



Künstliche Intelligenz zur Entscheidungsunterstützung in Leitstellen des Personennahverkehrs – Technische und sozio-technische Herausforderungen

Tobias Kopp · Robin Weitemeyer · Jens Beyer · Dominic Ziegler · Roxana Hess

Eingegangen: 13. April 2023 / Angenommen: 1. August 2023
© The Author(s) 2023

Zusammenfassung Im Kontext der Industrie 5.0 gilt die gezielte Kombination von menschlicher und maschineller Intelligenz – z. B. in Form von Entscheidungsunterstützungssystemen – als besonders vielversprechend. Während eine künstliche Intelligenz (KI) Entscheidungsvorschläge unterbreitet, obliegt im Sinne einer menschenzentrierten Gestaltung den beteiligten Personen die Entscheidungshoheit. Die KI-Vorschläge basieren zumeist auf umfangreichen historischen Datenbeständen, wie sie typischerweise im öffentlichen Personennahverkehr (ÖPNV) anfallen. Diese können bspw. genutzt werden, um Mitarbeitenden in ÖPNV-Leitstellen im Falle unvorhergesehener Störungen geeignete dispositive Maßnahmen wie Umleitungen oder Fahrplanänderungen vorzuschlagen. Dieser anspruchsvolle und oftmals zeitkritische Prozess verursacht eine hohe kognitive Belastung und bietet sich damit für eine KI-basierte Unterstützung an. Praktische Erfahrungen bei der Konzeption einer solchen hybriden Intelligenzlösung und Ergebnisse aus Gesprächen mit betroffenen Mitarbeitenden verdeutlichen allerdings, dass neben technischen auch zahlreiche sozio-technische Herausforderungen an der Mensch-Maschine-Schnittstelle bestehen. Diese betreffen u. a. die Akzeptanz des Systems oder die mangelnde Formalisierbarkeit der intuitiven und individuell unterschiedlichen menschlichen Entscheidungsabläufe. Der Artikel analysiert diese Herausforderungen und trägt damit zu einer

Die Verantwortung für den Inhalt dieser Veröffentlichung liegt bei den Autor:innen.

✉ Tobias Kopp · Robin Weitemeyer
Institut für Lernen und Innovation in Netzwerken, Hochschule Karlsruhe,
Moltkestr. 30, 76133 Karlsruhe, Deutschland
E-Mail: tobias.kopp@h-ka.de

Jens Beyer
Lavrio Solutions, Roonstraße 23a, 76137 Karlsruhe, Deutschland

Dominic Ziegler · Roxana Hess
INIT Innovative Informatikanwendungen in Transport-, Verkehrs- und Leitsystemen GmbH,
Käppelestraße 4–10, 76131 Karlsruhe, Deutschland

realistischeren Bewertung des praktischen Potenzials hybrider Intelligenzlösungen bei.

Schlüsselwörter Künstliche Intelligenz · Hybride Intelligenz · Entscheidungsunterstützung · Mensch-Maschine-Schnittstelle · Herausforderungen · Öffentlicher Personennahverkehr

Artificial Intelligence for Decision Support in Local Public Transport Control Centers—Technical and Socio-technical Challenges

Abstract In the context of Industry 5.0, the targeted combination of human and machine intelligence—e.g., in the form of decision support systems—is considered particularly promising. In the sense of a human-centred design, the persons involved have the authority to make decisions, whereas artificial intelligence (AI) makes data-based suggestions. To do this, AI requires extensive historical data, such as that which regularly accumulates in public transport. These can be used, for example, to suggest suitable measures such as diversions or timetable changes to employees in public transport control centers in the event of unforeseen disruptions. This demanding and often time-critical process involves a high cognitive load, which should be reduced through the use of AI. However, practical experience and results from discussions with affected employees make it clear that with such a hybrid intelligence solution poses not only technical but also numerous socio-technical challenges at the human-machine interface. These include the acceptance of the system or the lack of formalisation of intuitive human decision-making processes. The paper analyses these challenges and contributes to a more realistic evaluation of the practical potential of hybrid intelligence solutions.

Keywords Artificial intelligence · Hybrid intelligence · Decision support · Human-machine interface · Challenges · Public transportation

1 Hybride Intelligenz im öffentlichen Personennahverkehr (ÖPNV)

Während der Fokus bei der Industrie 4.0 auf der Automatisierung lag, rücken im Rahmen des Industrie 5.0-Paradigmas die beteiligten Menschen und deren spezifischen Stärken wieder in den Vordergrund (Golovianko et al. 2023). Mit dieser Aufwertung des menschlichen Einflussbereichs verschiebt sich der Fokus von der Etablierung maschineller Intelligenz zur Integration der komplementären Fähigkeiten von Menschen und Maschinen zu einer hybriden Intelligenz (Dellermann et al. 2019). Insbesondere nehmen Menschen eine aktive Rolle bei Entscheidungsprozessen ein und sind damit „back in the loop“. Gleichzeitig werden klassische Entscheidungsunterstützungssysteme (decision support systems, DSS) mit künstlicher Intelligenz (KI) zu intelligenten DSS (IDSS) oder auch KI-getriebenen DSS (AI-DSS) aufgewertet, die nicht nur entscheidungsrelevante Daten bereitstellen, sondern sogar Entscheidungen vorschlagen (Braun et al. 2020; Phillips-Wren 2012). Da-

mit einher gehen zahlreiche u. a. sozio-technische und ethische Fragestellungen und Herausforderungen bei der Gestaltung der Mensch-Maschine-Schnittstelle und der Konzeption des Entscheidungsprozesses.

Der Prozess der Entscheidungsfindung im Kontext der Industrie 5.0 lässt sich auch auf andere Anwendungsdomänen übertragen (Bregar 2022). Im öffentlichen Personennahverkehr (ÖPNV) finden sich DSS-Anwendungen bspw. in den Leitstellen der Verkehrsbetriebe. Dort überwachen Disponent:innen mithilfe eines rechnergestützten Betriebsleitsystems, des sogenannten *Intermodal Transport Control System* (ITCS), den Betriebsablauf, behandeln Störungen und stellen Informationen für Fahrgäste bereit (Bachmann et al. 2022). Im ÖPNV treten Betriebsstörungen wie Unfälle, Verkehrsbehinderungen, Fahrzeugdefekte etc. täglich auf. Die Hauptaufgabe der Leitstellendisponent:innen besteht darin, die Auswirkungen solcher Störungen auf die Fahrgäste möglichst gering zu halten. Zu diesem Zweck ordnen sie sogenannte dispositive Maßnahmen wie z. B. Umleitungen oder Fahrverkürzungen an, um einen möglichst zuverlässigen Betrieb aufrechtzuerhalten bzw. nach der Beseitigung der Störung zum normalen Fahrplan zurückzukehren. Die Entscheidung, ob und welche Maßnahmen ausgeführt werden, hängt von zahlreichen Faktoren und Rahmenbedingungen ab, bspw. von der Anzahl der betroffenen Fahrgäste sowie der vorgeschriebenen Lenk- und Ruhezeiten des Fahrpersonals (Carrel et al. 2010). Es handelt sich daher um komplexe multifaktorielle Entscheidungen, die situativ auf Basis der gegebenen Datenlage und häufig unter hohem Zeitdruck getroffen werden müssen. Da es sich bei der Tätigkeit der Disponent:innen um eine wissensbasierte Bildschirmarbeit handelt und in den verwendeten Systemen große Datenmengen vorliegen, erscheint der Einsatz einer KI naheliegend und aussichtsreich. Wie die folgenden Kapitel darstellen, entstehen bei der konkreten Umsetzung allerdings zahlreiche Herausforderungen. Die Ausführungen basieren auf Interviews, Workshops und Gesprächen mit Disponent:innen und Betriebsleitern in verschiedenen Verkehrsbetrieben sowie auf eigenen Erfahrungswerten bei der KI-Entwicklung im skizzierten Anwendungskontext. Detaillierte Informationen zur Datenerhebung finden sich im Online-Anhang.

2 Anwendungsidee und Umsetzung

2.1 Anwendungsidee und Ausgangsdaten

Das Ziel der geplanten KI-Anwendung besteht darin, die Disponent:innen in der Entscheidungssituation zu entlasten. Ein erster Prototyp sollte dabei situativ entscheiden, ob angesichts der aktuellen Verkehrssituation eine sog. Kurzwende sinnvoll ist. Kurzwenden stellen eine relativ häufig auftretende dispositive Maßnahme dar, die hauptsächlich eingesetzt wird, wenn auf gewissen Linien starke Verspätungen aufgelaufen sind. Sowohl angeordnete Kurzwenden als auch die aufgelaufenen Verspätungen sind in den vorhandenen Datensätzen einfach zu identifizieren. Aufgrund der vorliegenden historischen Verkehrsdaten, aus denen die situationsabhängigen Entscheidungen der Disponent:innen hervorgehen, kamen gängige Algorithmen des

überwachten Lernens (*supervised learning*) zum Einsatz (Kotsiantis 2007). Die Entwicklung verlief angelehnt an das CRISP-DM-Modell (Würth und Hipp 2000).

Für das initiale Training der KI wurde ein Datensatz herangezogen, der historische Daten aus einem Zeitraum von zweieinhalb Jahren umfasste. Dieser enthielt folgende Parameter: Uhrzeit und Datum der Fahrt, Haltestellenname, Linien- und Fahrt-ID, Verspätung der Fahrt, Abstand zur folgenden Fahrt derselben Linie. Für jede relevante Linie, Haltestelle und Fahrt wurde eine Zeile erstellt, die angibt, ob eine Kurzwende angeordnet wurde oder nicht. In einer weiteren Iteration wurde der Datensatz um die Anzahl an Haltestellen, die während der Fahrt bereits bedient wurden, sowie um die aktuelle Entfernung vom Startpunkt der Linie ergänzt. Naturgemäß enthielt der Datensatz nicht alle Daten, die den Disponent:innen im Betrieb zur Verfügung stehen. Insbesondere Informationen, die nicht digitalisiert vorliegen, wie bspw. Informationen aus Funkgesprächen der Disponent:innen mit den Fahrzeugführer:innen, konnten nicht berücksichtigt werden. Weitere unstrukturierte Daten lagen ebenfalls nicht vor.

Ein Beispiel mit leicht abgeänderten Daten ist in Abb. 1 zu sehen.

Der Fokus bei der Datenanalyse lag auf Haltestellen und Linien, an denen in der Vergangenheit Kurzwenden aufgetreten sind. Dadurch konnte die Datenmenge auf eine niedrige achtstellige Anzahl an relevanten Datensätzen reduziert werden.

Zur Untersuchung der Daten wurden klassische Methoden der Datenexploration angewendet, wie z. B. das Erstellen diverser statistischer und visueller Auswertungen, anhand derer fehlende Datenpunkte und im ETL-Prozess entstandene Fehler identifiziert und iterativ ausgebessert werden konnten.

2.2 Feature Engineering

Ein weiterer wichtiger Teilschritt bei der Entwicklung der KI war das sogenannte Feature Engineering, das sich in CRISP-DM der Data Preparation zuordnen lässt. Hierbei handelt es sich um den Prozess der Erstellung neuer aussagekräftiger Merkmale aus den vorhandenen Daten (Zheng und Casari 2018). Dieser Prozess ist besonders wichtig, wenn der gewählte Machine Learning (ML)-Algorithmus nicht in der Lage ist, die zugrundeliegenden Zusammenhänge selbst zu lernen. Dies kann daran liegen, dass nicht genügend Daten zur Verfügung stehen oder entscheidende Faktoren in den Daten nicht enthalten sind, wie im vorliegenden Fall bspw. in Hinblick auf den Zeitstempel. Aus anderen Datenquellen und individuellem Erfahrungswissen ist bekannt, dass zu bestimmten Zeiten typischerweise ein überdurchschnittlich hohes Verkehrsaufkommen herrscht. Dies lässt sich rein aus der im Datensatz enthaltenen Uhrzeit durch den ML-Algorithmus nicht lernen, sondern bedarf einer durch die Entwickler:innen hinzugefügten Spalte für „Hohes Verkehrsaufkommen“, die abhängig von der Zeitspalte automatisch berechnet wird. Informationen über das Verkehrsaufkommen wurden bei dem entwickelten KI-Prototyp nicht berücksichtigt; dies ist jedoch für zukünftige Iterationen vorgesehen. Zusätzlich wurden Entfernungen zu Start- und Zielhaltestellen aus der aktuellen Haltestellennummer seit Beginn der Fahrt und der Gesamtzahl der Haltestellen berechnet und als weitere relevante Merkmale dem Datensatz hinzugefügt. Die Information über die Strecke, die ein Passagier im aktuellen Fahrzeug maximal noch zurücklegen kann, hatten die

Linie	Pattern	Ereignis	Datum	Abfahrzeit	Halte- stelle	Diff	Haltezeit	Gesamt- verspätung	Vor- sprung	Nr. Halt	Gesamt- halte	Distanz von Start	Gesamt- strecke
7	3452	29938	01-11	41727	Mer4:4	17	360.0	70.0	NaN	10	19	1252	7423
8	34562	109378	01-11	21217	pcgh:5	907	780.0	340.0	150.0	12	23	12121	23100

Abb. 1 Beispieldaten, leicht abgeändert

Disponent:innen als entscheidungsrelevant hervorgehoben. Ebenso wurde das Datum zerlegt in Wochentag, Monatstag und Monat als Merkmal integriert, um etwaige Abhängigkeiten zu erkennen.

Dieser Prozess wurde iterativ gemäß CRISP-DM mit den Prozessschritten Modell trainieren und auswerten durchgeführt. Durch das Feature Engineering ließen sich die KI-Metriken iterativ verbessern, allerdings bisher ohne ein zufriedenstellendes und praxistaugliches Gesamtergebnis zu erzielen.

2.3 Algorithmenauswahl

Es wurden verschiedene gängige ML-Algorithmen auf reduzierten Daten trainiert und mittels der jeweils passenden Methoden die Hyperparameter optimiert. Die Datenreduktion erfolgte mittels sog. Downsamplings der Normalfahrten, d. h. es wurde nur ein Teil der Normalfahrten in die Trainingsdaten übernommen. Dies war für die meisten Algorithmen notwendig, um den Speicherbedarf im Vorhersagefall (K Nearest Neighbours), die Trainingszeit oder den Trainingspeicherbedarf ausreichend gering zu halten. Ebenfalls können diese Algorithmen nicht mit so einem großen Ungleichgewicht der Daten umgehen. Die Evaluation erfolgte auf einem repräsentativen, d. h. nicht reduzierten, Datensatz. Die Metriken zeigten bei allen Algorithmen einen nicht zufriedenstellenden Wertebereich (vergleiche Tab. 1).

Als konkreter Algorithmus aus dem Bereich des überwachten Lernen wurde LightGBM (Ke et al. 2017), ein spezieller Typ von Boosting-Algorithmen, verwendet. Boosting-Algorithmen gehören zu den wichtigsten Algorithmen des maschinellen Lernens und zeichnen sich durch ihre Fähigkeit aus, eine hohe Vorhersagegenauigkeit zu erzielen (Ke et al. 2017; Natekin und Knoll 2013; Schapire 2003). Insbesondere ist LightGBM für die Klassifizierung von Datensätzen geeignet, die eine große Anzahl an Variablen aufweisen. LightGBM verwendet eine Kombination von Entscheidungsbäumen, um die Bedeutung jeder Variablen für die Vorhersage des Zielwerts zu bestimmen. Darüber hinaus können auch nicht-lineare Zusammenhänge zwischen den Variablen berücksichtigt werden, was insbesondere für die geplante

Tab. 1 Ausgewählte Algorithmen und Metriken

Algorithmus	Datenreduktion erforderlich?	Accuracy	F1 Score	Precision	Recall	Matthews	ROC-AUC
Decision Tree	Ja	0,98	0,07	0,01	0,73	0,07	0,83
K Nearest Neighbour	Ja	0,99	0,01	0,01	0,34	0,04	0,76
Gaussian NB	Ja	0,98	0,01	0,00	0,07	0,05	0,82
Random Forest	Ja	0,98	0,03	0,01	0,73	0,09	0,85
SVM	Ja	0,99	0,04	0,02	0,33	0,08	Nicht anwendbar
Tabnet	Ja	0,99	0,04	0,01	0,34	0,06	0,73
LightGBM	Nein	0,99	0,07	0,03	0,48	0,10	0,90

Vorhersage der Kurzwenden relevant ist. Zusammenfassend bietet LightGBM also viele Vorteile für das vorliegende Anwendungsszenario. Dieser Eindruck bestätigte sich durch Tests alternativer Algorithmen wie neuronaler Netze (z. B. in der Implementierung TabNET, Arik und Pfister 2021). So schnitten diese Algorithmen bei gängigen Metriken wie F-Maß und ROC-AUC (Sokolova und Lapalme 2009; Hosin und Sulaiman 2015) deutlich schlechter ab als LightGBM. Konkret wurden viel mehr Normalfahrten als Kurzwenden klassifiziert und tatsächliche Fahrten, in denen Kurzwenden angeordnet wurden, wurden nicht als solche erkannt. Darüber hinaus dauerten Trainingsläufe bei Verwendung von neuronalen Netzen deutlich länger als bei Verwendung von LightGBM, teilweise konfigurationsabhängig mehr als 100-mal so lang. Im Kontext unseres Anwendungsszenarios und bei gegebener Datenlage zeigte sich LightGBM also als technisch beste Wahl, da dieser im Gegensatz zu den restlichen in Tab. 1 aufgeführten Algorithmen keine Datenreduktion benötigte und wesentlich kürzere Trainingszeiten aufwies, was sich in der Retrospektive angesichts zahlreicher Iterationen als vorteilhaft aus praktischer Hinsicht herausstellte.

2.4 Erklärbare KI mit SHAP

Das Ziel des Projekts bestand darin, die Disponent:innen bei ihren Entscheidungen zu unterstützen. Im Sinne einer transparenten Unterstützung und einer hohen Akzeptanz wurde dabei besonders auf die Erklärbarkeit der KI-Empfehlungen geachtet. Bei jedem Entscheidungsvorschlag sollte ersichtlich sein, welche Faktoren das Modell maßgeblich berücksichtigt hat. Hierzu bietet sich die Erweiterung von LightGBM um Methoden der Erklärbaren KI (XAI) an. Im Zuge der Entwicklung fiel die Entscheidung auf die XAI-Methode SHAP (SHapley Additive exPlanations) (Lundberg und Lee 2017). Diese Methode hat den Vorteil, dass sie modell-agnostisch und somit auf jedes KI-Modell anwendbar ist. Weiterhin ermöglicht sie es, sowohl lokale Erklärungen für einzelne KI-Entscheidungen als auch globale Erklärungen mit Korrelationen zwischen Ein- und Ausgaben der KI zu erzeugen, um das gesamtethliche Verhalten des KI-Modells nachvollziehen zu können (Ekanayake et al. 2022).

Eine beispielhafte Erklärung eines Entscheidungsvorschlags ist in Abb. 2 ersichtlich. In diesem Beispiel empfiehlt die KI das Einleiten einer Kurzwende. Insbesondere die Verspätung der Bahn, aber auch die Distanz zum Linienende ist für die KI von großer Bedeutung. Die Distanz vom Start hingegen wirkt sich aus Sicht des Modells gegenläufig aus.

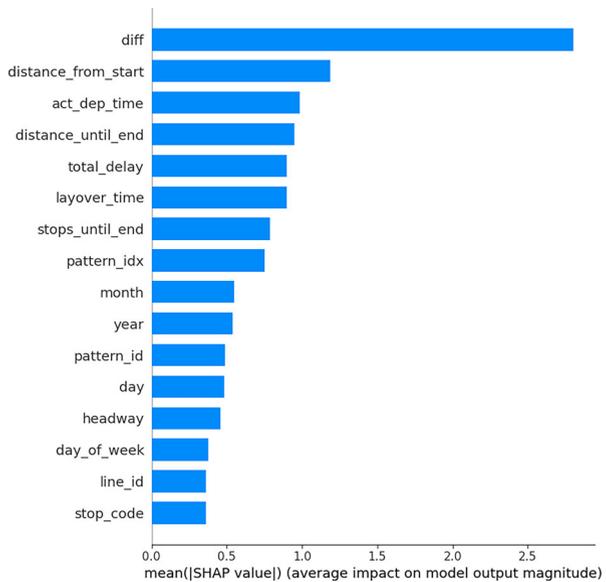
Die durch SHAP berechneten Shapley Values lassen sich auch über die gesamte Population betrachten. Daraus ergibt sich ein transparentes Bild, auf welche Faktoren der Algorithmus beim Training geachtet hat (Abb. 3). Für die praktische Anwendung in der Interaktion mit den Disponent:innen erscheint eine auf die wichtigsten Faktoren reduzierte Liste der laut SHAP einflussreichsten Faktoren als am zielführendsten und handhabbarsten.

Die Bestimmung des Einflusses einzelner Merkmale durch SHAP kann einerseits das Vertrauen in die KI-Anwendung erhöhen und andererseits von KI-Entwickler:innen genutzt werden, um ein besseres Verständnis über die Problemstellung und die Daten zu erhalten. Die Verwendung von XAI-Methoden stellt daher einen



Abb. 2 Erklärung einer einzelnen Vorhersage mittels SHAP

Abb. 3 Einfluss der Faktoren laut SHAP auf das Modell



relevanten Bestandteil der Evaluation dar und liefert wichtige Informationen für die Business und Data Understanding-Phasen der nächsten Iteration des CRISP-DM-Prozesses.

2.5 Ergebnisse anhand ausgewählter Metriken

Eine große Herausforderung bestand darin, Metriken zu definieren, die sowohl aussagekräftig als auch für geplante Diskussionen mit Domänenexpert:innen verwendbar waren. Klassische Metriken wie F-Maß oder ROC-AUC erwiesen sich für Fachfremde als schwer nachvollziehbar (Hand et al. 2021; Fawcett 2004). Alternativ wird üblicherweise die sogenannte Konfusionsmatrix verwendet, die alle Vorhersagen der KI vier verschiedenen Kategorien zuordnet und folgende Angaben enthält:

- Die richtig vorhergesagten Normalfahrten, d. h. die Fahrten ohne Kurzwende, die die entwickelte KI ebenfalls nicht als Kurzwende eingestuft hat (oben links in der Matrix),
- die korrekt vorhergesagten Kurzwenden (unten rechts),
- die tatsächlichen Kurzwenden, die die KI nicht identifiziert hat (unten links)
- die falsch vorhergesagten Kurzwenden bei Fahrten, bei denen in Wirklichkeit keine Kurzwende durchgeführt wurde (oben rechts).

Abb. 4 und 5 zeigen zwei Matrizen basierend auf unterschiedlichen Modellkonfigurationen, die beide auf richtig vorhergesagte Kurzwenden normiert sind, also verdeutlichen, wie viele Fehler pro richtig erkannter Kurzwende entstehen. Anhand solcher Matrizen lässt sich die Performance der KI auch gegenüber KI-Laien relativ gut erklären. Entscheidend ist hierbei der notwendige Abwägungsprozess zwischen einer höheren Anzahl an falschen Vorschlägen von Kurzwenden und einer höhe-

		Vorhersage	
		Normalfahrt	Kurzwende
Realität	Normalfahrt	> 40.000	> 50
	Kurzwende	2.5	1.0

Abb. 4 Konfusionsmatrix für die Kurzwenden-Vorhersage (erstes Modell), normiert auf richtig vorhergesagte Kurzwenden. Die Matrix zeigt mehr als 50 Fehllarmer (*rot* hinterlegt) pro richtig erkannter Kurzwende. 2,5 Kurzwenden pro richtig vorhergesagter Kurzwende bleiben unerkannt

		Vorhersage	
		Normalfahrt	Kurzwende
Realität	Normalfahrt	> 100.000	0.4
	Kurzwende	> 30	1.0

Abb. 5 Konfusionsmatrix für die Kurzwenden-Vorhersage (alternatives Modell), normiert auf richtig vorhergesagte Kurzwenden. Mehr als 30 Kurzwenden pro richtig erkannter Kurzwende bleiben unerkannt. Gleichzeitig werden weniger als halb so viele Fehllarmer ausgelöst als richtige Alarme. Insgesamt werden somit extrem wenig Kurzwenden erkannt

ren Anzahl nicht erkannter Kurzwenden. Je nach Anwendungssituation sind diese Klassifikationsfehler als unterschiedlich kritisch einzuschätzen, sodass die Disponent:innen keine allgemeingültige und situationsübergreifende Abwägung treffen konnten. Da das übergeordnete Ziel in der Reduktion der Auswirkung auf die Fahrgäste besteht, kann sich eine fälschlicherweise vorgeschlagenen Kurzwende ähnlich negativ auswirken wie eine nicht erkannte, aber notwendige Kurzwende.

Die Diskussionen mit den Verkehrsexpert:innen zeigten schnell, dass angesichts der erzielten Vorhersagegenauigkeit noch keine praktischen Vorteile entstehen. Neue Merkmale, die aus weiteren Datenquellen zugeführt bzw. berechnet werden können, sollen weitere Verbesserungen bringen. Dazu zählt u. a. die Auslastung der Wagen, die entweder durch Fahrgastzähler ermittelt oder statistisch errechnet werden. Diese Verbesserungsansätze werden künftig zunächst iterativ erprobt, bevor der praktische Einsatz des Modells (in CRISP-DM „Deployment“) vorgenommen wird.

Tab. 2 gibt einen Überblick zu den Schritten des CRISP-DM Prozesses, die bisher im Projekt durchlaufen wurden. Für jeden Schritt wird jeweils die zentrale Fragestellung, die getroffene Entscheidung und eine dazugehörige Begründung bzw. Erläuterung aufgeführt.

Zusammenfassend haben sich im Projektverlauf vielfältige praktische Herausforderungen herauskristallisiert, die sich auf vergleichbare KI-Projekte übertragen lassen. Durch die Ausgestaltung als hybride Intelligenz mit der entsprechend entstehenden Mensch-Technik-Interaktion sind diese Herausforderungen nicht nur technischer, sondern mitunter auch sozio-technischer Natur. Die Herausforderungen lassen sich auf gängige Dimensionen der Datenqualität beziehen, insbesondere auf Zugänglichkeit (accessibility), Vollständigkeit (completeness) und Aktualität (timeliness) in Hinblick auf technische Herausforderungen sowie auf Objektivität (objectivity) und

Tab. 2 Übersicht der Fragestellungen und Entscheidungen entlang der Phasen von CRISP-DM

Phase aus CRISP-DM	Fragestellung	Entscheidung	Begründung/Erläuterung
Business Understanding	Für welche dispositive Maßnahme soll der erste Prototyp des Assistenzsystems Vorschläge generieren?	Kurzwenden	Vglw. einfacher Anwendungsfall, beste Datenlage, einfache und nützliche praktische Anwendung
Data Understanding	Welche zusätzlichen Informationen sind für die Entscheidungsfindung relevant und können aus den vorhandenen Daten extrahiert werden?	Entfernung zu Start- und Zielbahnhöfen	Berechnung aus aktueller Haltestelle und Gesamtzahl der Haltestellen auf der Linie
Data Preparation	Wie soll mit unausgeglichene Trainingsdaten umgegangen werden?	Undersampling	Reduzierung der Normalfahrten, Reduktion der Datenmenge für manche KI-Algorithmen notwendig
		Anwendung robuster KI-Algorithmen	keine Manipulation der Verteilung der Trainingsdaten notwendig
Modelling	Welcher KI-Algorithmus eignet sich am besten für die Entwicklung des Assistenzsystems?	LightGBM	Keine Datenreduktion notwendig, kürzeste Trainingsdauer, vglw. gute Klassifikationsergebnisse
Evaluation	Welche Metriken sollen als primäre Bewertungskriterien zur Beurteilung der trainierten KI-Modelle betrachtet werden?	Konfusionsmatrix	Gute Verständlichkeit für Laien, Optimierung des KI-Modells anhand falsch-positiver und falsch-negativer Klassifizierungen

Glaubwürdigkeit (believability) in Hinblick auf sozio-technische Herausforderungen (Wang und Strong 1996). Diese werden im folgenden Kapitel detailliert erläutert.

3 Technische Herausforderungen

3.1 Unzureichende Dokumentation wichtiger qualitativer Daten

Eine der wichtigsten Informationsquellen für die Disponent:innen in der Leitstelle sind Gespräche per Funk oder Telefon mit dem Fahrpersonal, dem Aufsichtspersonal im Außendienst sowie externen Stellen, z. B. der Polizei oder der Feuerwehr. Hierüber erhalten sie insbesondere Informationen über Unfälle, blockierte Streckenabschnitte und die voraussichtliche Störungsdauer. Die mündlich mitgeteilten Informationen werden zwar in der Regel dokumentiert, jedoch teilweise unvollständig und kaum strukturiert, was eine automatische Auswertung stark erschwert. Mitunter enthalten die Dokumentationen außerdem sensible personenbezogene Daten, wie z. B. Angaben zu Unfallbeteiligten. Dadurch ergeben sich datenschutzrechtliche Hürden bei der Analyse der Daten, obwohl der Personenbezug für das KI-Training irrelevant wäre. Unter Umständen müssen diese Daten erst aufwändig manuell bereinigt werden, bevor sie sich für die Auswertung nutzen lassen. Die Dokumentation erfolgt außerdem oft mit großem Zeitverzug, da den Disponent:innen in der

Erstphase des Störungsmanagements keine Zeit für die Erfassung bleibt. Das Einleiten zeitkritischer Sofortmaßnahmen (wie z. B. das Anfordern des Rettungsdiensts) genießt Vorrang. In der Folge fehlen für das Training der KI wichtige Informationen, die bei den Entscheidungen der Disponent:innen eine Rolle spielen. Ein möglicher Lösungsansatz, um die kritischen Informationen aus den Funkgesprächen für das Training und den Einsatz des KI-Modells nutzen zu können, wäre die Aufzeichnung und Verarbeitung des Funkverkehrs durch Methoden der automatischen Spracherkennung. So können Spracheingaben in Textausgaben umgewandelt werden (*Speech-to-Text*). Diese unstrukturierten Daten lassen sich mit Hilfe von Verfahren der natürlichen Sprachverarbeitung (*natural language processing, NLP*) ohne manuelle Vorverarbeitung von einer KI interpretieren (Otter et al. 2021). Diese Daten standen im konkreten Projekt jedoch nicht zur Verfügung, da der Funkverkehr nicht aufgezeichnet wird und auch kein Zugriff auf den Live-Betrieb der Leitstellen bestand.

3.2 Integration verschiedener Datenbasen

Im vorliegenden Fall sind die relevanten Daten über verschiedene und zwischen unterschiedlichen Verkehrsbetrieben teils variierende Systeme verteilt. Dazu gehören beispielsweise zum einen das Betriebsleitsystem, aus dem auch die Disponent:innen Informationen – etwa zum Standort der Fahrzeuge – beziehen, und zum anderen ein System zur Dokumentation von Betriebsmeldungen, in das die Disponent:innen mündlich in Erfahrung gebrachte Informationen (s. Abschn. 3.1) eingeben. Die unterschiedlichen Systeme arbeiten weitgehend getrennt voneinander, sodass die erhaltenen Daten zunächst zusammengeführt und zueinander in Beziehung gesetzt werden müssen. Davon abhängig, ob auch unstrukturierte Daten verwendet werden, wäre die Implementierung eines Data Warehouse oder Data Lake ein geeigneter Lösungsansatz, um die unterschiedlichen Datenbasen in einer zentralen Datenbank zusammenzuführen (Miloslavskaya und Tolstoy 2016). Die besondere Herausforderung für die geplante KI-Anwendung bestand jedoch darin, dass ein Zugriff auf die Systeme anderer IT-Lösungsanbieter nicht möglich war.

3.3 Unausgeglichene Datensätze

Die Datenbasis an sich ist mit Millionen eingetragener Datenpunkte pro Jahr zwar sehr groß, jedoch ist die Menge an relevanten dispositiven Maßnahmen im Betrachtungszeitraum relativ gering (Größenordnung wenige Tausend). Diese Beobachtung reflektiert im Fall eines Assistenzsystems für dispositive Maßnahmen im ÖPNV die reale Situation, dass Störungen im Vergleich zur gesamten Betriebsdauer in geringerem Maße auftreten. Unausgeglichene Datensätze sind ein häufig vorkommendes Problem bei vielen Anwendungsfällen im Bereich der KI und setzen eine individuelle Lösungsfindung voraus (Kaur et al. 2020; Krawczyk 2016). Typische Lösungsansätze sind dabei im binären Klassifikationsfall das Undersampling der größeren bzw. das Oversampling der kleineren Klasse (Krawczyk 2016). Für diese KI-Anwendung wurde das Training auf dem unausgebalancierten Datensatz mit dem Undersampling und Oversampling verglichen. Der verwendete ML-Algorithmus LightGBM hat sich

hier im direkten Vergleich jedoch als von sich aus geeignet im Umgang mit solchen Datenverteilungen herausgestellt, sodass ein Under- bzw. Oversampling obsolet war.

3.4 Veraltete historische Daten

Aufgrund eines jahrelangen Umbaus des Schienennetzes sind große Teile der historischen Daten aus mehreren Gründen veraltet. Zum einen entsprechen sie nicht mehr dem aktuellen Schienennetz, da sich Haltestellen und Linienfahrpläne geändert haben, und zum anderen mussten Linienverläufe über die Jahre hinweg immer wieder geändert werden, um die wandernden Baustellen zu umfahren. Diese Daten eignen sich daher nicht mehr zum Trainieren einer KI, die konkrete Haltestellen und Linien betrachtet, weshalb zunächst eine neue Datenbasis auf dem aktuellen Schienennetz aufgebaut werden muss. Diese lag zum momentanen Zeitpunkt jedoch noch nicht vor, weshalb eine weitere Iteration des KI-Trainings nicht durchgeführt werden konnte. Ferner bestehen für geänderte Streckenabschnitte mitunter neue Vorgaben in Hinblick auf die im Störfall einzuleitenden dispositiven Maßnahmen. Die vergangenen Maßnahmen entsprechen daher nicht zwangsläufig den aktuellen Anforderungen. Die mitunter begrenzte Lebensdauer der gesammelten Daten (*big data volatility*) ist im Kontext von Big Data eine typische Problematik (Khan et al. 2014).

4 Sozio-technische Herausforderungen

4.1 KI-Training auf Basis suboptimaler historischer Daten

Disponent:innen in den Leitstellen treffen ihre dispositiven Entscheidungen mitunter unter hohem Zeitdruck. Entsprechend ist davon auszugehen, dass diese Entscheidungen in vielen Fällen suboptimal sind. Zwar werden bei größeren Störfällen im Nachgang häufig Review-Gespräche in der Leitstelle durchgeführt, allerdings sind deren Ergebnisse in den historischen Daten nicht hinterlegt. Eine nachträgliche Bewertung der eingeleiteten dispositiven Maßnahmen anhand der vorliegenden Daten ist ohne Kontextwissen nicht möglich. Ferner werden aufgrund der bestehenden Arbeitsbelastung und des hohen Zeitdrucks in Störungssituationen dispositive Maßnahmen mitunter erst zeitlich verzögert oder gar nicht eingeleitet, wenn gerade wichtigere Störfälle die volle Aufmerksamkeit erfordern. Somit lässt sich im Nachhinein nicht herausfinden, ob zu einem gegebenen Zeitpunkt bewusst auf eine Maßnahme verzichtet oder diese nur aufgrund zeitlicher Rahmenbedingungen nicht oder zu einem späteren Zeitpunkt ausgelöst wurde. Entsprechend würde eine KI, die dispositive Maßnahmen auf Basis der historischen Daten vorschlägt, dazu beitragen, dass suboptimale Entscheidungen aus der Vergangenheit reproduziert werden. Damit würden eher suboptimale Entscheidungsmuster manifestiert, anstatt dass sich eine Qualitätsverbesserung der dispositiven Maßnahmen einstellt. Diese Herausforderung wird in der KI-Forschung regelmäßig im Kontext der sog. Wahrhaftigkeit von Daten (*big data veracity*) thematisiert (Gandomi und Haider 2015; L'Heureux et al. 2017).

4.2 Heterogene Entscheidungen der Disponent:innen

Erfahrungsgemäß treffen die jeweiligen Disponent:innen individuell unterschiedliche Entscheidungen bei dispositiven Maßnahmen, da – je nach Verkehrsunternehmen und der jeweiligen Situation – keine einheitlichen und vordefinierten Handlungsanweisungen vorliegen. In die Entscheidungen fließen neben Kontextfaktoren wie den aktuellen Wetterdaten, die sich potenziell als Kriterium in das KI-Training einbeziehen lassen, auch bspw. subjektive Eindrücke aus telefonischen Gesprächen mit dem Fahrpersonal oder schwer formalisierbare Faktoren wie spezifische Ortskenntnisse mit ein, die mögliche Fahrtwege beeinflussen. Für das Training einer KI ergibt sich daraus die Herausforderung, dass sich keine gleichförmigen Handlungsmuster in den Daten zeigen, sondern die subjektiven Entscheidungspräferenzen des jeweiligen Disponent:innen als Störvariable einfließen. Insofern sind entsprechende Reaktionsmuster in den Daten durch subjektive Einflüsse verzerrt und für eine KI nur schwerlich generisch zu extrahieren. Um nicht von verzerrten historischen Daten abhängig zu sein, könnte ein regelbasiertes Expertensystem einen vielversprechenden alternativen Ansatz darstellen (Grosan und Abraham 2011). Die fehlende Vorgabe von Handlungsanweisen erschwert dies jedoch, da ein solches Expertensystem eine Wissensbasis benötigt, die das formalisierte Expertenwissen repräsentiert. Diese meist in Form von Wenn-dann-Regeln dargestellte Wissensbasis müsste daher zunächst aufgebaut werden. Dies kann z. B. in Form von Experteninterviews erfolgen (Yoon et al. 1994; Clancey 1983), was aufgrund der starken Heterogenität in den Entscheidungen der Disponent:innen jedoch einen erheblichen Aufwand bedeuten würde.

4.3 Unklare Zielhierarchien bei dispositiven Maßnahmen

Die Zufriedenheit der Fahrgäste ist typischerweise die oberste Prämisse der Verkehrsbetriebe. Insbesondere soll vermieden werden, dass Fahrgäste an einer gewissen Straßenbahnstation nicht mehr weiterbefördert werden können. Konkrete Störungssituationen sind allerdings häufig komplexer und produzieren Zielkonflikte, die sich nicht eindeutig auflösen lassen. So existieren bspw. keine festgeschriebenen Regeln in Hinblick darauf, wie eine etwaige Umleitungsbedingt längere Fahrzeit gegenüber einzelnen Fahrtausfällen abzuwägen ist. Aufgrund dieser konzeptionellen Unschärfe bei den Zieldefinitionen lässt sich für eine KI nicht eindeutig festlegen, was in einem konkreten Fall die beste dispositive Maßnahme darstellt. Besonders kritisch wäre diese Herausforderungen für alternative KI-Lösungsansätze nach Methoden des bestärkenden Lernens (*reinforcement learning*) (Sutton und Barto 2018). Diese würden sich zwar prinzipiell für den betrachtenden Anwendungsfall eignen (Haydari und Yilmaz 2022), bedürften allerdings einer klar vorgegebenen Zielhierarchie, um eine geeignete Belohnungsfunktion definieren zu können (Dulac-Arnold et al. 2021). An dieser Stelle besteht demnach zunächst konzeptioneller Nachbesserungsbedarf vor dem KI-Training.

4.4 Objektivität und Glaubwürdigkeit relevanter Informationsquellen

Wie bereits beschrieben, stellen Funkgespräche mit dem Fahrpersonal eine der wichtigsten Informationsquellen für die Disponent:innen dar. Besonders bei Störungen und Unfällen, die durch Personen verursacht wurden, sind Disponent:innen grundlegend auf die Einschätzungen des Personals vor Ort angewiesen. Objektivität (objectivity) und Glaubwürdigkeit (believability) stellen zwei wichtige Dimensionen der Datenqualität dar (Wang und Strong 1996), die in diesen (Unfall-)Situationen allerdings nicht garantiert werden können. Das Fahrpersonal ist möglicherweise direkt in die Situation involviert und kann je nach Schwere des Unfalls unter Schock stehen. Um dieser Problematik entgegenzuwirken, wird von den Verkehrsbetrieben mobiles Aufsichtspersonal entsendet, um eine objektivere Einschätzung der Lage direkt vor Ort zu erhalten. Aufgrund der begrenzten Kapazitäten kann dies jedoch nicht immer gewährleistet werden. Umso wichtiger sind daher die zwischenmenschlichen Fähigkeiten der Disponent:innen als Empfänger:innen der Informationen. Diese können im Dialog und in der Interaktion mit ihren Gesprächspartner:innen die Situation besser einordnen und die Belastbarkeit der erhaltenen Informationen abschätzen. Selbst wenn der KI die über Funk erhaltenen Informationen zur Verfügung stünden, stellt die Einschätzung der Disponent:innen auf Basis ihrer sozialen Fähigkeiten weiterhin einen kritischen Vorverarbeitungsschritt dar.

4.5 Mangelnde Akzeptanz und Angst vor technologischer Ersetzung

In den Gesprächen wiesen die Disponent:innen auf die Notwendigkeit eines gewissen Bauchgefühls hin, das für gute Entscheidungen unerlässlich sei. Möglicherweise hegen die Disponent:innen Vorbehalte dagegen, ihr prozedurales und tazites Erfahrungswissen explizit und verfügbar zu machen, um weiterhin als Schlüsselpersonen mit exklusivem Wissen unersetzbar zu bleiben (Schiedermaier et al. 2023). Dieses Verhalten ist getrieben von Bedenken, perspektivisch komplett durch eine KI ersetzt zu werden, wenngleich der Fokus bei Industrie 5.0 explizit auf der Ergänzung anstatt der Ersetzung menschlicher Intelligenz liegt (Bregar 2022). Vergleichbare Ängste zeigen sich auch regelmäßig bei der Einführung assistierender Technologien im Kontext der Industrie 4.0 (Kopp 2022; Kopp et al. 2022). Darüber hinaus sind tazites Wissen bzw. tazite Fähigkeiten häufig unbewusst und selbst bei bestehender Bereitschaft nur schwer zu artikulieren und zu formalisieren (Ambrosini und Bowman 2001).

Ferner stellt ein mangelndes Vertrauen in die Leistungsfähigkeit einer entscheidungsunterstützenden KI ein weiteres Akzeptanzhemmnis dar, insbesondere weil die Disponent:innen das subjektive Bauchgefühl als wichtigen positiven Einflussfaktor betrachten, der sich nicht quantitativ erfassen und einberechnen lässt. Die Vertrauenswürdigkeit der KI durch derzeit intensiv diskutierte Ansätze der erklärbaren KI (XAI) (Confalonieri et al. 2021) zu erhöhen, stieß in den Expert:innengesprächen ebenfalls auf wenig Zuspruch. XAI als technischer Lösungsansatz scheint nicht ohne Weiteres die komplexer gelagerten sozio-technische Herausforderungen beseitigen zu können (Dvorak et al. 2022). An dieser Stelle wird auch ein gewisser Widerspruch deutlich. Dieser besteht zwischen der allgemeinen Forderung nach

Transparenz und Erklärbarkeit von KI-Systemen einerseits und des aus Sicht der Anwendenden unverzichtbaren und naturgemäß nicht letztgültig erklär- und formalisierbaren Bauchgefühls andererseits. XAI impliziert aus Sicht der Disponent:innen letztlich, dass ihr Erfahrungswissen in konkrete Kriterien und Formeln überführbar sei und kann somit als Abwertung ihrer als einzigartig eingeschätzten Intuition erfahren werden. Laut den Disponent:innen fehle in den Entscheidungssituationen außerdem schlicht die Zeit, um ausführliche Erklärungen einer KI wahrzunehmen und in die Entscheidung einzubeziehen. Des Weiteren besteht die Gefahr, dass das wichtige Bauchgefühl durch ein Übermaß an KI-Unterstützung verloren geht. Der Einsatz von XAI-Methoden könnte neben der Vertrauenssteigerung jedoch trotzdem einen Mehrwert für die Disponent:innen darstellen, indem sie von der KI erkannte Muster in den Daten offen legt und Erkenntnisse dazu ermöglicht, in welchen Situationen dispositive Maßnahmen besonders oft getroffen und welche Kriterien dafür primär herangezogen werden. Diese Erkenntnisse könnten auch ohne KI-Einsatz im laufenden Betrieb die manuellen Entscheidungen der Disponent:innen informieren und optimieren.

5 Fazit

Hybrider Intelligenz als Bestandteil des Industrie 5.0-Paradigmas wird ein hohes Potenzial zugeschrieben, u. a. im Kontext der KI-basierten Unterstützung menschlicher Entscheidungsprozesse, die unter hoher kognitiver Belastung und hohem Zeitdruck ablaufen. Allerdings zeigen praktische Erfahrungen, dass zusätzliche Herausforderungen an der wieder verstärkt in den Fokus rückenden Schnittstelle zwischen menschlicher und maschineller Intelligenz entstehen, die eine breite Adoption von KI-Lösungen in realistischen Anwendungsszenarien noch behindern. Zu den typischen technischen Herausforderungen gesellen sich bei hybriden Intelligenzlösungen weitere sozio-technische Herausforderungen, wie in diesem Artikel auf Basis eines Entscheidungsunterstützungssystems im Kontext des ÖPNV exemplarisch aufgezeigt. Es ist davon auszugehen, dass vergleichbare Herausforderungen bei ähnlich gearteten Systemen im Bereich der industriellen Fertigung auftreten. Für die Ausgestaltung hybrider Intelligenz benötigt es demnach neuartige Umsetzungs- und Lösungsansätze, um den beschriebenen Herausforderungen zu begegnen und das Potenzial in der Praxis zu erschließen.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (<https://doi.org/10.1365/s40702-023-00996-8>) enthalten.

Danksagung Ein ganz besonderer Dank gilt den Disponent:innen, die uns im Rahmen von Expert:inneninterviews und Workshops wertvolle Einsichten gewährten.

Förderung Die Publikation entstand im Projekt „Kompetenzzentrum KARL – Künstliche Intelligenz für Arbeit und Lernen in der Region Karlsruhe“. Das Projekt wird durch das Bundesministerium für Bildung und Forschung (BMBF) im Programm „Zukunft der Wertschöpfung – Forschung zu Produktion, Dienstleistung und Arbeit“ (Förderkennzeichen: 02L19C250) gefördert.

Funding Open Access funding enabled and organized by Projekt DEAL.

Interessenkonflikt T. Kopp, R. Weitemeyer, J. Beyer, D. Ziegler und R. Hess geben an, dass kein Interessenkonflikt besteht.

Open Access Dieser Artikel wird unter der Creative Commons Namensnennung 4.0 International Lizenz veröffentlicht, welche die Nutzung, Vervielfältigung, Bearbeitung, Verbreitung und Wiedergabe in jeglichem Medium und Format erlaubt, sofern Sie den/die ursprünglichen Autor(en) und die Quelle ordnungsgemäß nennen, einen Link zur Creative Commons Lizenz beifügen und angeben, ob Änderungen vorgenommen wurden.

Die in diesem Artikel enthaltenen Bilder und sonstiges Drittmaterial unterliegen ebenfalls der genannten Creative Commons Lizenz, sofern sich aus der Abbildungslegende nichts anderes ergibt. Sofern das betreffende Material nicht unter der genannten Creative Commons Lizenz steht und die betreffende Handlung nicht nach gesetzlichen Vorschriften erlaubt ist, ist für die oben aufgeführten Weiterverwendungen des Materials die Einwilligung des jeweiligen Rechteinhabers einzuholen.

Weitere Details zur Lizenz entnehmen Sie bitte der Lizenzinformation auf <http://creativecommons.org/licenses/by/4.0/deed.de>.

Literatur

- Ambrosini V, Bowman C (2001) Tacit knowledge: some suggestions for operationalization. *J Management Studs* 38:811–829. <https://doi.org/10.1111/1467-6486.00260>
- Arik SÖ, Pfister T (2021) TabNet: attentive interpretable tabular learning. *AAAI* 35:6679–6687. <https://doi.org/10.1609/aaai.v35i8.16826>
- Bachmann FR, Briem L, Busch F, Vortisch P (2022) Dynamics and processes in operations control centers in Urban public transport: potentials for improvement. *IEEE Trans Intell Transport Syst* 23:17819–17834. <https://doi.org/10.1109/TITS.2022.3158821>
- Braun M, Hummel P, Beck S, Dabrock P (2020) Primer on an ethics of AI-based decision support systems in the clinic. *J Med Ethics* 47:e3. <https://doi.org/10.1136/medethics-2019-105860>
- Bregar A (2022) Decision support for industry 5.0. In: Lobo C, Oliveira S (Hrsg) Proceedings of TAKE 2022 conference. Universidade Portucalense, Lisbon, Porto
- Carrel A, Mishalani RG, Wilson NHM, Attanucci JP, Rahbee AB (2010) Decision factors in service control on high-frequency metro line: importance in service delivery. *Transp Res Rec* 2146:52–59. <https://doi.org/10.3141/2146-07>
- Clancey WJ (1983) The epistemology of a rule-based expert system—a framework for explanation. *Artif Intell* 20:215–251. [https://doi.org/10.1016/0004-3702\(83\)90008-5](https://doi.org/10.1016/0004-3702(83)90008-5)
- Confalonieri R, Coba L, Wagner B, Besold TR (2021) A historical perspective of explainable artificial intelligence. *WIREs Data Min Knowl Discov*. <https://doi.org/10.1002/widm.1391>
- Dellermann D, Ebel P, Söllner M, Leimeister JM (2019) Hybrid intelligence. *Bus Inf Syst Eng* 61:637–643. <https://doi.org/10.1007/s12599-019-00595-2>
- Dulac-Arnold G, Levine N, Mankowitz DJ, Li J, Paduraru C, Goyal S, Hester T (2021) Challenges of real-world reinforcement learning: definitions, benchmarks and analysis. *Mach Learn* 110:2419–2468. <https://doi.org/10.1007/s10994-021-05961-4>
- Dvorak J, Kopp T, Kinkel S, Lanza G (2022) Explainable AI: A key driver for AI adoption, a mistaken concept, or a practically irrelevant feature? In: Reich C, Mescheder U (Hrsg) Artificial Intelligence—Applications in Medicine and Manufacturing. UR-AI 2022—The Upper Rhine Artificial Intelligence Symposium, S 88–97
- Ekanayake IU, Meddage D, Rathnayake U (2022) A novel approach to explain the black-box nature of machine learning in compressive strength predictions of concrete using Shapley additive explanations (SHAP). *Case Stud Constr Mater* 16:e1059. <https://doi.org/10.1016/j.cscm.2022.e01059>
- Fawcett T (2004) ROC graphs: Notes and practical considerations for researchers. *Mach Learn* 31:1–38
- Gandomi A, Haider M (2015) Beyond the hype: Big data concepts, methods, and analytics. *Int J Inf Manage* 35:137–144. <https://doi.org/10.1016/j.ijinfomgt.2014.10.007>
- Golovianko M, Terziyan V, Branytskyi V, Malyk D (2023) Industry 4.0 vs. Industry 5.0: co-existence, transition, or a hybrid. *Procedia Comput Sci* 217:102–113. <https://doi.org/10.1016/j.procs.2022.12.206>

- Grosan C, Abraham A (2011) Rule-based expert systems. In: Kacprzyk J, Jain LC, Grosan C, Abraham A (Hrsg) Intelligent systems. Springer, Berlin, Heidelberg, S 149–185
- Hand DJ, Christen P, Kirielle N (2021) F*: an interpretable transformation of the F-measure. *Mach Learn* 110:451–456. <https://doi.org/10.1007/s10994-021-05964-1>
- Haydari A, Yilmaz Y (2022) Deep reinforcement learning for intelligent transportation systems: a survey. *IEEE Trans Intell Transp Syst* 23:11–32. <https://doi.org/10.1109/itis.2020.3008612>
- Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. *IJDKP* 5:1–11. <https://doi.org/10.5121/ijdkp.2015.5201>
- Kaur H, Pannu HS, Malhi AK (2020) A systematic review on Imbalanced data challenges in machine learning. *ACM Comput Surv* 52:1–36. <https://doi.org/10.1145/3343440>
- Ke G, Meng Q, Finley T, Taifeng W, Chen W, Weidong M, Qiwei Y, Liu T-Y (2017) LightGBM: a highly efficient gradient boosting decision tree. *Advances in neural information processing systems*
- Khan MA, Uddin MF, Gupta N (2014) Seven V's of Big Data understanding Big Data to extract value. In: *Proceedings of the 2014 Zone 1 Conference of the American Society for Engineering Education*. IEEE, S 1–5
- Kopp T (2022) Vertrauen in Roboter und dessen Beeinflussbarkeit durch sprachliches Framing. *KIT Scientific*
- Kopp T, Baumgartner M, Kinkel S (2022) How linguistic framing affects factory workers' initial trust in collaborative robots: the interplay between anthropomorphism and technological replacement. *Int J Hum Comput Stud* 158:102730. <https://doi.org/10.1016/j.ijhcs.2021.102730>
- Kotsiantis SB (2007) Supervised machine learning: a review of classification techniques. *Informatica* 31:249–268
- Krawczyk B (2016) Learning from imbalanced data: open challenges and future directions. *Prog Artif Intell* 5:221–232. <https://doi.org/10.1007/s13748-016-0094-0>
- L'Heureux A, Grolinger K, Elyamany HF, Capretz MAM (2017) Machine learning with big data: challenges and approaches. *IEEE Access* 5:7776–7797. <https://doi.org/10.1109/access.2017.2696365>
- Lundberg SM, Lee S-I (2017) A unified approach to interpreting model predictions. In: Guyon I, Von Luxburg U, Bengio S, Wallach H, Fergus R, Vishwanathan S, Garnett R (Hrsg) *Advances in neural information processing systems*. Curran Associates,
- Milosavlaskaya N, Tolstoy A (2016) Big data, fast data and data lake concepts. *Procedia Comput Sci* 88:300–305. <https://doi.org/10.1016/j.procs.2016.07.439>
- Natekin A, Knoll A (2013) Gradient boosting machines, a tutorial. *Front Neurobot* 7:21. <https://doi.org/10.3389/fnbot.2013.00021>
- Otter DW, Medina JR, Kalita JK (2021) A survey of the usages of deep learning for natural language processing. *IEEE Trans Neural Netw Learning Syst* 32:604–624. <https://doi.org/10.1109/TNNLS.2020.2979670>
- Phillips-Wren G (2012) AI tools in decision making support systems: a review. *Int J Artif Intell Tools* 21:1240005. <https://doi.org/10.1142/S0218213012400052>
- Schapire RE (2003) The boosting approach to machine learning: an overview. In: Denison DD, Hansen MH, Holmes CC, Mallick B, Yu B (Hrsg) *Nonlinear estimation and classification*. Springer, New York, S 149–171
- Schiedermaier I, Kick E, Baumgartner M, Kopp T, Kinkel S (2023). *Wissensmanagement in KMU*. ZWF 118(6):395–399. <https://doi.org/10.1515/zwf-2023-1087>
- Sokolova M, Lapalme G (2009) A systematic analysis of performance measures for classification tasks. *Inf Process Manag* 45:427–437. <https://doi.org/10.1016/j.ipm.2009.03.002>
- Sutton RS, Barto A (2018) *Reinforcement learning; An introduction*. MIT Press, Cambridge, London
- Wang RY, Strong DM (1996) Beyond accuracy: what data quality means to data consumers. *J Manag Inf Syst* 12:5–33. <https://doi.org/10.1080/07421222.1996.11518099>
- Würth R, Hipp J (2000) CRISP-DM: Towards a standard process model for data mining. In: *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, S 29–39
- Yoon Y, Guimaraes T, Swales G (1994) Integrating artificial neural networks with rule-based expert systems. *Decis Support Syst* 11:497–507. [https://doi.org/10.1016/0167-9236\(94\)90021-3](https://doi.org/10.1016/0167-9236(94)90021-3)
- Zheng A, Casari A (2018) *Feature engineering for machine learning; Principles and techniques for data scientists*. O'Reilly, Beijing