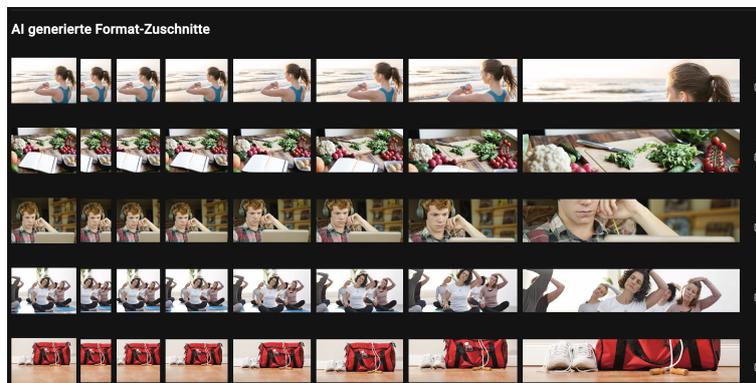


Abbildung 16: Level-5-Konfiguration

7. Implementierung

Im Rahmen der Implementierung wurden die erarbeiteten Figma-Prototypen vom Autor dieser Arbeit in funktionalen Programmcode überführt. Dieser basiert neben den notwendigen Webtechnologien HTML, CSS und JavaScript auf dem JavaScript Framework Angular. Angular ermöglicht eine gute Codestruktur und verfügt durch dessen tiefe TypeScript-Integration über eine statische Typisierung und schafft somit die Voraussetzungen für eine effiziente, sichere und modulare Entwicklung. Der entwickelte Programmcode wird mithilfe des Versionsverwaltungsprogramms Git organisiert und über die Webanwendung GitLab in einem Repository¹ verwaltet. Zu Beginn dieser Arbeit existierte der Programmcode der Ausgangskonfiguration bereits. Basierend darauf wurden die drei Prototypen jeweils in einer eigenständigen Branch entwickelt und verwaltet.

Für die algorithmische Lösung des Zuschnittproblems werden zwei bestehende Web-Services genutzt, die vom Hersteller des Digital Asset Management Systems als Feature bereitgestellt werden. Diese ermöglichen den Zuschnitt von im DAM-System gespeicherten Bildmedien, werden im aktuellen Workflow allerdings nicht genutzt. Die algorithmische Generierung der Zuschnitte selbst wird dabei von bestehenden Azure-Services übernommen. Bei der Auslieferung der Zuschnitte gibt es zwei grundlegend verschiedene Möglichkeiten. Zum einen können die Zuschnitte selbst als Bilddateien ausgeliefert werden. Zum anderen kann lediglich die Information über die Position des Bildausschnitts relativ zum Originalbild in Koordinatenform ausgeliefert werden. Wie in Kapitel 3 bereits erläutert, setzt sich diese zusammen aus einer x- und einer y-Koordinate sowie einer Breite und einer Höhe. Die Auswahl über die Ausgabeform erfolgt über einen URL-Parameter in der API-Anfrage. Wie in Kapitel 3 bereits erläutert, sieht die gewählte Systemarchitektur nicht vor, dass Bildzuschnitte als Bilddatei gespeichert werden. Vielmehr wird nur die Information darüber gespeichert, wo sich die Zuschnittskordinaten in Bezug auf das Originalbild befinden. Für die Einbindung der Web-Services wurde sich für die zweite beschriebene Möglichkeit entschieden, in der Bildzuschnitte in Koordinatenform ausgespielt werden. Allerdings wäre eine qualitative Beurteilung der generierten Zuschnitte nur auf Grundlage der Zuschnittskordinaten und ohne visuelle Unterstützung für Anwender:innen unzumutbar gewesen. Daher sieht der Figma Prototyp für diesen Anwendungsfall eine Lösung vor, in der die Bildausschnitte innerhalb der Webanwendung für den Bildzuschnitt zur Laufzeit generiert werden können. Mithilfe eines HTML Canvas Elements, dem Originalbild sowie den Zuschnittskordinaten wird eine Vorschau des generierten Zuschnitts erzeugt. Auf Basis dieser Vorschau ist es für Anwender:innen möglich, eine Entscheidung über die Qualität zu treffen. Gleichzeitig kann das Format der Zuschnittskordinaten auch direkt genutzt und über den DAM-Service im DAM-System abgespeichert werden.

.

¹<https://git.imp.fu-berlin.de/sonnik91/bachelor-thesis-project-smart-image-cropping>

8. Evaluation

8.1. Aufbau und Durchführung der User-Tests

Die Datengrundlage für die in diesem Kapitel beschriebene Evaluation wurde im Rahmen von praktisch durchgeführten Usertests erhoben. Um eine möglichst diverse Proband:innengruppe hinsichtlich Fachkompetenz und Erfahrungswerten zu schaffen, wurden für die Tests gezielt Personen aus den Bereichen UI- und UX-Design sowie aus den Bereichen Backend-Entwicklung und Contentdesign ausgewählt. Drei der Proband:innen waren sehr vertraut mit dem bestehenden Prozess des Bildzuschnitts sowie den möglichen Problemen und Herausforderungen (Heavy Users). Die anderen zwei Proband:innen hingegen kannten zwar den grundlegenden Prozess des Bildzuschnitts, führten diesen aber nicht auf einer alltäglichen Basis selbst durch (Non-Heavy-Users). Aufgrund der kleinen Proband:innengruppe sowie deren unterschiedlichen fachlichem Hintergrund, wurde sich bei der Durchführung der Tests für das „Within-Subjects-Design“ entschieden. Hierbei testen alle Proband:innen jede der vier Konfigurationen. Auf diese Weise können differierende Kompetenzen und Erfahrungen innerhalb der Proband:innengruppe besser abgefangen werden, was heißt, dass diese Merkmale weniger Gewichtung finden als beispielsweise bei einem „Between-Subjects-Design“.

Für die Usertests wurden zwanzig Bilder zur Bearbeitung ausgewählt. Diese wurden für die Durchführung der Tests in vier Untergruppen mit jeweils fünf Bildern untergliedert. Jede dieser Untergruppen ähnelte sich in der Zusammenstellung bezüglich Komposition und Inhalt. Jede Bildergruppe enthielt sowohl Bilder mit nur einem möglichen AoI sowie Bilder mit mehreren möglichen AoIs. Die Proband:innen hatten die Aufgabe, sukzessive mit jeweils einem der vier Prototypen die Bilder aus jeweils einer dieser Bilduntergruppen zuzuschneiden, sodass jede:r Proband:in im Gesamtverlauf alle vier Prototypen genutzt und insgesamt alle zwanzig Bilder zugeschnitten hatte. Trotz der Tatsache, dass die Bilduntergruppen möglichst analog gestaltet wurden, waren Unterschiede dennoch nicht auszuschließen. Um mögliche qualitative Abweichungen bei den algorithmisch generierten Bildzuschnitten zwischen den Untergruppen auszugleichen, sah das Testdesign für alle Proband:innen andere Kombinationen aus Prototyp und Bilduntergruppe vor.

Um im Nachgang die tatsächliche objektive Zuschnittszeit zu ermitteln, lief während der Tests parallel eine Bildschirmaufnahme. Zur Messung der subjektiv wahrgenommenen Arbeitsbelastung wurden die Proband:innen nach dem Test jedes Prototyps dazu aufgefordert, den Nasa-TLX-Fragebogen (siehe Anhang A.1) sowie den für diese Studie angepassten ResQue-Fragebogen (siehe Anhang A.4) auszufüllen.

8.2. Ergebnisse

8.2.1. Zuschnittszeit

Für jeden der durchgeführten 100 Tests (5 Proband:innen x 4 Konfigurationen x je 5 Bilder) wurde im Anschluss eine Analyse der Bildschirmaufnahmen durchgeführt, um die jeweiligen Zuschnittszeiten zu ermitteln. Da sich beim Zuschnitt der insgesamt sieben Formate pro Bild bei den Proband:innen ein potenzieller Lerneffekt einstellte (z.B. hinsichtlich der Lokalisation wesentlicher Bildinhalte), war der Vergleich der Zuschnittszeit pro Format nicht zielführend. Durch den potenziellen Lerneffekt wären die Ergebnisse verfälscht worden. Aus diesem Grund wurde nicht die Zuschnittszeit je Format als Messgröße herangezogen, sondern die Gesamtzeit für die Bearbeitung aller sieben Formate pro Bild. Zur Ermittlung einer durchschnittlichen Gesamtbearbeitungszeit pro Bild, sprich zur Erstellung der sieben Formate, wurde für jeden der Automatisierungsgrade (Manuell, P1, P2, P3) das arithmetische Mittel aus den jeweiligen Bearbeitungszeiten aller Bilder und aller Proband:innen gebildet. Die Ergebnisse sind in Tabelle 7 dargestellt.

Konfigurationen	Manuell	Prototyp P_1	Prototyp P_2	Prototyp P_3
LoA	1	2	3	5
Teilnehmende	5	5	5	5
Anzahl bearbeiteter Bilder	25	25	25	25
\emptyset Zeit pro Bild [s]	118	132	86	45

Tabelle 7: Evaluationsergebnisse der Zeiterfassung

Die Ergebnisse zeigen, dass die durchschnittlichen Zuschnittszeiten pro Bild in den höheren Automatisierungsgraden deutlich sinken. In der P_2 -Konfiguration zeigt sich im Vergleich zur Ausgangskonfiguration mit rein manueller Bearbeitung eine Verringerung der Bearbeitungszeit um rund 30 Prozent, in der P_3 -Konfiguration sogar um fast 62 Prozent. Lediglich die P_1 -Konfiguration bricht aus diesem Trend aus (siehe graue Markierung in Tabelle 7). Diese erforderte im Vergleich zur Ausgangskonfiguration eine um rund zehn Prozent höhere durchschnittliche Zuschnittszeit pro Bild.

8.2.2. Wahrgenommene Arbeitsbelastung

Die Ergebnisse des Nasa-Raw-TLX-Fragebogens geben Einblicke in die subjektiv erlebte Arbeitsbelastung der Proband:innen. Raw-TLX-Kennwerte bewegen sich im Intervall zwischen 0 und 100. Ein Raw-TLX-Kennwert von 0 entspricht der kleinstmöglichen wahrgenommenen Arbeitsbelastung, während der Wert 100 der größtmöglichen wahrgenommenen Arbeitsbelastung entspricht. Auch hier wurde aus den Ergebnissen der jeweiligen Proband:innen das arithmetische Mittel gebildet. Tabelle 8 zeigt die Ergebnisse dieser

Befragung. Die grau hinterlegten Tabellenfelder kennzeichnen den jeweils schlechtesten Wert in der befragten Dimension.

Konfigurationen	Manuell	Prototyp P_1	Prototyp P_2	Prototyp P_3
LoA	1	2	3	5
Teilnehmende	5	5	5	5
Geistige Beanspruchung	51	59	30	22
Körperliche Beanspruchung	49	48	27	16
Erlebter Zeitdruck	30	39	20	9
Leistung	24	30	28	21
Grad der Anstrengung	38	54	25	14
Grad der Frustration	34	61	23	11
Nasa RAW-TLX	38	49	26	16

Tabelle 8: Evaluationsergebnisse des Nasa Raw TLX Fragebogens

Die Ergebnisse zeigen in den höheren Automatisierungsgraden LoA3 (P_2) und LoA5 (P_3) eine von den Proband:innen deutlich geringer erlebte Arbeitsbelastung. Die erlebte Arbeitsbelastung, die sich im Gesamtwert des Nasa RAW-TLX abbildet, sinkt bei P_2 im Vergleich zum vollständig manuellen Zuschnitt um rund 32 Prozent, bei P_3 zeigt sich sogar eine um fast 60 Prozent geringere Arbeitsbelastung. Auch hier ist es erneut die LoA2-Konfiguration (P_1), die einen entgegengesetzten Trend zeigt. Die bei P_1 subjektiv eingeschätzte Arbeitsbelastung ist rund 30 Prozent höher als beim rein manuellen Zuschnitt. Vergleicht man die Einzelergebnisse, so erzielt die P_1 -Konfiguration in nahezu allen Dimensionen die schlechtesten Werte. Lediglich in der Dimension „Körperliche Beanspruchung“ ist der Wert des manuellen Zuschnitts geringfügig größer. Am wenigsten belastend empfanden die Proband:innen die P_3 -Konfiguration mit dem höchsten Automatisierungsgrad (LoA5). Sie erhielt in allen Dimension die besten Bewertungen hinsichtlich der mit dem Raw-TLX gemessenen Arbeitsbelastung

8.2.3. Eingesetzter Aufwand zur Finalisierung von Bildzuschnitten

Um neben menschlichen auch algorithmische Perspektiven einzubeziehen, wurde, wie in Abschnitt 5.3 bereits erläutert, der Aufwand ermittelt, der notwendig ist, um die algorithmisch generierten Zuschnitte in eine Form zu bringen, die dem Qualitätsverständnis der Proband:innen entspricht. Dazu wurde die relative Zuschnittverschiebung sowie die relative Zuschnittsvergrößerung beziehungsweise -verkleinerung ermittelt. Diese beide Kennzahlen messen die von den Proband:innen vorgenommenen Korrekturen an den algorithmisch generierten Bildzuschnitten. Sie werden in Prozent ausgedrückt; die Bezugsgröße ist dabei die maximal mögliche Veränderung.

Zur besseren Vergleichbarkeit der Prototypen wurden die Ergebnisse der beiden Kennzahlen wiederum in drei Kategorien zusammengefasst. In die erste Kategorie „keine Änderungen“ fallen alle Ergebnisse, bei denen die algorithmisch generierten Zuschnitte ohne Anpassung direkt übernommen wurden. Ihre Qualität war so hoch, dass die Proband:innen keinen Interventionsbedarf sahen. Sie entsprechen einer Zuschnittsveränderung von 0%. Eine zweite Kategorie mit dem Label „kleine Änderungen“ umfasst alle Ergebnisse mit einer Veränderung größer 0% und kleiner oder gleich 10%. Die dritte Kategorie „große Änderungen“ umfasst alle Werte größer 10%. Die Kennzahlen Zuschnittsverschiebung und Zuschnittsvergrößerung bzw. -verkleinerung konnten nur für die Prototypen mit einem Automatisierungsgrad von mindestens LoA2 erhoben werden (P1, P2, P3). Für die Ausgangskonfiguration mit manuellem Zuschnitt spielt sie keine Rolle. Für jeden Prototypen wurde im Anschluss berechnet, wie groß der prozentuale Anteil der ermittelten Werte an der Gesamtzahl der Bearbeitungen in jeder der drei Änderungskategorien ist.

Tabelle 9 zeigt die Ergebnisse hinsichtlich vorgenommener Zuschnittsverschiebungen des Bildmittelpunktes durch die Proband:innen. Konkret wurde gemessen, wie stark der Bildmittelpunkt bei jedem einzelnen Bildzuschnitt versetzt wurde („keine Änderungen“, „kleine Änderungen“, „große Änderungen“). Die Prozentzahlen geben, an, wie groß der Anteil der Ergebnisse an der Gesamtzahl der Bearbeitungen hinsichtlich des Umfangs der Verschiebung in den jeweiligen Änderungskategorien ist.

Konfigurationen	Prototyp P_1	Prototyp P_2	Prototyp P_3
LoA	2	3	5
Teilnehmende	5	5	5
keine Änderungen	3%	19%	30%
kleine Änderungen	56%	58%	48%
große Änderungen	41%	23%	24%

Tabelle 9: Eingesetzter Aufwand zur Finalisierung von Bildzuschnitten - Vergleich der Prototypen hinsichtlich der Verschiebung des Bildmittelpunktes

Wie man dort insbesondere in der Kategorie „keine Änderungen“ sehen kann, sinkt der Anpassungsaufwand (Aufwand für Verschiebungen) mit einem steigenden Automatisierungsgrad kontinuierlich ab (von 3% in P_1 über 19% in P_2 zu 30% in P_3). Die häufigsten „großen Änderungen“ ergeben sich mit 41% beim Prototyp P_1 mit LoA2; in den Konfigurationen P_2 (LoA3) und P_3 (LoA5) reduzierte sich die Häufigkeit dieser Kategorie um mehr als 40%. Dies deutet darauf hin, dass durch die automatisierte Wahl des AoI in P_2 und P_3 offensichtlich eine deutliche Reduktion des Aufwands für größere manuelle Anpassungen hinsichtlich des Bildmittelpunktes verbunden war. Bei schlechter algorithmenbasierter Wahl des AoI hätte man ausgeprägte Verschiebungen der Bildmitte erwartet, was sich in den vorliegenden Ergebnissen aber umgekehrt abbildet. Auch

der bereits beschriebene Trend in der ersten Kategorie „keine Änderungen“ geht in diese Richtung. Innerhalb der Kategorie „kleine Änderungen“ zeigt sich am deutlichsten eine Tendenz zur Aufwandsreduktion in Prototyp P_3 . Diese fällt zwar geringer aus als in P_1 und P_2 , ist jedoch noch immer auffällig hoch. Letztlich weisen auch die Ergebnisse der zweiten Kategorie „kleine Änderungen“ auf die Konfiguration P_3 als diejenige hin, die hinsichtlich der Verschiebung des Bildmittelpunktes den geringsten zusätzlichen Anpassungsaufwand benötigte. Damit kann die P_3 -Konfiguration mit dem Automatisierungsgrad LoA5 betreffend der Kennzahl „Zuschnittsverschiebung“ als die zielführendste bewertet werden.

Die zweite Messgröße erfasst die relative Zuschnittsvergrößerung beziehungsweise -verkleinerung. Mithilfe dieser Kennzahl wurde ermittelt, wie stark das Zuschnittsfenster bei jedem einzelnen Bildzuschnitt angepasst wurde („keine Änderungen“, „kleine Änderungen“, „große Änderungen“). Die Prozentzahlen in Tabelle 10 geben an, wie groß der Anteil der Ergebnisse in den jeweiligen Änderungskategorien ist.

Prototypen	Prototyp P_1	Prototyp P_2	Prototyp P_3
LoA	2	3	5
Teilnehmende	5	5	5
keine Änderung	11%	24%	32%
kleine Änderung	52%	50%	35%
große Änderung	36%	27%	24%

Tabelle 10: Eingesetzter Aufwand zur Finalisierung von Bildzuschnitten - Vergleich der Prototypen hinsichtlich der Zuschnittsvergrößerung bzw. -verkleinerung

Auch für dieses Kennwertes verbessern sich die Ergebnisse mit steigendem Automatisierungsgrad. In der Kategorie „keine Änderung“ zeigt sich von P_1 über P_2 bis P_3 eine kontinuierliche Reduktion des Aufwands mit zunehmender Automatisierung (von 11% in P_1 über 24% in P_2 auf 32% in P_3). Betrachtet man die Kategorie „große Änderungen“ zeigt sich im direkten Vergleich von P_1 mit sowohl P_2 als auch P_3 in den beiden höheren Automatisierungsstufen eine Reduktion des Anteils der Zuschnitte, in denen größere Änderungen notwendig waren. Die Häufigkeit von „kleinen Änderungen“ reduzierte sich erst zwischen P_2 und P_3 in auffälligerem Maße. Diese Ergebnisse ähneln stark dem Veränderungsmuster der vorausgehend besprochenen Metrik „Zuschnittsverschiebung“. Auch hinsichtlich der hier betrachteten Kennzahl der „Zuschnittsverkleinerung bzw. -vergrößerung“ erreicht die P_3 -Konfiguration mit dem Automatisierungsgrad LoA5 die besten Ergebnisse.

8.2.4. ResQue Fragebogen

Der ResQue-Fragebogen zielt darauf ab, die integrative, menschlich-algorithmische Perspektive zu erfassen. Die Proband:innen wurden aufgefordert, nach dem Test jedes Prototyps jeweils zwölf Fragen zu beantworten. Diese Fragen sollen die Erfahrungen der Proband:innen im Umgang mit den entwickelten Prototypen erfassen. Wie in Kapitel 5.3 beschrieben, ermöglicht das Rahmenwerk zum ResQue-Fragebogen [33] individuelle Anpassungen je nach Verwendungszweck. Für das vorliegende Automatisierungsproblem wurden die Dimensionen „Novelty“, „Control“ und „Purchase Intention“ des ResQue-Fragebogens und die damit verbundenen Fragen gestrichen. Die Dimensionen „Novelty“ und „Purchase Intention“ erschienen unpassend, da es sich in dem Untersuchungskontext um keine klassische Empfehlungsfunktion wie beispielsweise auf Onlinehändler-Webseiten handelt. Die Dimension „Control“ konzentriert sich auf ein anpassbares Geschmacksprofil für die generierten Vorschläge. So müssten sich die generierten Vorschläge dem subjektiven Ästhetikempfinden der Anwender:innen anpassen und die Vorschläge basierend auf den Entscheidungen der Anwender:innen generiert werden. Da es sich bei den genutzten Zuschnitt-Services nicht um Reinforcement-Learning Technologien handelt, erschien auch diese Dimension unpassend. Die verbliebenen Dimensionen sind in der Tabelle 11 aufgeführt. Jede Dimension wird dabei durch eine Frage repräsentiert (siehe Anhang A.4). Die Beantwortung der Fragen erfolgte in einer fünfstufigen Likert-Skala von -2 („strongly disagree“) bis 2 („strongly agree“). Auch hier konnte die Befragung nur zu den Konfigurationen erfolgen, die einen Automatisierungsgrad von mindestens LoA2 aufweisen. Deshalb wird die Ausgangskonfiguration in den Ergebnissen nicht aufgeführt. Zur ganzheitlichen Betrachtung wurde aus den Ergebnissen aller Proband:innen das arithmetische Mittel gebildet. Tabelle 11 zeigt diese Durchschnittswerte für alle Prototypen und Dimensionen. Die grau hinterlegten Felder markieren jeweils den besten Wert der entsprechenden Dimension.

Prototypen	Prototyp P_1	Prototyp P_2	Prototyp P_3
LoA	2	3	5
Teilnehmende	5	5	5
Q_1 Recommendation Accuracy	-0,4	0,8	0,8
Q_2 Diversity	-0,4	0,6	0
Q_3 Interface Adequacy	-0,6	0,2	0,6
Q_4 Explanations	-1,6	-1,0	-1,6
Q_5 Information	0,2	1,2	0,6
Q_6 Interaction	-0,4	0,8	-0,2
Q_7 Perceived Ease of Use	0,4	1,0	1,6
Q_8 Transparency	-0,8	-0,2	0,4
Q_9 Usefulness	-0,4	0,4	1,4
Q_{10} Satisfaction	-0,4	1,2	1,6
Q_{11} Confidence and Trust	-1	-0,2	-0,2
Q_{12} Use Intention	-0,4	0,8	1,4
Gesamtbewertung	-0,5	0,5	0,5

Tabelle 11: Ergebnisse des ResQue-Fragebogens

In allen befragten Dimensionen erzielt die P_1 -Konfiguration die schlechtesten Ergebnisse. Dies spiegelt sich auch in der durchschnittlichen Gesamtbewertung wider (-0,5). Sowohl die P_2 - als auch die P_3 -Konfiguration erzielen in der Gesamtbewertung das gleiche Resultat (0,5), während sich die Einzelergebnisse zum Teil deutlich unterscheiden. So erzielten beide Konfigurationen eine gleichermaßen gut bewertete Genauigkeit bei den generierten Zuschnittsvorschlägen („Recommendation Accuracy“), während in den Dimensionen „Diversity“, „Information“ und „Interaction“ vor allem die P_2 -Konfiguration die besseren Ergebnisse erzielte. Die P_3 -Konfiguration erzielte dagegen deutlich bessere Ergebnisse in den Bereichen „Perceived Ease of Use“, „Usefulness“ und „Use Intention“. Abschließend lässt sich aus den Ergebnissen des ResQue-Fragebogens kein eindeutiger Favorit identifizieren. Sowohl Prototyp P_2 als auch der Prototyp P_3 erzielten im Vergleich zu Prototyp P_1 die besseren Ergebnisse. Die beiden besser bewerteten Konfigurationen unterscheiden sich dennoch stark in ihren jeweiligen Stärken und Schwächen. P_2 punktet besonders hinsichtlich Funktionalität und Qualität („Diversity“, „Information“, „Interaction“), während die P_3 -Konfiguration durch eine höhere Benutzer:innenfreundlichkeit bzw. Nutzungsbereitschaft überzeugt („Perceived Ease of Use“, „Usefulness“, „Use Intention“, „Satisfaction“).

8.3. Diskussion der Ergebnisse

Der in Abschnitt 5.1 identifizierte „point of friction“ beschreibt zwei gegenläufige Zielsetzungen. Einerseits zielt die im Zuge dieser Arbeit angestrebte Automatisierung auf eine

Verringerung der Zuschnittszeit und eine Schonung von Personalressourcen ab, andererseits bestehen hohen Anforderungen an die Qualität der Bildzuschnitte. Aus diesem Spannungsfeld leiten sich zwei Erwartungen ab: Die erforderliche Zuschnittszeit sollte mit steigendem Automatisierungsgrad sinken, übernimmt das intelligente System doch immer mehr Teilaufgaben innerhalb des Prozesses, je höher der Automatisierungsgrad ist. Analog dazu sinkt auch die Qualität der generierten Bildzuschnitte, da sich ein menschlicher Beitrag grundsätzlich als qualitätssteigernd auswirkt. Die zuvor beschriebenen Ergebnisse werden nachfolgend gemäß dieser beiden Erwartungen eingeordnet und bewertet.

Die erzielten durchschnittlichen Zuschnittszeiten folgen im Wesentlichen der Annahme, dass ein steigender Automatisierungsgrad eine Beschleunigung des Prozesses bewirkt. Wie in Kapitel 8.2.1 bereits festgestellt, erzielte die LoA1-Konfiguration des Prototypen P_1 hinsichtlich der Zuschnittszeiten die schlechtesten Ergebnisse im Test. Selbst die Ausgangskonfiguration mit einem vollständig manuellen Zuschnitt erzielte im Durchschnitt eine schnellere Zuschnittszeit. Nun basieren Ausgangs- und P_1 -Konfiguration auf derselben Aufgabenzuweisung; in beiden Konfigurationen führt der Mensch weiterhin alle Teilschritte selbst manuell durch. Beide Konfigurationen unterscheiden sich im Wesentlichen durch das Interaktionsdesign und die Tatsache, dass die P_1 -Konfiguration bereits Zuschnittsvorschläge auf sehr niedrigem Automatisierungsniveau algorithmisch generiert. Eine mögliche Erklärung für die Verschlechterung der Ergebnisse liegt im Interaktionsdesign. Während in der Ausgangskonfiguration die Wahl des AoI und des Bildausschnitts vollständig kognitiv seitens der Anwender:innen stattfindet, fordert die P_1 -Konfiguration Anwender:innen dazu auf, den AoI manuell im Zuschnittsfenster auszuwählen und einen Skalierungsfaktor über das zur Verfügung stehende Slider-Element festzulegen. Es sind also zusätzliche manuelle Interaktionen erforderlich, die bei der Ausgangskonfiguration nicht notwendig sind, sondern intuitiv und ohne zusätzliche Bedienflächen umgesetzt werden. Hinzu kommt die Wartezeit, die dadurch entsteht, dass die manuell eingegebenen Parameter an eine API übergeben werden müssen und der algorithmisch generierte Zuschnitt erst mit einer Verzögerung zurückgeliefert wird. Aufgrund dieser beiden nachteiligen Effekte ist die LoA2-Konfiguration des Prototyps P_1 als ungeeignete Automatisierungslösung zu bewerten. Diese Einschätzung wird gestützt von den Ergebnissen bezüglich der wahrgenommenen Arbeitsbelastung. Die mithilfe des standardisierten Nasa-Raw-TLX-Fragebogens erfassten Ergebnisse verzeichnen hier für Prototyp 1 ebenfalls die schlechtesten Werte. Speziell das um rund 80 Prozent höhere Frustrationslevel („Frustration Level“), sowie der um 40 Prozent höhere Grad der erlebten Anstrengung („Effort“) unterstreichen die zuvor formulierten Vermutungen. Letztlich konnte die als sehr hoch empfundene Arbeitsbelastung auch durch die Ergebnisse der Befragung mit dem ResQue-Fragebogen bestätigt werden. Hier erzielte die LoA2-Konfiguration in allen untersuchten Dimensionen die schlechtesten Werte.

Des Weiteren wurde erwartet, dass sich der menschliche Beitrag insbesondere in niedrigeren Automatisierungsgraden qualitätssteigernd auswirkt. Die Ergebnisse zeigten jedoch, dass der erforderliche Aufwand für die Herstellung der Zuschnittsqualität mit steigen-

dem Automatisierungsgrad sinkt. Nachfolgend werden mögliche Ursachen für diese abweichenden Ergebnisse für jeden der drei Prototypen einzeln diskutiert.

Eine Ursache für diese widersprüchlichen Qualitätsergebnisse der LoA2-Konfiguration des Prototyps P_1 könnte in der Art und Weise liegen, wie der AoI ausgewählt wird. Diese wird für jedes Bild initial nur einmal festgelegt. Daraufhin werden basierend auf dem ausgewählten Bereich alle sieben Format-Zuschnitte generiert. Wurde ein verhältnismäßig großer Bereich markiert, so wurde für alle algorithmusgenerierten Formate versucht, diesen Bereich im Zuschnitt zu berücksichtigen. Nun variieren die vordefinierten Formate von hochformatigen bis hin zu sehr breitformatigen Varianten. Die Vermutung liegt nahe, dass der ausgewählte AoI nicht immer optimal für jedes Format passte und so zu zusätzlichem Nachbearbeitungsaufwand führte. Weiterhin zeigten die Analysen der Bildschirmaufnahmen, dass überhaupt nur wenige Proband:innen die Möglichkeit des Slider-Elements überhaupt nutzten, um den Skalierungsfaktor manuell zu ändern. Viele Proband:innen veränderten den initial generierten Vorschlag mit einem Skalierungsfaktor von 0,5 nicht und nahmen auf dieser Basis weitere Anpassungen des Zuschnittfensters manuell vor, was wiederum den Nachbearbeitungsaufwand erhöhte. Die Anwendung wurde zum Teil also nicht so genutzt, wie sie ursprünglich gedacht war und programmtechnisch gestaltet wurde. Möglicherweise resultierte das aus fehlender Vertrautheit mit dem angebotenen Tool, sodass viele Anwender:innen ihren Lösungsweg in bekannten Handlungsschemata suchten. Diese Fehlnutzung könnte zu einer Verschlechterung der Ergebnisse beigetragen haben. Im Nachhinein lässt sich aus diesen Beobachtungen die Erkenntnis ziehen, dass zukünftig vor einer Überprüfung und Bewertung von Automatisierungsvarianten ausreichende Schulungs- und Trainingsphasen für die Nutzung der neuen Werkzeuge essenziell sind.

Im Folgenden soll der verringerte Aufwand zur Herstellung der Zuschnittsqualität von der LoA3 zu der LoA5-Konfiguration der Prototypen P_2 und P_3 kritisch betrachtet werden. Eine naheliegende Schlussfolgerung hierfür wäre, dass die Automatisierung beim Erreichen einer zufriedenstellenden Zuschnittqualität sowohl die Bearbeitungszeit als auch den objektiv erfassten und subjektiv erlebten Aufwand reduzieren konnte. Dieses Ergebnis könnte aber auch auf ein gestiegenes Vertrauen in die Qualität algorithmisch generierter Zuschnitte zurückgeführt werden. Dies förderte folglich eine Verhaltenstendenz des Zögerns hinsichtlich eines Veränderungseingriffes. Das gewählte Interaktionsdesign könnte dabei die Problematik zusätzlich verschärft haben. So präsentiert die LoA5-Konfiguration des Prototyps P_3 nach der Bildauswahl lediglich eine Übersicht aller generierten Format-Zuschnitte zu allen ausgewählten Bildern (siehe Kapitel 6.3). Der Computer legte also bereits fest, welcher Zuschnitt für welches Format vorausgewählt wird. Anwender:innen wurden somit nicht weiter dazu aufgefordert, jeden Bildzuschnitt einzeln auszuwählen und zu bestätigen. Zuschnitte, an denen zuvor möglicherweise noch kleine manuelle Anpassungen vorgenommen worden wären, wurden nun bereits in der Übersichtsansicht als gut befunden. Andererseits wurden schlechte Zuschnitte in der Übersichtsansicht weiterhin erkannt und anschließend manuell ausgebessert. Dieser Effekt würde vor allem auch die Verbesserungen in den Kategorien „keine Änderungen“ und „kleine Änderungen“ erklären.

8.4. Ableitung der Sweet-Spot-Konfiguration

Im Zuge der Analyse in Kapitel 5 wurden für den untersuchten Prozess des Bildzuschnitts zwei gegensätzliche Zielsetzungen identifiziert. Einerseits sollte Anwender:innen durch die Integration intelligenter Automatisierungselemente eine Möglichkeit geboten werden, ihre Aufgaben im Bereich des Bildzuschnitts schneller und komfortabler durchzuführen. Andererseits sollte die Qualität der Bildzuschnitte erhalten bleiben hinsichtlich einer inhaltlich und grafisch korrekten und zum Kontext passenden Auswahl des AoI. Die Zielsetzung dieser Arbeit war es, einen angemessenen Automatisierungsgrad zu identifizieren, der diese scheinbar gegensätzlichen Anforderungen bestmöglich vereint.

In nahezu allen erfassten Leistungsmetriken erzielte die LoA5-Konfiguration die besten Ergebnisse. Lediglich die Ergebnisse aus der Befragung mit dem ResQue-Fragebogen waren weniger eindeutig. Hier erzielten sowohl die LoA3- als auch die LoA5-Konfiguration dieselbe Gesamtbewertung. Mit Blick auf die zuvor formulierte Zielsetzung eines komfortableren und ressourcenschonenderen Prozesses kommen den Dimensionen „Perceived Ease of Use“, „Usefulness“, „Use Intention“ und „Satisfaction“ eine hohe Bedeutung zu. Die positiven Ergebnisse in diesen Dimensionen des ResQue-Fragebogens deuten darauf hin, dass die Proband:innen im Umgang mit der LoA5-Konfiguration zielsetzungskonform am zufriedensten waren. Somit lässt sich abschließend die LoA5-Konfiguration des Prototypen P3 als Sweet-Spot-Konfiguration für das untersuchte Automatisierungsproblem bestimmen.

Für eine finale Implementierung sollten die aufgeführten Verbesserungspotenziale dieser Arbeit berücksichtigt und bei der Umsetzung einbezogen werden. Vor allem die Ergebnisse des ResQue-Fragebogens zeigen offene Potenziale vor allem in der Gestaltung des Prototyps.

8.5. Sweet-Spot-Konfiguration und der Aspekt der Dequalifizierung

Wie in Kapitel 2 bereits erläutert, ist es wichtig, bei der Automatisierung und der Unterstützung von bestehenden komplexen Prozessen durch intelligente Systeme stets mögliche Dequalifizierungseffekte und erforderlichen Qualifizierungsbedarf zu berücksichtigen. Dadurch, dass Menschen die automatisierten Tätigkeiten nicht weiter routinemäßig ausführen, verlieren sie schrittweise damit verbundenen Fähigkeiten. So stellen auch in der untersuchten Problemstellung dequalifizierende Effekte ein reales Risiko dar. Menschen lernen oft schon früh durch ihre schulische Ausbildung über Bildbeschreibungen, Bildanalysen und Bildinterpretationen die Fähigkeit, Bilder und Bildinhalte zu lesen und ihren Fokus zu identifizieren. Nun erfordert der manuelle Zuschnitt von Bildmedien diese Fähigkeiten und trainiert sie routinemäßig in spezieller Weise. Die erarbeitete Aufgabenzuweisung hingegen implementiert einen stark automatisierten Prozess, innerhalb dessen diese Zuschnittsfähigkeiten auf den ersten Blick weniger intensiv genutzt werden.

Die als Ergebnis dieser Arbeit vorgeschlagene Konfiguration verlangt jedoch eher eine Bewahrung von Fähigkeiten, anstatt diese zu verschieben oder gänzlich zu substituieren. Die algorithmisch generierten Zuschnitte müssen weiterhin von Menschen geprüft und validiert werden. Um Zuschnitte zu kontrollieren, sind eben jene Analyse- und Interpretationsfähigkeiten weiterhin notwendig. Es ändert sich lediglich der Rahmen, in dem sie angewandt werden. Ein weiterer Effekt, der dem Deskillung entgegenwirkt, ist die weiter bestehende schulische Ausbildung, die den Erwerb grundlegender Fähigkeiten sicherstellt. Diese Tatsache entspricht einem der drei von Rafner et al. [34] definierten Ansätze, um Effekte des Deskillung zu bekämpfen. Im Rahmen der erarbeiteten Lösung scheinen Auswirkungen im Sinne eines Deskillung sehr begrenzt zu sein, dennoch sollte dieser Aspekt in weiteren Ausbaustufen immer wieder neu miteinbezogen werden.

9. Fazit

In der vorliegenden Arbeit wurde eine Aufgabenzuweisung für den Zuschnitt von Bildmedien erarbeitet, welche eine kollaborierende Arbeitsweise zwischen Mensch und intelligentem System ermöglicht. Mithilfe des siebenstufigen Prozesses nach Mackeprang [29] wurde der Gestaltungsraum für die Entwicklung des interaktiven, intelligenten Systems systematisch untersucht. Die zentrale Herausforderung dabei war, eine Balance zwischen zwei sehr gegensätzlichen Zielsetzungen zu schaffen. So sollten Nutzer:innen einerseits entlastet werden und der Bildzuschnitt sollte möglichst wenig ihrer Arbeitszeit binden, andererseits sollte die gestaltete Aufgabenzuweisung den Anforderungen an die Zuschnittsqualität standhalten. Im Rahmen der Untersuchungen stellte sich die Aufgabenzuweisung der Level-5-Konfiguration nach Sheridan [35] als die leistungsfähigste heraus. Das interaktive, intelligente System übernahm dabei bereits zentrale Teilaufgaben beim Zuschnitt von Bildmedien und verschob die Rolle der zuschneidenden Person in eine beobachtende und kontrollierende Richtung. Die Aufgabenzuweisung ermöglichte dabei eine Beschleunigung des Prozesses um rund 60 Prozent und garantierte dennoch beste Qualitätsergebnisse im Rahmen der untersuchten Automatisierungsgrade. Die empfundene Arbeitsbelastung der befragten Anwender:innen sank dabei ebenfalls um rund 60 Prozent.

Das Ergebnis dieser Arbeit zeigt, dass durch die Kombination menschlicher und algorithmischer Fähigkeiten eine Steigerung der Leistungsfähigkeit beim Lösen komplexer Probleme möglich ist. Dies gilt insbesondere für Probleme, die – zumindest zur heutigen Zeit – noch nicht vollständig von intelligenten Systemen übernommen werden können. Die Arbeit verdeutlicht darüber hinaus, dass ein zielgerichtetes methodisches Vorgehen erforderlich ist, um die optimale Verbindung menschlicher und algorithmischer Fähigkeiten im Sinne hybrider Intelligenz zu gestalten und die richtige Auswahl aus der Vielfalt möglicher Kombinationen zu treffen. Gleichzeitig demonstriert sie beispielhaft die erfolgreiche Nutzung einer bereits existierenden strukturierten Methodologie für die anwendungspraktische Implementierung von KI-gestützter Automatisierung [29].

Im Hinblick auf die aktuellen Entwicklungen von KI, die bis dato in Gänze nicht überschaubar und greifbar zu sein scheinen, bleibt das Zusammenspiel von Mensch und intelligenten Systemen auch zukünftig ein sehr relevantes Spannungsfeld. Die Anforderungen an interaktive, intelligente Systeme werden sich mit den technologischen Innovationen weiter anpassen müssen. Dabei ist es essenziell, die menschliche Perspektive bei der Gestaltung dieser Systeme miteinzubeziehen, um negativen Effekten wie beispielsweise dem Deskillung vorzubeugen.

A. Anhang

A.1. Fragen des NASA TLX Fragebogens

Rating Scale Definitions

Title	Descriptions	
MENTAL DEMAND	How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?	<p>Place a mark at the desired point on each scale:</p> <p>MENTAL DEMAND</p> <p>Low High</p>
PHYSICAL DEMAND	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?	<p>PHYSICAL DEMAND</p> <p>Low High</p>
TEMPORAL DEMAND	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?	<p>TEMPORAL DEMAND</p> <p>Low High</p>
PERFORMANCE	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?	<p>PERFORMANCE</p> <p>Good Poor</p>
EFFORT	How hard did you have to work (mentally and physically) to accomplish your level of performance?	<p>EFFORT</p> <p>Low High</p>
FRUSTRATION LEVEL	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?	<p>FRUSTRATION</p> <p>Low High</p>

A.2. Fragen des angepassten ResQue Fragebogens

Dimension	Question	strongly disagree	disagree	neither agree nor disagree	agree	strongly agree
1	Recommendation Accuracy	The items recommended to me match my expectations.				
2	Diversity	The items recommended to me are diverse.				
3	Interface Adequacy	The layout and labels of the recommender interface are adequate.				
4	Explanation	The recommender explains why the products are recommended to me.				
5	Information	The information provided for the recommended items is sufficient for me to make a selection.				
6	Interaction	The recommender allows me to tell what I like/dislike				
7	Perceived Ease of Use	I became familiar with the recommender system very quickly.				
8	Transparency	I understood why the items were recommended to me.				
9	Usefulness	The recommender helped me find the ideal item.				
10	Satisfaction	Overall, I am satisfied with the recommender.				
11	Confidence & Trust	The recommender can be trusted.				
12	Use Intentions	I will use this recommender frequently.				

A.3. Ergebnisse der Usertests

<https://box.fu-berlin.de/s/J96pFBAMji5rZQ6>

A.4. GIT Repository

<https://git.imp.fu-berlin.de/sonnik91/bachelor-thesis-project-smart-image-cropping>

Literatur

- [1] Cutting and Bordering – ImageMagick Examples.
- [2] Intelligent zugeschnittene Miniaturbilder – maschinelles Sehen - Azure Cognitive Services | Microsoft Learn.
- [3] NASA Task Load Index | Digital Healthcare Research.
- [4] R. Achanta, F. Estrada, P. Wils, and S. Süsstrunk. Salient region detection and segmentation. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5008 LNCS, 2008.
- [5] Z. Akata, D. Balliet, M. De Rijke, F. Dignum, V. Dignum, G. Eiben, A. Fokkens, D. Grossi, K. Hindriks, H. Hoos, H. Hung, C. Jonker, C. Monz, M. Neerincx, F. Oliehoek, H. Prakken, S. Schlobach, L. Van Der Gaag, F. Van Harmelen, H. Van Hoof, B. Van Riemsdijk, A. Van Wynsberghe, R. Verbrugge, B. Verheij, P. Vossen, and M. Welling. A Research Agenda for Hybrid Intelligence: Augmenting Human Intellect with Collaborative, Adaptive, Responsible, and Explainable Artificial Intelligence. *Computer*, 53(8), 2020.
- [6] M. Anthimopoulos, S. Christodoulidis, L. Ebner, A. Christe, and S. Mougiakakou. Lung Pattern Classification for Interstitial Lung Diseases Using a Deep Convolutional Neural Network. *IEEE Transactions on Medical Imaging*, 35(5), 2016.
- [7] C. Babbage. On the economy of machinery and manufactures . *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 1(3), 1832.
- [8] A. Braga and R. K. Logan. The emperor of strong AI has no clothes: Limits to artificial intelligence. *Information (Switzerland)*, 8(4), 2017.
- [9] H. Braverman. Labor and Monopoly Capital. *Monthly Review*, 26(3), 1974.
- [10] F. Brugger and C. Gehrke. Skilling and deskilling: technological change in classical economic theory and its empirical evidence. *Theory and Society*, 47(5), 2018.
- [11] C. J. Cai, S. Winter, D. Steiner, L. Wilcox, and M. Terry. “Hello Ai”: Uncovering the onboarding needs of medical practitioners for human–AI collaborative decision-making, 2019.
- [12] A. Christe, A. A. Peters, D. Drakopoulos, J. T. Heverhagen, T. Geiser, T. Stathopoulou, S. Christodoulidis, M. Anthimopoulos, S. G. Mougiakakou, and L. Ebner. Computer-Aided Diagnosis of Pulmonary Fibrosis Using Deep Learning and CT Images. *Investigative Radiology*, 54(10), 2019.

- [13] L. Colligan, H. W. Potts, C. T. Finn, and R. A. Sinkin. Cognitive workload changes for nurses transitioning from a legacy system with paper documentation to a commercial electronic health record. *International Journal of Medical Informatics*, 84(7), 2015.
- [14] S. W. A. Dekker and D. D. Woods. MABA-MABA or Abracadabra? Progress on Human-Automation Co-ordination. *Cognition, Technology & Work*, 4(4), 2002.
- [15] D. Dellermann, P. Ebel, M. Söllner, and J. M. Leimeister. Hybrid Intelligence. *Business and Information Systems Engineering*, 61(5), 2019.
- [16] M. Dikmen and C. Burns. Trust in autonomous vehicles: The case of tesla autopilot and summon. In *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017*, volume 2017-January, 2017.
- [17] Erik Brynjolfsson and Andrew McAfee. *Race Against the Machine*. Digital Frontier Press, Lexington, 2011.
- [18] T. Ferris, N. Sarter, and C. D. Wickens. Chapter 15 – Cockpit Automation: Still Struggling to Catch Up. . . . In *Human Factors in Aviation*. 2010.
- [19] J. R. Glissen Brown, N. M. Mansour, P. Wang, M. A. Chuchuca, S. B. Minchenberg, M. Chandnani, L. Liu, S. A. Gross, N. Sengupta, and T. M. Berzin. Deep Learning Computer-aided Polyp Detection Reduces Adenoma Miss Rate: A United States Multi-center Randomized Tandem Colonoscopy Study (CADET-CS Trial). *Clinical Gastroenterology and Hepatology*, 20(7), 2022.
- [20] K. Goddard, A. Roudsari, and J. Wyatt. Automation bias - A hidden issue for clinical decision support system use. In *Studies in Health Technology and Informatics*, volume 164, 2011.
- [21] K. Goddard, A. Roudsari, and J. C. Wyatt. Automation bias: Empirical results assessing influencing factors. *International Journal of Medical Informatics*, 83(5), 2014.
- [22] R. R. Hoffman, T. M. Cullen, and J. K. Hawley. The myths and costs of autonomous weapon systems. *Bulletin of the Atomic Scientists*, 72(4), 2016.
- [23] J. Johnson. Die Vermischung von Menschen und Nicht-Menschen: Die Soziologie eines Türschließers. In *ANThology. Eine einführendes Handbuch zur Akteur-Netzwerk-Theorie.*, number 1977. 2006.
- [24] Karl Marx. *Lohnarbeit und Kapital / Lohn, Preis und Profit*. Das Freie Buch, 2002.
- [25] G. B. Kim, K. H. Jung, Y. Lee, H. J. Kim, N. Kim, S. Jun, J. B. Seo, and D. A. Lynch. Comparison of Shallow and Deep Learning Methods on Classifying the Regional Pattern of Diffuse Lung Disease. *Journal of Digital Imaging*, 31(4), 2018.
- [26] D. Lazer, R. Kennedy, G. King, and A. Vespignani. The parable of google flu: Traps in big data analysis, 2014.

- [27] D. Li, H. Wu, J. Zhang, and K. Huang. A2-RL: Aesthetics Aware Reinforcement Learning for Image Cropping. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018.
- [28] Y. Li, R. Kumar, W. S. Lasecki, and O. Hilliges. Artificial intelligence for HCI: A modern approach. In *Conference on Human Factors in Computing Systems - Proceedings*, 2020.
- [29] M. Mackeprang, C. Müller-Birn, and M. Stauss. Discovering the sweet spot of human—computer configurations: A case study in information extraction. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW), 2019.
- [30] Mark Watkins. Automatic image cropping with Amazon Rekognition | AWS Machine Learning Blog.
- [31] R. Parasuraman, T. B. Sheridan, and C. D. Wickens. A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans.*, 30(3), 2000.
- [32] F. Prayer, S. Röhrich, J. Pan, J. Hofmanninger, G. Langs, and H. Prosch. Künstliche Intelligenz in der Bildgebung der Lunge. *Der Radiologe*, 60(1), 2020.
- [33] P. Pu and L. Chen. A user-centric evaluation framework of recommender systems. In *CEUR Workshop Proceedings*, volume 612, 2010.
- [34] J. Rafner, D. Dellermann, A. Hjorth, D. Verasztó, C. Kampf, W. Mackay, and J. Sherson. Deskilling, Upskilling, and Reskilling: a Case for Hybrid Intelligence. *Morals & Machines*, 1(2), 2021.
- [35] T. B. Sheridan and W. L. Verplank. Human and Computer Control of Undersea Teleoperators. Technical Report DECEMBER 1978, 1978.
- [36] B. Shneiderman. Design Lessons From AI’s Two Grand Goals: Human Emulation and Useful Applications. *IEEE Transactions on Technology and Society*, 1(2), 2020.
- [37] B. Shneiderman. Human-Centered Artificial Intelligence: Reliable, Safe & Trustworthy. *International Journal of Human-Computer Interaction*, 36(6), 2020.
- [38] E. Sinagra, F. Rossi, and D. Raimondo. Use of Artificial Intelligence in Endoscopic Training: Is Deskilling a Real Fear?, 2021.
- [39] S. G. Sutton, V. Arnold, and M. Holt. How much automation is too much? Keeping the human relevant in knowledge work. *Journal of Emerging Technologies in Accounting*, 15(2), 2018.
- [40] S. Trösterer, M. Gärtner, A. Mirnig, A. Meschtscherjakov, R. McCall, N. Louveton, M. Tscheligi, and T. Engel. You never forget how to drive: Driver skilling and deskilling in the advent of autonomous vehicles. In *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Proceedings*, 2016.

- [41] Ulrich Heisig. The Deskillling and Upskilling Debate. In R. Maclean and D. Wilson, editors, *International Handbook of Education for the Changing World of Work*, chapter X.5, pages 1639–1651. Springer Science, 2009.
- [42] P. Wang, P. Liu, J. R. Glissen Brown, T. M. Berzin, G. Zhou, S. Lei, X. Liu, L. Li, and X. Xiao. Lower Adenoma Miss Rate of Computer-Aided Detection-Assisted Colonoscopy vs Routine White-Light Colonoscopy in a Prospective Tandem Study. *Gastroenterology*, 159(4), 2020.
- [43] W. Wang and J. Shen. Deep Cropping via Attention Box Prediction and Aesthetics Assessment. In *Proceedings of the IEEE International Conference on Computer Vision*, volume 2017-October, 2017.
- [44] W. Xu. Toward human-centered AI: A perspective from human-computer interaction. *Interactions*, 26(4), 2019.
- [45] Y. Zhang, Q. Vera Liao, and R. K. Bellamy. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 2020.
- [46] N. n. Zheng, Z. y. Liu, P. j. Ren, Y. q. Ma, S. t. Chen, S. y. Yu, J. r. Xue, B. d. Chen, and F. y. Wang. Hybrid-augmented intelligence: collaboration and cognition, 2017.
- [47] L. Zhong, F. H. Li, H. Z. Huang, Y. Zhang, S. P. Lu, and J. Wang. Aesthetic-guided outward image cropping. *ACM Transactions on Graphics*, 40(6), 2021.
- [48] J. Zihlsler, D. Schwager, P. Hock, P. Szauer, M. Walch, E. Rukzio, and K. Dzuba. Carvatar: Increasing trust in highly-Automated driving through social cues. In *AutomotiveUI 2016 - 8th International Conference on Automotive User Interfaces and Interactive Vehicular Applications, Adjunct Proceedings*, 2016.