

Entwicklung einer Lernumgebung zu Maschinellern Lernen und algorithmischer Verzerrung für den Mathematikunterricht der Primarstufe

MELANIE PLATZ, SAARBRÜCKEN

Zusammenfassung: Im vorliegenden Beitrag wird eine mittels Design Science Research entwickelte Lernumgebung zu maschinellern Lernen und algorithmischer Verzerrung evaluiert. Dazu wird eine multiple Fallstudie mit qualitativen Erhebungs- und Auswertungsmethoden durchgeführt. Als Rahmen für die Unterstützung des didaktischen Designs wird auf Theory of Didactical Situations zurückgegriffen und mit Abstraction in Context kombiniert, um Lernprozesse auf mikroanalytischer Ebene untersuchen zu können. Die Ergebnisse zeigen, dass einzelne Lernende zentrale Aspekte algorithmischer Verzerrung konstruieren, dennoch sollte die Funktionsweise sowie die Güte der Ausgabe eines KI-Systems intensiver mit den Lernenden besprochen werden.

Abstract: This paper evaluates a learning environment for machine learning and algorithmic bias developed using Design Science Research. For this purpose, a multiple case study is conducted using qualitative survey and evaluation methods. Theory of Didactical Situations is used as a framework to support the didactic design combined with Abstraction in Context to analyse learning processes on a micro-analytical level. The results indicate that some learners construct central aspects of algorithmic bias; nevertheless, the functionality and the quality of the output of an AI system should be discussed more intensively with the learners.

1. Einleitung

In KMK (2024) wird das „Lernen über KI“ (S. 5) thematisiert, bei dem, neben einer grundlegenden informatischen Bildung, Wissen über KI sowie über ihre Wirkungsweisen und die Klärung ihrer ethischen und rechtlichen Rahmenbedingungen vermittelt werden soll. Dieser Aufbau digitaler Kompetenzen und ICT-Literacy (inkl. Grundkenntnisse über KI) sollte bereits in der Primarstufe stattfinden (SWK, 2024). Die KMK (2016) formuliert Kompetenzen für ein Leben in einer digitalen Welt, zu deren Entwicklung jedes einzelne Fach mit seinen spezifischen Zugängen zur digitalen Welt seinen Beitrag leisten soll, also auch Mathematik. Um eine Umsetzbarkeit der „neue[n] Gestaltung der Lehr- und Lernprozesse“ (KMK, 2016, S. 13) im Mathematikunterricht der Grundschule zu gewährleisten, müssen Konzepte

entwickelt werden, die im regulären Primarstufenunterricht in Verknüpfung mit traditionellen Unterrichtsthemen umgesetzt werden können. Die kindgerechte Analyse und Reflexion der Funktionsweise von KI-Systemen kann dazu beitragen, Kinder dabei zu unterstützen die Ausgaben von KI-Systemen kritisch zu betrachten und diese verantwortungsvoll zu nutzen.

Touretzky et al. (2019) halten fünf ‚Big Ideas of AI‘ fest, die KI-Bildung abdecken sollte. Eine davon ist ‚Computers can learn from data‘ und umfasst Maschinelles Lernen (ML). Eine weitere ‚Big Idea‘ lautet ‚AI applications can impact society in both positive and negative ways.‘ In ähnlicher Weise werden diese Ideen auch im ‚AI competency framework for students‘ der UNESCO (Miao & Shiohira, 2024) formuliert. Beide ‚Big Ideas‘ werden im vorliegenden Beitrag aufgegriffen. Damit die Lernenden nicht nur Konsumenten von KI, sondern auch Gestalter intelligenter Lösungen werden, muss die Anwendung von ML-Konzepten vermittelt werden, z. B. durch die Entwicklung von Bilderkennungsmodellen. Denn Lernende, die die Möglichkeiten, Stärken und Schwächen von ML praktisch erleben, erwerben eher ein tieferes Verständnis (Gresse von Wangenheim et al., 2021; Touretzky et al., 2019).

Sanusi et al. (2023) stellen in einem systematischen Literature Review zum Lehren und Lernen von ML in der K-12-Bildung u. a. fest, dass (1) zusätzliche ML-Ressourcen für den Kindergarten bis zur Mittelstufe benötigt werden – auch in einem aktuelleren systematischen Literature Review stellen Tan und Tang (2025) fest, dass sich die meisten empirischen Studien zur KI-Kompetenz auf die Sekundarstufe beziehen –, (2) weitere Studien darüber durchgeführt werden müssen, wie ML in andere Fachbereiche als die Informatik integriert werden kann, (3) es an Programmen zur beruflichen Weiterbildung von Lehrkräften mangelt, und (4) in Forschungsarbeiten mehr Belege für die gesellschaftlichen und ethischen Auswirkungen von ML berücksichtigt werden sollten. Trotz der wachsenden Besorgnis über algorithmische Verzerrungen gibt es einen Mangel an Forschung darüber, wie das Bewusstsein und Verständnis für diese Verzerrungen bei Kindern und Jugendlichen gefördert werden kann (Vartiainen et al., 2024). Der vorliegende Beitrag versucht genau an

diesen Stellen anzusetzen: (1) Durch die Entwicklung einer Lernumgebung für Viertklässler:innen sollen ML-Ressourcen für die Primarstufe entwickelt und erprobt werden. (2) Die Lernumgebung soll im regulären Mathematikunterricht der Primarstufe durchführbar sein, als Kontext wird die Klassifizierung von Dreiecken und Vierecken gewählt. Die Kinder sollen einen Klassifizierer trainieren selbst gezeichnete Dreiecke und Vierecke zu unterscheiden und werden zu ML-Modifikator:innen. (3) Die entwickelte Lernumgebung kann Lehrpersonen als Grundlage für eine Unterrichtsplanung zu ML dienen. (4) Insbesondere die algorithmische Verzerrung soll in der Lernumgebung in den Blick genommen werden.

2. Theoretischer Hintergrund

2.1 KI und Maschinelles Lernen (ML)

Künstliche Intelligenz (KI) zu definieren ist schwierig, da Definitionen eher schwammig bleiben und viele Systeme und Technologien (z. B. auch handelsübliche Taschenrechner) nach diesen Definitionen zu KI zählen müssten. Z. B. definiert Zweig (2019) ähnlich wie Rich (1983) KI als „[...] eine Software, mit deren Hilfe ein Computer eine kognitive Tätigkeit ausführt, die normalerweise Menschen erledigen.“ (S. 126). Um KI besser fassen zu können, schlagen Bender und Hanna (2025) vor eher von Automatisierung, statt von KI, zu sprechen und zu betrachten, was genau automatisiert wird.

Im vorliegenden Beitrag geht es um ein System, das die Klassifikation automatisiert, d. h. Eingaben sollen klassifiziert werden. Grundlage dafür ist das Maschinelle Lernen (ML) – ein Teilgebiet der KI –, es geht also um datenbasierte Systeme, bei denen die Regeln vorrangig durch einen (lernenden) Algorithmus generiert und durch eine automatisierte Datenanalyse entstanden sind (Eickhoff-Schachtebeck & Strecker, 2025). Beim ML lernen die Algorithmen anhand von (ausreichend) Daten verschiedene Zusammenhänge. ML folgt einem zyklischen, iterativen und probabilistischen Prozess (Shearer, 2000; Amershi et al., 2019). Der ML-Prozess beginnt mit der Datenakquisition und -exploration, bei der Rohdaten gesammelt und auf ihre Eignung für das Lernziel analysiert werden. In der Datenaufbereitungsphase werden diese zu strukturierten Trainingsdatensätzen transformiert, die als primäre Lernquelle für Algorithmen dienen (Provost & Fawcett, 2013). Das Modelltraining erfolgt durch iterative Optimierung, wobei Algorithmen Hypothesen über Datenstrukturen bilden und diese gegen Validierungsdaten testen (Bishop, 2006). Dieser Prozess ist inhärent probabilistisch: Modelle treffen Vorhersagen mit quantifizierbaren

Unsicherheiten und lernen aus Fehlern durch Rückkopplungsmechanismen (Murphy, 2012). Das somit entstandene Modell kann anschließend auf neue, unbekannte Daten derselben Art angewendet werden (Botsch, 2023). Ein zentrales Merkmal des ML-Prozesses sind die Feedback-Schleifen zwischen verschiedenen Phasen (Sculley et al., 2015). Evaluationsergebnisse fließen zurück in die Modellarchitektur und Hyperparameter-Optimierung. Im produktiven Einsatz generierte Vorhersagen können als neue Trainingsdaten zur kontinuierlichen Modellverbesserung verwendet werden (Settles, 2009; Hoi et al., 2021). Die Modellbewertung erfolgt mehrdimensional durch verschiedene Metriken und berücksichtigt sowohl statistische Leistung als auch praktische Anwendbarkeit (Flach, 2019). Besonders in produktiven Systemen erfordert dies kontinuierliches Monitoring und Re-Training, da sich Datenverteilungen über die Zeit ändern können (Concept Drift) (Gama et al., 2014).

In der vorliegenden Lernumgebung soll das System lernen, Vierecke und Dreiecke unterscheiden. Dafür wird das überwachte Lernen als Lernverfahren angewandt (Abb. 1).

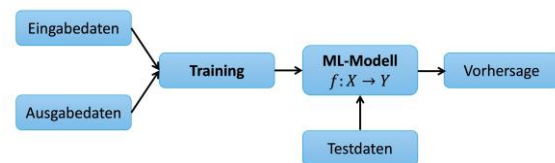


Abb. 1: Überwachtes Lernen (nach Botsch, 2023, S. 3)

Überwachtes Lernen (Supervised Learning) bildet eine der drei Hauptkategorien des ML und zeichnet sich durch die Verwendung gelabelter Trainingsdaten aus (Russell & Norvig, 2020). Überwachtes Lernen nutzt Datensätze, bei denen sowohl die Eingabefeatures (aus X) als auch die korrekte Zielvariable (aus Y) bekannt sind (Hastie et al., 2009). Formal ausgedrückt besteht ein gelabelter Trainingsdatensatz aus Paaren $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$, wobei $x_i \in X$ den Eingabevektor und $y_i \in Y$ die zugehörige korrekte Ausgabe darstellt (Bishop, 2006; Domingos, 2012). Das Ziel ist es, eine Funktion $f: X \rightarrow Y$ zu erlernen, die bei neuen, unbekannten Eingaben (in unserem Fall Zeichnungen von Dreiecken und Vierecken) zuverlässige Vorhersagen treffen kann. Der gewünschte Funktionswert aus Y ist in unserem Fall eine Klasse (Dreieck oder Viereck). Konkret sollen Viertklässler:innen ein Modell trainieren, das gezeichnete Dreiecke und Vierecke klassifiziert. Unsere Trainingsdaten sind also selbst gezeichnete Drei- und Vierecke und deren Klassifikation. Es handelt sich dabei – zwecks didaktischer Reduktion

– um einen sehr kleinen Datensatz. In der Realität werden KI-Modelle mit wesentlich größeren Datensätzen trainiert.

2.2 Klassifizierer-Training in der Primarstufe

Grundschul- und sogar Kindergartenkinder sind durchaus in der Lage, grundlegende Konzepte des Klassifizierertrainings zu verstehen, insbesondere, wenn visuelle und spielerische Ansätze verwendet werden. Z. B. die Studie von Vartiainen et al. (2020) zeigte, dass Kinder (3-9 Jahre) ML verstehen, erkunden und erklären können. Die Kinder lernten, indem sie einen Computer lehrten (Papert, 1980): sie trainierten ein ML-Modell mit emotionalen Ausdrücken. Vartiainen et al. (2020) stellten außerdem fest, dass die sofortige Reaktion des Systems die Exploration fördert und dass das Peer Teaching in diesem Kontext gut funktionierte. In einer Studie von Dwivedi et al. (2021) trainierten 7-13-jährige Bildklassifikatoren und testeten gegenseitig die Robustheit ihrer Modelle. Folgende Gestaltungskriterien für ML-Erfahrungen für Kinder konnten abgeleitet werden: (1) Die Sichtbarmachung und Dynamisierung von Metriken wie Konfidenzwerten kann Experimente, Schlussfolgerungen und Diskussionen begünstigen. (2) Die Ermöglichung von Partneraktivitäten, bei denen Kinder Trainingssätze und Strategien mit anderen vergleichen können, kann das Nachdenken und Erkennen von Mustern zur Verbesserung von Modellen unterstützen. (3) Der Einsatz von für Kinder zugänglichen Modalitäten (z. B. Bilder) in ML-Aktivitäten kann die Mustererkennung fördern und eine schnelle Datenüberprüfung und Trainingsanpassung ermöglichen.

Da die Objekterkennung derzeit eine der typischsten ML-Aufgaben ist, unterstützen viele Tools die Entwicklung von ML-Modellen für Erkennungsaufgaben, vor allem für die Bilderkennung mit Nutzung von Deep-Learning-Modellen wie TensorFlow (Gresse von Wangenheim et al., 2021). Im vorliegenden Fall wird mit Machine Learning for Kids (2017) gearbeitet, dessen Grundlage darin besteht, Kindern im Schulalter die Werkzeuge an die Hand zu geben, mit denen sie ihre eigenen Modelle für ML erstellen können, und sie in kinderfreundliche Plattformen wie Scratch (eine blockbasierte visuelle Programmiersprache inklusive ihrer Entwicklungsumgebung und Online-Community-Plattform) zu integrieren, damit sie Projekte mit ihren Modellen erstellen können (Lane, 2024). Die von Dwivedi et al. (2021) identifizierten Gestaltungskriterien für ML-Erfahrungen für Kinder können mit Machine Learning for Kids (2017) umgesetzt werden. Der Klassifizierer ist

vollständig webbasiert und wurde speziell für den Einsatz im Unterricht an Schulen entwickelt. Über eine Administrator-Seite kann die Lehrperson Accounts für die Lernenden anlegen und deren Zugang verwalten. Persönliche Daten der Lernenden werden dabei nicht erhoben. Für die Bilderkennung wird aktuell auf TensorFlow.js, eine Bibliothek für ML in JavaScript, zurückgegriffen (Lane, 2024). TensorFlow arbeitet mit Tensoren, das sind mehrdimensionale Arrays (oder Vektoren), die Bilddaten repräsentieren. Ein Farbbild wird typischerweise als 3D-Tensor gespeichert (Höhe×Breite×Farbkanäle). Für Bilderkennung verwendet TensorFlow hauptsächlich Convolutional Neural Networks (CNNs) (Krizhevsky et al., 2012), die speziell für visuelle Daten entwickelt wurden. TensorFlow bzw. CNNs analysieren Bilder statistisch durch Pixelmuster. Das Netzwerk lernt, bestimmte Kombinationen von Pixelwerten mit Labels zu verknüpfen, ohne dabei geometrische Konzepte zu ‚verstehen‘. Menschliche Wahrnehmung arbeitet hingegen mit abstrakten geometrischen Konzepten. Menschen erkennen z. B. ein Dreieck unabhängig von Größe, Orientierung oder Zeichenstil, wir sind robust gegenüber Variationen. CNNs sind dagegen anfällig für ‚adversarial examples‘ – winzige, für Menschen unsichtbare Pixelveränderungen können bei CNNs zu einer falschen Klassifizierung führen (Wiyatno et al., 2019).

Die Qualität überwachter Lernmodelle hängt also kritisch von der Repräsentativität und Qualität der gelabelten Trainingsdaten ab (Provost & Fawcett, 2013). Probleme wie Selection Bias (wenn die Trainingsdaten nicht repräsentativ für die Grundgesamtheit sind, auf die das Modell später angewendet werden soll), Label Noise (bei fehlerhaften oder inkonsistenten Bezeichnungen in den Trainingsdaten) oder Concept Drift (die Veränderung der statistischen Eigenschaften der Zielvariable über die Zeit, wodurch die erlernten Zusammenhänge ungültig werden) können die Generalisierungsfähigkeit erheblich beeinträchtigen (Quionero-Candela et al., 2009).

2.3 Klassifizieren von Dreiecken und Vierecken in der Primarstufe

Ein empirisch fundiertes Modell zur Entwicklung des geometrischen Denkens wurde 1957 von Pierre und Diana van Hiele-Geldorf erarbeitet. Zahlreiche Untersuchungen zeigen, dass das Modell eingesetzt werden kann, um den Begriffserwerb und das geometrische Denken von Kindern zu beschreiben. Das Modell besteht aus fünf Denkebenen, die sich qualitativ unterscheiden lassen (Franke & Reinhold,

2016). Für den vorliegenden Beitrag sind insbesondere die Stufen (1) bis (3) relevant: (1) Räumlichanschauungsgebundenes Denken: Ein zu identifizierendes geometrisches Objekt wird ganzheitlich erfasst und noch nicht anhand einzelner Eigenschaften erkannt. Kinder können auf dieser Stufe z. B. verschiedene Viereckformen erkennen und benennen, sie sind aber noch nicht in der Lage, diese Zuordnung über Eigenschaftszuweisungen zu begründen. Figuren werden aufgrund der Ähnlichkeit mit früher betrachteten Vierecken identifiziert. (2) Geometrisch-analysierendes Denken: Eine erste Analyse der geometrischen Objekte kann durch die Kinder erfolgen, dabei richtet sich die Aufmerksamkeit erstmals gezielt auf die geometrischen Eigenschaften der Objekte. Die Kinder erkennen Eigenschaften, beschreiben diese und können auch begründete Klassifizierungen vornehmen. Allerdings sind die Klasseninklusionen zwischen Figuren (z.B. zwischen Rechteck und Quadrat) noch nicht einsehbar. (3) Geometrisch-abstrahierendes Denken: Kinder können Beziehungen zwischen den Eigenschaften einer Figur und den Eigenschaften verwandter Figuren feststellen und diese Erkenntnisse in Begründungen einbeziehen. Es werden nicht nur Klassifikationen, sondern nun auch Klasseninklusionen verstanden (Franke & Reinhold, 2016, S. 136f).

Das Definieren und Klassifizieren von Vierecken ist für viele Lernende ein schwieriges Thema (Currie & Pegg, 1998; de Villiers, 1994; Erez & Yerushalmy, 2006; Monaghan, 2000; Pickreign, 2007). Die Gründe für diese Schwierigkeiten liegen in der Komplexität zu lernen die Eigenschaften verschiedener Vierecke zu analysieren und zwischen kritischen und nicht kritischen Aspekten zu unterscheiden. Es zeigt sich, dass Lernende häufig Schwierigkeiten mit den Definitionen von Formen haben und dass ihr geometrisches Denken oft erheblich von ihren mentalen Vorstellungen von Formen beeinflusst wird (Fujita & Jones, 2007; Erez & Yerushalmy, 2006). Das Identifizieren gelingt Kindern am leichtesten, wenn die Figur dem Prototyp möglichst nahekommt und – da alle Quadrate ähnlich sind, bereitet das Erkennen von Quadraten kaum Probleme. Schwierigkeiten treten lediglich durch die Lage der Figur auf dem Arbeitsblatt auf. Rechtecke werden am leichtesten als solche erkannt, wenn das Seitenverhältnis bei 1:2 liegt und die langen Seiten parallel zu den Blattkanten verlaufen oder wenn das Rechteck näherungsweise den DIN-Papierformaten entspricht (Franke & Reinhold, 2016; Monaghan, 2000). Monaghan (2000) stellte in einer Studie mit 11-Jährigen fest, dass es diesen schwerfällt zu akzeptieren, dass ein

Quadrat eine besondere Art von Rechteck ist. Das Verständnis hierarchischer Beziehungen umfasst mindestens: Die Fähigkeit, eine Form auf verschiedene Weise zu klassifizieren und mit verschiedenen Namen zu bezeichnen (z. B., dass eine Raute auch als Viereck, eine besondere Art von Parallelogramm oder Drachen usw. bezeichnet werden kann); Die Notwendigkeit, die transitiven Beziehungen zwischen den Begriffen der Formen zu verstehen (z. B., dass, wenn ein Quadrat eine Raute ist und eine Raute ein Parallelogramm ist, dann ist ein Quadrat auch ein Parallelogramm); Die Notwendigkeit, die Asymmetrie der Beziehungen zwischen Vierecken zu verstehen (z. B., dass jedes Rechteck ein Parallelogramm ist, aber nicht jedes Parallelogramm ein Rechteck ist); Die Notwendigkeit, die entgegengesetzte Asymmetrie und die transitiven Beziehungen der kritischen Attribute von Formkonzepten zu verstehen (z. B., dass die Attribute des Rechtecks in den Attributen des Quadrats enthalten sind, die Attribute des Quadrats jedoch nicht in denen des Rechtecks.) Damit Lernen effektiv ist, sind deduktive Schlussfolgerungen sowie eine angemessene Interaktion zwischen Konzepten und Bildern erforderlich (Fujita & Jones, 2007). Beim Verständnis der Beziehungen zwischen Vierecken gibt es möglicherweise eine hierarchische Reihenfolge der Schwierigkeitsgrade. So werden Rauten z. B. eher als besondere Art von Parallelogrammen erkannt als Quadrate oder Rechtecke, da sie Parallelogrammen ‚ähneln‘ (Fujita & Jones, 2007).

Wie bei den Vierecken spielt für Kinder auch bei den Dreiecken häufig die Lage auf dem Blatt eine wichtige Rolle: „Die Grundseite muss immer unten sein.“ (Pielsticker, 2020, S. 232). Bei der Aufgabe verschiedene Dreiecke zu zeichnen, stellten Burger & Shaughnessy (1986) fest, dass Kinder Dreiecke häufig auf Basis irrelevanter Attribute wie die Lage auf dem Blatt oder durch unsauber gezeichnete Linien unterscheiden. Wie im vorhergehenden Abschnitt 2.2 erwähnt, spielt aber genau das eine wichtige Rolle für TensorFlow bzw. CNNs.

2.4 Algorithmische Verzerrung in der Primarstufe

Algorithmische Prozesse, die zur Automatisierung oder Unterstützung von Entscheidungen über Menschen eingesetzt werden, können diskriminierende Ergebnisse hervorbringen, die gegen die Normen von Gerechtigkeit und Gleichheit verstoßen und sich nachteilig auf bestimmte Personen oder Gemeinschaften am Arbeitsplatz oder in der Gesellschaft auswirken. Dieses Phänomen wird als algorithmische Verzerrung bezeichnet und tritt auf, wenn die

Ergebnisse eines Algorithmus bestimmte Einzelpersonen oder Gruppen stärker begünstigen oder benachteiligen als andere, ohne dass es einen gerechtfertigten Grund für eine solche ungleiche Behandlung gibt (Kordzadeh & Ghasemaghaei, 2022).

Vartiainen et al. (2024) führten Workshops mit Lernenden der 4. und 7. Klasse zu den möglichen Ursachen algorithmischer Verzerrungen innerhalb generativer Text-zu-Bild-Modelle durch. Dabei wurden statistisch signifikante Fortschritte bei den datengestützten Erklärungen der Kinder beobachtet. Verbesserungen in den Erklärungen von Kindern können durch praktische Lernerfahrungen (Osterhaus et al., 2021) und durch kreatives Gestalten (Kafai et al., 2019) gefördert werden, bei denen Kinder angeleitet werden, gemeinsam einige Schlüsselkonzepte der Informatik auf kontextbezogene Weise zu erforschen und zu erklären (Vartiainen et al., 2020).

In der vorliegenden Lernumgebung bearbeiten Kinder das Problem einen Klassifizierer zu trainieren, wobei zunächst Trainingsdaten verwendet werden, die absichtlich ‚unfair‘ sind (siehe Abschnitt 4.1), um so die Idee der algorithmischen Verzerrung aufgreifen zu können. Bei der Problemlösung machen sie praktische Lernerfahrungen, können kreativ tätig werden und werden zum Argumentieren angeregt.

3. Forschungsdesign

3.1 Design Science Research (DSR)

Im vorliegenden Beitrag wird Design Science Research (DSR; Drechsler & Hevner, 2016) angewandt, um eine Lernumgebung zu ML und algorithmischer Verzerrung bei der Klassifizierung von Dreiecken und Vierecken zu entwickeln. DSR ist ein in der Philosophie des Pragmatismus verwurzeltes Paradigma, das seinen Ursprung in den Wissenschaften und der Konstruktion des Künstlichen („The Sciences of the Artificial“: Simon, 1970) hat. DSR bringt nützliche Artefakte durch problemlösende Forschung zur Beantwortung von Forschungsfragen in Bezug auf menschliche Probleme hervor (Drechsler & Hevner, 2016). Wittmann (1995) hält mit Bezug auf Simon (1970) fest, dass sich der Kern des Mathematikunterrichts auf die Konstruktion künstlicher Objekte wie Unterrichtseinheiten konzentriert.

Es gibt viele DSR-Modelle, die in verschiedenen Disziplinen entwickelt wurden und Anwendung finden. Im vorliegenden Beitrag werden Modelle aus den Bereichen ‚Education‘ und ‚Information Systems‘ herangezogen, um Synergieeffekte dieser Modelle zu nutzen. Die Modelle im Bereich ‚Education‘ haben

alle die Betonung von Zyklen oder Iterationen gemein (Bakker, 2018, S. 59), wobei die Terminologie zur Beschreibung der verschiedenen Phasen innerhalb solcher Zyklen unterschiedlich ist. Dennoch scheinen folgenden Phasen (oder Arbeitsbereiche, vgl. Prediger et al., 2012) jedes so genannten Makrozyklus der Designforschung (Bakker, 2018, S. 59, vgl. auch Gravemeijer & Prediger, 2019, S. 35) bei allen Designforschungsprojekten recht einheitlich zu sein: (1) Vorbereitung und Design, (2) Durchführung und (3) Analyse und Re-Design. Auch in den Modellen aus dem Bereich ‚Information Systems‘ sind diese wiederzufinden (z. B. Peffers et al., 2007).

DSR ist ein Problemlösungsprozess und das Verständnis für ein Designproblem und dessen Lösung können durch die Entwicklung und Anwendung eines Artefakts erworben werden (Hevner et al., 2004). D. h., DSR erfordert die Entwicklung eines innovativen, zweckmäßigen Artefakts (Leitlinie 1) für einen bestimmten Problembereich (Leitlinie 2). Da das Artefakt zweckmäßig sein soll, muss es für das angegebene Problem einen Nutzen erbringen. Daher ist eine gründliche Bewertung des Artefakts von entscheidender Bedeutung (Leitlinie 3). Ebenso entscheidend ist die Neuheit, da das Artefakt innovativ sein muss und ein bisher ungelöstes Problem oder ein bekanntes Problem auf effektivere oder effizientere Weise lösen muss (Leitlinie 4). Das Artefakt selbst muss klar definiert, formal dargestellt, kohärent und in sich konsistent sein (Leitlinie 5). Der Prozess, durch den es entsteht, und oft auch das Artefakt selbst, beinhaltet oder ermöglicht einen Suchprozess, durch den ein Problemraum konstruiert und ein Mechanismus vorgeschlagen oder umgesetzt wird, um eine effektive Lösung zu finden (Leitlinie 6). Schließlich müssen die Ergebnisse der DSR didaktischem Publikum (Wissenschaftler:innen, die sie weiterentwickeln werden, und Praktiker:innen, die sie umsetzen werden) effektiv vermittelt werden (Leitlinie 7) (Hevner et al., 2004). Diese sieben Leitlinien wurden festgelegt, um dabei zu unterstützen, die Anforderungen an eine effektive DSR zu verstehen. Dabei sollte jede dieser Richtlinien berücksichtigt werden, damit die DSR vollständig ist (Hevner et al., 2004, S. 82).

3.2 Forschungsfragen

Für die Lernumgebungsentwicklung ist eine Verknüpfung von Mathematikunterricht und informatischer Bildung notwendig. Kenntnisse in der Mathematik werden vertieft, während in der Informatik neue Kenntnisse erworben werden. Es werden folgende Forschungsfragen adressiert:

(FF1) Welche epistemischen Aktionen nach Abstraction in Context (siehe Abschnitt 3.4) sind bei Lernenden bei der Erprobung der Lernumgebung erkennbar?

(FF2) Welche Hürden können rekonstruiert werden?

Eine multiple Fallstudie (Yin, 2018) mit qualitativen Erhebungs- und Auswertungsmethoden wird durchgeführt. Als Rahmen für die Unterstützung des didaktischen Designs wird auf die Theory of Didactical Situations (Artigue et al., 2014) zurückgegriffen und mit Abstraction in Context (Dreyfus & Kidron, 2014) kombiniert, um Lernprozesse auf mikroanalytischer Ebene untersuchen zu können.

3.3 Theory of Didactical Situations (TDS)

Die Theory of Didactical Situations (TDS) begann sich in den 1960er Jahren in Frankreich zu entwickeln, initiiert von Guy Brousseau, der sie seither federführend weiterentwickelt hat. TDS bietet einen systemischen Rahmen für die Untersuchung von Lehr- und Lernprozessen in der Mathematik und für die Unterstützung der didaktischen Gestaltung (Artigue et al., 2014). TDS verfolgt die Ansicht, dass die Bedingungen für die Entstehung des angestrebten mathematischen Wissens aus der Interaktion der Schüler:innen mit dem Umfeld als optimale Lösung für ein mathematisches Problem geklärt und geschaffen werden müssen. Das Milieu ist das System, mit dem die Lernenden in einer a-didaktischen Situation (die Vorsilbe ‚a‘ weist darauf hin, dass die Situation zeitlich von ihrer didaktischen Intentionalität befreit wurde) interagieren. Eine wesentliche Aufgabe der Lehrperson ist es, dieses Milieu zu organisieren (Artigue et al., 2014).

Folgende Bedingungen sollten erfüllt sein, damit eine Situation als a-didaktisch erlebt werden kann (Brousseau, 2008, S. 249; Bessot, 2024, S. 9): Das angestrebte mathematische Wissen sollte die einzige gute Methode zur Lösung der Aufgabe sein. Der Lernende kann eine Antwort in Betracht ziehen, aber diese anfängliche Antwort (Grundverfahren, das sich auf früheres Wissen und Kenntnisse bezieht) ist nicht die Antwort, die gelehrt werden soll: Wenn die Antwort bereits bekannt wäre, wäre dies keine Lernsituation. Das Grundverfahren muss sich schnell als unzureichend oder ineffektiv erweisen, sodass die Lernenden gezwungen sind, Anpassungen/Änderungen an ihrem Wissenssystem vorzunehmen. Die Lernenden können also mit einem unzureichenden ‚Grundwissen‘ an die Arbeit gehen. Es besteht eine Unsicherheit der Lernenden hinsichtlich der zu treffenden Entscheidungen. Sie können eine rasche

Reihe von Versuchen nach dem Prinzip ‚Versuch und Irrtum‘ durchführen, sie können also selbst feststellen, ob ihr Versuch gelungen oder gescheitert ist. Ohne die Lösung zu bestimmen, sind diese Überprüfungen suggestiv (sie begünstigen einige Hypothesen, bringen einige geeignete Informationen ein, weder zu offen noch zu geschlossen). Das angestrebte Wissen ist a priori erforderlich, um von der grundlegenden zur optimalen Strategie zu gelangen. Die Lösung kann von einem Teil der Lernenden in angemessener Zeit gefunden und getestet werden und von den anderen rasch weitergegeben und überprüft werden, d. h. es gibt ein Milieu für die Validierung: Das Milieu ermöglicht Rückkopplungen. Die Situation kann wiederverwendet werden und liefert dann einige Fragen, die den gesamten Prozess wieder in Gang setzen. Diese Bedingungen ermöglichen ein hohes Maß an Autonomie der Lernenden. Natürlich beschreiben sie ein Ideal und werden von realen Szenarien selten erfüllt. Sie stellen eine theoretische Referenz für Forschende dar, die ihnen hilft, das a-didaktische Potenzial eines gegebenen Szenarios und dessen Grenzen feststellen zu können und die Kontingenz der Realisierungen im Klassenzimmer besser zu verstehen (Artigue et al., 2014).

Bei der TDS wird eine a-priori-Analyse (d. h. die Analyse dessen, was möglich war) einer Situation durchgeführt und mit der a-posteriori-Analyse (d. h. der Analyse dessen, was tatsächlich geschehen ist) verglichen. Eine a-priori-Analyse erfolgt in Abschnitt 4, die a-posteriori-Analyse in Abschnitt 5. Leitfragen können dabei sein (Mangiante-Orsola et al., 2018): Was ist die didaktische Absicht der Lehrperson (das mathematische Wissen, das er/sie den Lernenden vermitteln will)? Können wir das objektive Milieu identifizieren, das den Lernenden zur Verfügung steht? (Unter objektivem Milieu werden alle Daten verstanden, die unabhängig von den Interventionen der Lehrperson und vom Wissen der Lernenden sind und die für das Handeln oder die Reflexion der Lernenden zur Verfügung stehen.) Gibt es in diesem Milieu etwas, das für die Lernenden problematisch ist? Wie können die Lernenden dieses Problem lösen? Welches Wissen steht für die Lernenden im Vordergrund? Welcher Einsatz von Wissen ist notwendig, um mit dem Milieu zu interagieren und das Problem zu lösen? (Wird es benötigt, um: Fortschritte bei der Lösung des Problems zu erzielen? Um die Lösung so zu formulieren, dass jemand anders das Problem lösen kann? Um zu zeigen, dass diese Lösung ein guter Weg ist, um das Problem zu lösen?) Welchen Status hat dieses Wissen für die Lernenden (ganz neues Wissen, Wissen im Laufe des Lernens, vermeintlich

bekanntes Wissen)? Bei dieser Frage geht es um die Beziehungen zwischen dem Wissen, um das es geht (neu oder alt), und dem didaktischen Vertrag (was von der Lehrperson und von den Lernenden erwartet wird). Welche Möglichkeiten hat der Lehrer, das Umfeld so zu verändern, dass sich das Wissen, das für die Lernenden auf dem Plan steht, verändert (didaktische Variablen)? Um Lernprozesse darüber hinaus auf mikroanalytischer Ebene untersuchen zu können, die (für den Lernenden) zu neuen Konstrukten führen, wird auf Abstraction in Context (AiC) zurückgegriffen.

3.4 Abstraction in Context (AiC)

AiC stellt ein Modell verschachtelter epistemischer Aktionen bereit. Das zentrale theoretische Element von AiC ist das dynamisch verschachtelte epistemische Aktionsmodell (oder RBC-Modell), ein theoretisch-methodisches Modell, nach dem die Entstehung eines neuen Konstrukts anhand von drei beobachtbaren epistemischen Aktionen beschrieben und analysiert wird (Dreyfus, Hershkowitz & Schwarz, 2015, S. 188): (1) Identifizieren situationsrelevanter Wissenskonstrukte (R–Recognizing), (2) kombinieren bekannter Konstrukte zur Lösung eines lokalen Problems (B–Building-with) und (3) entwickeln neuer Konstrukte (C–Constructing) (Schnell, 2013, S. 78). Es können auch Partially Correct Constructs (PaCC; Ron et al, 2010) auftreten. PaCC zeichnen sich dadurch aus, dass sie nur teilweise mit mathematisch tragfähigen Konstrukten übereinstimmen. Sie werden z. B. genutzt, um individuelle Lernwege zu Beschreiben und um inkonsistente Antworten im Verlauf von Lerneinheiten zu erklären (Schnell, 2013). Auch bei AiC wird eine a-priori und a-posteriori-Analyse durchgeführt (siehe Abschnitte 4 und 5).

3.5 Synergieeffekte von AiC und TDS

AiC fokussiert den Lernenden und seine kognitive Entwicklung, TDS hingegen didaktische Systeme. Die beiden theoretischen Ansätze sind sensibel für Fragen des Kontexts, aber aufgrund der unterschiedlichen Schwerpunkte wird dieser nicht auf dieselbe Weise theoretisiert und behandelt (Kidron et al., 2014). In Bezug auf die Rolle der Lehrperson könnten die TDS-Forschenden z. B. fragen, welches Milieu die Lehrperson den Lernenden zur Verfügung stellt und wie sie dessen Entwicklung steuert, um eine sinnvolle Verbindung mit dem angestrebten mathematischen Wissen herzustellen. AiC-Forschende könnten fragen, wie die Intervention der Lehrperson den Konstruktionsprozess der

Lernenden beeinflusst und wie dieser durch die epistemischen Aktionen des RBC-Modells beschrieben werden kann. Die beiden Theorien weisen einige Gemeinsamkeiten bzgl. der zentralen Rolle, die dem Aufbau von mathematischem Wissen in der Analyse zugewiesen wird, auf. TDS sucht z. B. nach Bedingungen dafür, dass ein Lernprozess nicht auf eine einfache Kopie von zuvor erarbeiteten Antworten reduziert wird. Auch in AiC wird Abstraktion als eine Aktivität der vertikalen Reorganisation vorheriger mathematischer Konstrukte innerhalb der Mathematik und mit mathematischen Mitteln, die zu einem für den Lernenden neuen Konstrukt führen, definiert (Kidron et al., 2014). Dennoch gibt es einen entscheidenden Unterschied: Bei AiC liegt der Schwerpunkt auf der Art und Weise, wie die Lernenden ihre Antwort, ihre Strategie oder ihr ‚Denken‘ entwickeln; bei TDS liegt er auf den Bedingungen, die diese Entwicklung ermöglichen, und den Einschränkungen, die andere mögliche Antworten oder Strategien behindern (Kidron et al., 2014). Im vorliegenden Beitrag werden die Stärken beider Theorien verknüpft: Während AiC eine differenzierte Analyse der Konstruktionsprozesse der Lernenden durchführen kann, sind TDS-Instrumente wirksamer, wenn die Lehrer:innen-Schüler:innen-Interaktion untersucht werden soll.

4. A-priori-Analyse

4.1 Planung und Einbettung

Die Lernumgebung ist für einen Zeitumfang von 45 Minuten geplant: 10 min Einführung, 25 min Arbeit mit dem Klassifizierer in Kleingruppen (2-3 Kinder) und 10 min Reflexion. Es soll ein Modell trainiert werden, das gezeichnete Dreiecke und Vierecke klassifiziert. Für diese Aufgabe in der Lernumgebung wurde ein Klassifizierer, der Vierecke und Dreiecke automatisch unterscheiden soll, von der Lehrperson (Autorin des Beitrags) vortrainiert. Die Trainingsdaten wurden bewusst so ausgewählt, um hinterher bestimmte Effekte (Verzerrung) zu erhalten.

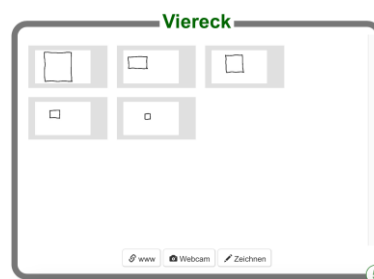


Abb. 2: Datensatz Vierecke: drei Quadrate und zwei Rechtecke, längste Seite parallel zur unteren Blattkante

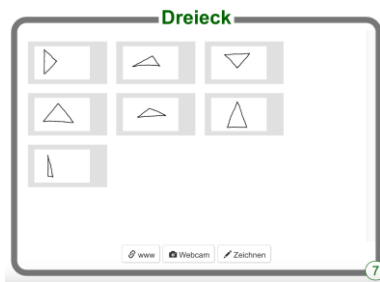


Abb. 3: Datensatz Dreiecke: fünf annähernd gleichschenklige Dreiecke, drei davon mit der längsten Seite parallel zur unteren, eins zur linken und eins zur oberen Blattkante mit der Spitze unten. Zwei annähernd rechtwinklige Dreiecke, eins mit der längsten Seite parallel zur unteren Blattkante, eins mit den kurzen Seiten (Katheten) parallel zur unteren und linken Blattkante

In der Einführung erklärt die Lehrperson, wie das Tool zu bedienen ist. Dabei werden die Kinder zu jedem der beiden Datensätze (Viereck und Dreieck) gefragt: Wie viele Bilder sind enthalten? Wie ähnlich sind die Bilder? Inwiefern sind die Bilder unterschiedlich? (Beantwortet werden diese Fragen in Abschnitt 4.2.) Anschließend wird gezeigt, wie ‚gut‘ der Klassifizierer mit diesem Datensatz funktioniert (siehe Tab. 1).

Bild	Klassifikation (Dreieck/ Viereck)	Zuverlässigkeitsgrad	Richtig?
	Viereck	93%	ja
	Viereck	100%	ja
	Dreieck	95%	nein
	Viereck	76%	ja
	Dreieck	96%	ja
	Dreieck	99%	ja
	Dreieck	94%	nein

Tab. 1: Qualität des Klassifizierers. Getestet wurden: Rechteck, Quadrat, Trapez, Rechteck mit der kurzen Seite parallel zur unteren Blattkante, Seitenverhältnis 1:5; gleichschenkliges Dreieck mit der längsten Seite parallel zur rechten Blattkante, stumpfwinkliges Dreieck und ein Drache (von den Lernenden eher als Raute identifiziert)

Die Lehrperson erklärt im Detail wie die Tabelle entstanden ist. Die Kinder werden gefragt: Bei welcher Klasse hat der Klassifizierer besser funktioniert? Bei den Dreiecken oder den Vierecken? Woran liegt das eurer Meinung nach? (Auf diese Fragen wird in Abschnitt 4.2 eingegangen.) Dabei stellt die Lehrperson auch einen Bezug zwischen den Trainings- und Testdaten her. Es kann festgehalten werden: Der Klassifizierer ist zugunsten der Dreiecke verzerrt, zum Nachteil für Vierecke. Die Trainingsdaten sind ‚unfair‘ und spielen eine wichtige Rolle in einem KI-System.

Anschließend wird ein Bezug zur Lebenswelt hergestellt: Beispielsweise hat die MIT-Forscherin Joy Buolamwini in einem Forschungsprojekt gezeigt, dass die Gesichtserkennungsprodukte von populären Herstellern wesentlich schlechter darin sind, das Geschlecht einer Person zu bestimmen, wenn es sich um Frauen handelt, vor allem Frauen mit dunkler Haut (Buolamwini, 2017). Auch auf KI-gestützte Bewerbungssysteme wird verwiesen, die durch algorithmische Verzerrung unfair gegenüber bestimmten Bewerbergruppen wurden (Wilke, 2018).

In Kleingruppen erhalten die Kinder nun den Auftrag: Versucht den Klassifizierer zu optimieren, indem ihr neue Trainingsdaten – Vierecke und Dreiecke – hinzufügt. Jede Gruppe erhält dafür ein Tablet mit Stift, mit dem an einem eigenen webbasierten Projekt gearbeitet werden kann, sowie eine Infoseite zur Bedienung des Tools (s. Anhang 1) und ein Arbeitsblatt (s. Anhang 2), auf dem jeweils zu den neuen Datensätzen gefragt wird: Wie viele Bilder sind enthalten? Wie ähnlich sind die Bilder? Inwiefern sind die Bilder unterschiedlich? Ihren Klassifizierer können die Kinder über ‚Lernen & Testen‘ und ‚Trainiere ein neues maschinelles Lernmodell‘ trainieren und ihren Datensatz mit den Bildern in der Tabelle (siehe Tab. 1) testen. (Jedes Bild wird einzeln getestet, indem man es mit ‚Test durch Zeichnen‘ abzeichnet, es dürfen auch neue Bilder getestet werden). Wie in Tab. 1 wird die Qualität des Klassifizierers in der Tabelle auf dem Arbeitsblatt (Anhang 2) festgehalten.

In der Reflexion wird gefragt: Bei welcher Klasse hat der Klassifizierer nun besser funktioniert? Bei den Dreiecken oder den Vierecken? Woran liegt das eurer Meinung nach? Was würdet ihr im nächsten Schritt optimieren?

4.2 Didaktische Intention

In der Lernumgebung geht es darum, dass Kinder konzeptionelles Wissen darüber entwickeln, wie KI auf der Grundlage von Daten und Algorithmen trainiert wird. Mittels einer beispielbasierten Abstraktion wird Wissen darüber vermittelt, wie Modelle für ML mithilfe von Daten und Algorithmen trainiert werden. Das überwachte Lernen wird angewandt und es wird erklärt, wie Daten erfasst und gekennzeichnet werden und welche Rolle Trainingsdaten in einem KI-System spielen. Außerdem wird die Behauptung entkräftet, dass KI die Programmierung von Algorithmen automatisieren wird und dass Menschen nichts über Algorithmen lernen müssen (Miao & Shiohira, 2024). Es geht also darum, dass die

Kinder den Begriff ‚Optimierung‘ kennen lernen und erkennen, dass der Mensch über die Ziele der von ihm geschaffenen soziotechnischen Systeme entscheidet (Payne, 2019). Folgende Computational Thinking Praktiken werden angeregt (Shamir & Levin, 2022): Die Daten werden in Trainings- und Testdaten zerlegt. Bezogen auf die Mustererkennung müssen Daten gefiltert werden und Mutmaßungen über die Ausgabe des Systems getroffen werden. Zudem spielt die Abstraktion eine wichtige Rolle: Damit ein ML-Modell die Daten korrekt klassifizieren kann, dürfen nur diejenigen Eingabedaten ausgewählt werden, die die für die Lösung des Problems relevanten Informationen enthalten.

In der Mathematik geht es darum, Gesetzmäßigkeiten zu erkennen und zu beschreiben, Daten zu erfassen und zu strukturieren und Darstellungen von Daten zu interpretieren. Außerdem werden Zeichnungen geometrischer Figuren (mit und ohne Hilfsmittel) angefertigt sowie ebene Figuren nach Eigenschaften klassifiziert, Fachbegriffe zugeordnet und Beziehungen zwischen geometrischen Figuren beschrieben (KMK, 2022). In der Lernumgebung steht die Klassifikation (oder Kategorisierung) im Zentrum und somit die Begriffsbildung (Franke & Reinhold, 2016). Schaut man sich die Trainingsdaten genauer an (Abb. 2 und 3), stellt man fest, dass bei den Vierecken (Abb. 2) ausschließlich Bilder von Rechtecken enthalten sind, bei denen die lange Seite parallel zur unteren Blattkante verläuft. Unter den fünf Rechtecken sind drei Quadrate und zwei Rechtecke, bei denen die Seitenlängen allerdings nicht stark voneinander abweichen. Rechtecke werden am leichtesten als solche erkannt, wenn das Seitenverhältnis bei 1:2 liegt oder wenn das Rechteck näherungsweise den DIN-Papierformaten entspricht (Franke & Reinhold, 2016) – was hier nicht der Fall ist. Bei den Dreiecken (Abb. 3) sieht es etwas anders aus: hier sind verschiedene Dreiecke in verschiedenen Lagen zu finden – wobei eine Seite stets parallel zu einer der Blattkanten verläuft. Fünf der sieben Dreiecke sind (näherungsweise) rechtwinklig, eines ist spitzwinklig und eines stumpfwinklig – Kinder bezeichnen diese häufig als ‚spitzes‘ oder ‚flaches‘ Dreieck (Franke & Reinhold, 2016, S. 246 f.). Fünf der Dreiecke sind gleichschenkelig, die anderen beiden sind unregelmäßige Dreiecke. Schaut man sich nun den Test des Klassifizierers an (Tab. 1), stellt man fest, dass die Dreiecke zuverlässig mit einer Sicherheit von 99 % erkannt werden. Auch das Quadrat wird zuverlässig als Viereck erkannt (100 % Sicherheit). Etwas unsicherer ist sich das System (93 % Sicherheit) bei dem Rechteck mit dem näherungsweisen

Seitenverhältnis von 1:2 und dem Rechteck mit einem näherungsweisen Seitenverhältnis von 1:5, dessen kurze Seite parallel zur unteren Blattkante steht (76 % Sicherheit). Falsch erkannt werden das Trapez (mit einer Sicherheit von 95 % als Dreieck) sowie der Drache (mit 94 % Sicherheit als Dreieck). Das angestrebte mathematische Wissen der Kinder ist folgendes: Ein Viereck ist eine ebene Figur mit vier Ecken und vier Seiten. Vierecke gibt es in verschiedenen Formen: alle vier Seiten des Vierecks können gleich lang sein (Quadrat, Raute) oder nur zwei Seiten des Vierecks (Rechteck, symmetrisches Trapez, Parallelogramm, symmetrischer Drache) oder alle Seiten können unterschiedlich lang sein (z. B. unregelmäßiges Viereck). Alle vier gegenüberliegenden Seiten können parallel sein (Quadrat, Rechteck) oder zwei gegenüberliegende Seiten können parallel sein (symmetrisches Trapez, Parallelogramm). Benachbarte Seiten können im rechten Winkel zueinander stehen (Quadrat, Rechteck). Ein Dreieck ist eine ebene Figur mit drei Ecken und drei Seiten. Dreiecke gibt es in verschiedenen Formen, wobei alle Seiten gleich lang (gleichseitiges Dreieck), zwei Seiten gleich lang (gleichschenkliges Dreieck) oder alle Seiten unterschiedlich lang sein können (unregelmäßiges Dreieck) (Franke & Reinhold, 2016). Es gibt spitzwinklige Dreiecke (alle Winkel sind kleiner als 90°), stumpfwinklige Dreiecke (ein Winkel ist größer als 90°) und rechtwinklige Dreiecke (ein Winkel ist genau 90°). Um den Klassifizierer zu verbessern, müsste man also Vierecke in verschiedenen Formen, Größen (Ähnlichkeitsabbildungen) und Lagen (Kongruenzabbildungen) auf dem Papier hinzufügen, ohne allerdings die Dreiecke zu vernachlässigen (da sonst in die andere Richtung verzerrt wird.) Die Aufgabe der Kinder ist also die Ergänzung und ggf. Aufbereitung (Entfernung irrelevanter Daten oder Ausreißer) des Trainingsdatensatzes.

Folgende TDS-Kriterien sind erfüllt, die es ermöglichen eine Situation als a-didaktisch wahrzunehmen: Das mathematische Wissen in Verbindung mit dem Wissen darüber, wie Modelle für ML mithilfe von Daten und Algorithmen trainiert werden, scheint die beste Methode zur Lösung der Aufgabe zu sein. Lernende können zunächst mit unzureichendem ‚Basiswissen‘ starten. Die Lernenden können mit Hilfe von Machine Learning for Kids eine schnelle Abfolge von Versuchen nach dem ‚Trial and Error‘-Prinzip durchführen, da sie selbst feststellen können, ob ihr Versuch gelungen oder gescheitert ist. Eine Lösung kann von den Kindern in angemessener Zeit gefunden und getestet werden und von anderen Kindern überprüft

werden. Die Anwendung folgender Konstrukte könnte zielführend sein (a-priori-Analyse AiC):

K1: Klasseninklusion (R) – Vierecke sind ebene Figuren mit vier Ecken und vier Seiten. Quadrat, Rechteck, Parallelogramm, Trapez, Raute, Drache sind Vierecke. Dreiecke sind ebene Figuren mit drei Ecken und drei Seiten. Gleichseitige, gleichschenkelige, unregelmäßige, spitzwinklige, stumpfwinklige, rechtwinklige Dreiecke sind auch Dreiecke. Hier braucht die Lage auf einem Blatt (links oben/rechts unten/...) nicht betrachtet zu werden, gleiches gilt für die Größe eines Dreiecks/Vierecks und seine Drehung oder Spiegelung (s. Abschnitt 2.3).

K2: Passung zwischen Trainings- und Testdatensatz (B) – Die Klassifizierung des Systems basiert auf den eigenen Eingaben, interpretiert durch das ML-Modell, das die Kinder trainiert haben (Vartiainen et al., 2020). Im Trainingsdatensatz werden deshalb die Figuren ergänzt, die im Testdatensatz getestet werden.

K3: Diversität des Trainingsdatensatzes (C) – Eine fehlerhafte Funktionsweise des Systems wird auf Verzerrungen im Zusammenhang mit dem Training des Systems zurückgeführt, wie z. B. die Über- oder Unterrepräsentation bestimmter Gruppen in den Trainingsdaten (Vartiainen et al., 2024). Vierecks- und Dreiecksarten, die nicht in den Trainingsdaten vorhanden waren, werden ergänzt. Dabei wird auch die Lage auf dem Blatt variiert (nicht immer mit der längsten Seite parallel zur unteren Blattkante). Hier spielt nun also die Lage auf einem Blatt (links oben/rechts unten/...) und die Größe eines Dreiecks/Vierecks und seine Drehung oder Spiegelung eine Rolle.

Hintergrund von Konstrukt **K3** ist, dass es sinnvoll ist, dem Klassifizierer auch Figuren vorzugeben, die nicht auf den ersten Blick dem bekannten Prototyp entsprechen. Beim Zeichnen der Figuren können die Kinder allerdings lediglich die Anzahl der Seiten mit Sicherheit ohne Hilfsmittel feststellen. Zum Überprüfen der anderen Merkmale können sie ein Geodreieck als Hilfsmittel nutzen: Sie messen die Längen der Seiten und prüfen die Rechtwinkligkeit sowie die Parallelität (Franke & Reinhold, 2016) – das funktioniert auch auf dem Tablet.

K4: Iterative Optimierung (C) – Sollte der Klassifizierer bei bestimmten Zeichnungen nicht ‚gut‘ funktionieren, sollten diese im Trainingsdatensatz ergänzt werden.

4.3 Implementierung der Aufgaben

Die Lernumgebung wurde am 07.10.2024 mit elf mathematisch begabten und interessierten Kindern der 4. Klassenstufe im Lehr-Lern-Labor ‚Zahlenraum‘ der Universität des Saarlandes erprobt. Die Kinder konnten sich untereinander nicht und auch die Lehrperson kannte die Kinder im Vorfeld nicht. Ein Kind mit Autismus-Spektrum-Störung war anwesend (Lars, Gruppe 3). Sowohl die Gesamtsituation als auch die Gruppensituationen wurden videographiert und Screencasts auf den Tablets wurden aufgezeichnet, anschließend wurde das Audio-Bild-Material transkribiert. Die Arbeitsblätter wurden eingescannt.

5. Ergebnisse

Zunächst wird das Vorgehen der vier Lernendengruppen beschrieben (Abschnitt 5.1; die Namen der Kinder sind dabei frei erfunden) und anschließend analysiert und gedeutet. Dabei werden die Forschungsfragen adressiert (Abschnitt 5.2).

5.1 Beschreibung

Gruppe 1: Noah, Andreas, Felix

Die Kinder löschten im ersten Schritt alle vorhandenen Trainingsdaten und fügten mit dem Finger gezeichnet neue ein, anfangs ohne darüber zu sprechen (Abb. 4).

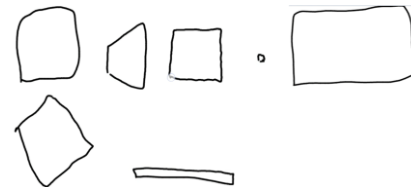


Abb. 4: Hinzugefügte Vierecke zum neuen Datensatz: drei Quadrate in unterschiedlichen Größen, zwei Rechtecke (Seitenverhältnis 2:1 und 5:1) alles parallel zur unteren Blattkante; Trapez mit der längsten Seite parallel zur rechten Blattkante; ein allgemeines Viereck (rautenähnlich)

An dieser Stelle wies die Lehrperson die Kinder darauf hin, dass sie auch mit dem Stift zeichnen können.



Abb. 5: Weitere hinzugefügte Vierecke zum neuen Datensatz: zwei weitere Quadrate in unterschiedlichen Größen und drei Rechtecke (zweimal Seitenverhältnis ca. 10:1, unsauber gezeichnet, einmal ca. 3:1), alle mit der längsten Seite parallel zur unteren Blattkante

Während die Kinder das vorletzte Viereck zeichnen (Abb. 5), teilte die Lehrperson das Arbeitsblatt aus und erklärte im Plenum kurz den Umgang damit. Die Kinder füllten den ersten Teil des Arbeitsblatts aus (Abb. 6). Anschließend gingen die Kinder zu den Dreiecken über. Sie sprachen weiterhin wenig, stimmten sich aber darüber ab, ob die jeweilige gezeichnete Figur für die Trainingsdaten geeignet ist und hinzugefügt werden soll oder nicht.

Abb. 6: Arbeitsblatt, Aufgabe 1

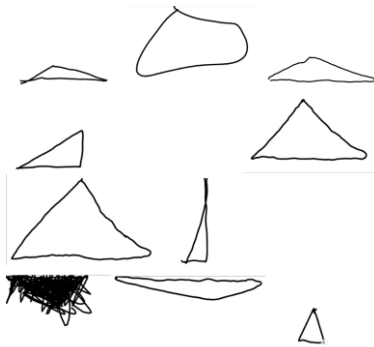


Abb. 7: Hinzugefügte Dreiecke zum neuen Datensatz: drei stumpfwinklige Dreiecke, davon eins mit der Spitze nach unten, drei annähernd gleichschenklige Dreiecke in unterschiedlichen Größen, zwei rechtwinklige Dreiecke mit den kurzen Seiten (Katheten) parallel zur unteren und rechten Blattkante; eine Figur, die eher an ein Viereck erinnert und ein ausgefülltes annähernd rechtwinkliges Dreieck ‚mit Fransen‘ und der Spitze nach unten

An dieser Stelle wurde aus versehen ein Rechteck aus dem Trainingsdatensatz der Vierecke gelöscht. Zwei weitere Dreiecke wurden hinzugefügt (Abb. 8) und eins der nahezu gleichschenkligen Dreiecke (Abb. 7) wurde gelöscht.



Abb. 8: Weitere hinzugefügte Dreiecke zum neuen Datensatz: zwei annähernd gleichseitige Dreiecke, davon eins mit der Spitze nach unten und eins mit weiteren Elementen im Hintergrund

Das Arbeitsblatt wurde weiter ausgefüllt (Abb. 9). Als gleich viele Dreiecke wie Vierecke enthalten

waren (nämlich jeweils elf, Abb. 10), wurde das Modell trainiert. Noah las die Anweisung des Tools laut vor:

Noah: Zeichne etwas, um das Modell zu testen.

Er setzte die Anweisung um (Abb. 11), ohne dies auf dem Arbeitsblatt zu dokumentieren.

Abb. 9: Arbeitsblatt, Aufgabe 2

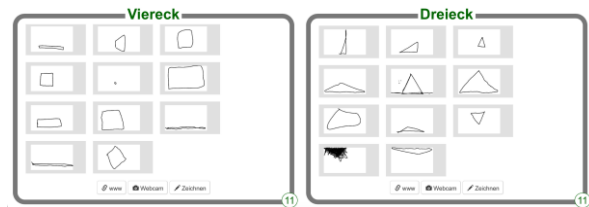


Abb. 10: Trainingsdaten



Abb. 11: 1. Test – erkannt als Dreieck (75 % Sicherheit)

Noah: Wie Dreieck, was stand da? Erkannt als Dreieck. Das war doch kein Dreieck. Mit 75 ähm ich mach gleich 100 %, ich sags euch (unv.)



Abb. 12: 2. Test – erkannt als Viereck (93 % Sicherheit)

Reihum wurde weitergetestet.



Abb. 13: 3. und 4. Test – links: erkannt als Dreieck (97 % Sicherheit), rechts: erkannt als Viereck (99 % Sicherheit)

Anschließend wurde versehentlich die leere Zeichenfläche getestet und mit 66 % Sicherheit als Dreieck erkannt.

Noah: Hm (fragend) wieso wird das die ganze Zeit erkannt/ Nichts wird als Dreieck erkannt?

Eine Diskussion wird allerdings nicht ausgelöst. Andreas nimmt ein Geodreieck zur Hand und zeichnet (Abb. 14, links).



Abb. 14: 5. und 6. Test – links: erkannt als Dreieck (95 % Sicherheit), rechts: erkannt als Viereck (100 % Sicherheit)

Noah: Ich hab 100 % Viereck jetzt muss ich noch 100 bei Dreieck.

Ein Quadrat wurde von Felix gezeichnet und ebenso mit 100% als Viereck erkannt.



Abb. 15: 7. und 8. Test – links: erkannt als Dreieck (97 % Sicherheit), rechts: erkannt als Dreieck (98 % Sicherheit)

Anschließend wurde folgendes Bild mit 100 % als Viereck erkannt:



Abb. 16: 9. Test – erkannt als Viereck (100 % Sicherheit)

Andreas: Er hat mit einem Smiley 100 Viereck als 100 % geschafft.

Noah: Wieso? Wieso? Ich mach das jetzt mal aus Spaß, ok? ich mach das jetzt mal aus Spaß. Und dann darf ich nochmal. Wieso hat der mit einem Smiley 100 % geschafft?

Abermals wird ein Smiley mit 100 % als Viereck erkannt. Die Kinder zeigen ihre Entdeckung der Lehrperson. Die Lehrperson fragt die Kinder nach ihren Trainingsdaten, die Kinder gehen allerdings nicht auf die Nachfrage ein. Sie testeten weiter Dreiecke und versuchen 100 % zu erreichen.

Andreas: Ich will mal 0 schaffen, wie kann man 0 schaffen?

[...]

Noah: Mal mal einen Kreis. Mal mal einen Kreis.

Andreas: Kreis?



Abb. 17: 10. Test – erkannt als Viereck (100 % Sicherheit)

Noah: Ich wusste es doch.

Allerdings erklärt Noah nicht, was er wusste. Im Plenum erklären die Kinder, dass sie Dreiecke und Vierecke gemacht haben, einen Smiley, einen Kreis, dass der Smiley mit 100 % als Viereck erkannt wurde und dass der Klassifizierer noch nicht verbessert werden konnte. Optimierungsmaßnahmen schlagen die Kinder keine vor.

Gruppe 2: Jakob und Dennis

Bevor das Arbeitsblatt ausgeteilt wurde, fügte die Gruppe ein Quadrat zu den Trainingsdaten der Vierecke und ein gleichschenkliges, rechtwinkliges Dreieck zu den Dreiecken hinzu (Abb. 18; 19). Bei beiden Figuren ist die längste Seite parallel zur unteren Blattkante.



Abb. 18: Hinzugefügte Figuren zum neuen Datensatz: ein Quadrat und ein annähernd gleichschenkliges, rechtwinkliges Dreieck mit der Spitze nach oben

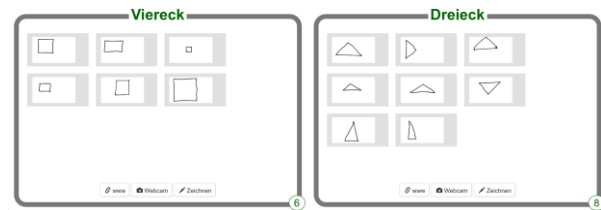


Abb. 19: Trainingsdaten

Anschließend wurde das Modell getestet (Abb. 20, erste Zeile), zuerst an einem Quadrat, das das Modell richtig erkannte (100 % Sicherheit). Anschließend testeten die Kinder ein Trapez, welches immer noch als Dreieck erkannt wurde (99 % Sicherheit). Mittlerweile wurde das Arbeitsblatt ausgeteilt, das allerdings nur kurz die Aufmerksamkeit der Kinder auf sich zog. Die Kinder testeten ein Sechseck (Abb. 20), welches als Dreieck erkannt wurde (91 % Sicherheit). Die Folgerung von Jakob ist:

Jakob: Guck, ok, wir müssen das besser zeichnen. Bisher erkennt er ja/ so schlecht erkennt er das nicht.

Die Kinder testeten weitere Figuren, die keine Dreiecke oder Vierecke waren (Abb. 20, zweite Zeile). Die meisten wurden als Dreiecke erkannt.

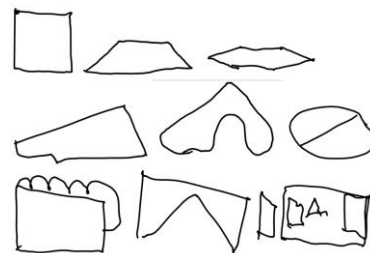


Abb. 20: Test 1-9

Die Kinder versuchten ab Zeile 3 (Abb. 20) die Zuordnung des Systems zu begründen, z. B.

Dennis: Viereck. Eigentlich passt es schon, weil es war auch ein Viereck.

Oder bei der nächsten Figur:

Jakob: Eigentlich/ das meiste war/ warte mal.

Dennis: Eigentlich war es auch ein Dreieck und ein Viereck.

Anschließend wurde das ‚beste Dreieck‘ (ein gleichschenkliges Dreieck mit der Spitze nach oben) gezeichnet, das auch als solches erkannt wurde. Die Lehrperson gab der Gruppe nochmal den Hinweis, dass sie ihren Test auf dem Arbeitsblatt dokumentieren können. Die Kinder gaben die Rückmeldung, dass ihr Klassifizierer bisher nicht so gut funktioniert. Sie testeten zunächst ohne Dokumentation weiter. Die Lehrperson schaute nochmals nach der Gruppe und die Kinder fragten, wie sie mit dem System automatisch eine solche Tabelle wie Tab. 1 erzeugen können. Die Lehrperson erklärte, dass man eine solche Tabelle selbst ausfüllen muss und verwies erneut auf das Arbeitsblatt. Sie fragte auch nach, ob die Kinder den Klassifizierer denn schon verbessert hätten, die Frage wurde allerdings ignoriert. Die Kinder begannen die Tabelle auf dem Arbeitsblatt (Abb. 21) auszufüllen.

Bild	Klassifikation	Zuverlässigkeitsgrad	Richtig?
	Viereck	98%	Ja!
	Viereck	100%	Ja!
	Dreieck	97%	Nein!
	Viereck	93%	Ja!
	Viereck	96%	nein!
	Dreieck	90%	Ja!
	Dreieck	93%	nein!
	Dreieck	63%	nein!
	Viereck	100%	Ja!
	Dreieck	96%	nein!
	Viereck	87%	Ja!

Abb. 21: Arbeitsblatt, Aufgabe 3

Jakob wollte unbedingt ein ‚Männchen‘ testen, das als Dreieck erkannt wurde (58 % Sicherheit, von den Kindern als 98 % gelesen):



Abb. 22: 11. Test – erkannt als Dreieck (58 % Sicherheit)

Dennis: Ach, nicht schon wieder dein Männchen. Ok, mach.

Jakob: Bist du nicht auch neugierig, als was der das erkennt?

Dennis: Eigentlich steht da nur Dreieck oder Viereck.

Jakob: Ich mal dem Dreieck-Hände.

Dennis: Hat der noch ein Gesicht oder nicht?

Jakob: Ja. So ein Viereckkopf.

Nun testeten die Kinder die Figuren in der Tabelle und stellten beim Testen des Trapezes vor der Ausgabe des Tools fest:

Dennis: Also, ich weiß schon ganz genau, dass das falsch ist.

Etwas später stellt Dennis nochmals fest:

Dennis: Ok, aber wir haben es ja auch so programmiert, dass es nur Dreieck oder Viereck sagt.

Auf die Sicherheit der einzelnen Tests wurde nur insofern geachtet, dass versucht wurde auf die gleichen Werte wie in Tab.1 zu kommen.

Jakob: Immer wenn der was nicht erkennt, dann sagt der Dreieck. Das sag ich dir. Ich sag der sagt Dreieck oder Viereck.

Dennis: Es kann ja nur Dreieck oder Viereck sein.

Jakob: Aber ich glaub eher Dreieck.

Dennis: Und jetzt testen. Dreieck 63 %.

Jakob: Ich wusste es.



Abb. 23: 12. Test – erkannt als Dreieck (63 % Sicherheit)

Jakob: Eigentlich ist das ja eine Raute und kein Viereck, oder? (Deutet auf die letzte Zeile in Tab.1)

Besprechung im Plenum:

L: Hat eurer denn besser funktioniert als meiner?

Jakob: //Ja manchmal.

Dennis: Ja manchmal.//

[...]

L: Was würdet ihr im nächsten Schritt noch machen, um den noch besser zu machen?

Dennis: Ähm. Also. (4)

L: Ausgefüllte Formen vielleicht noch, gell?

Dennis: Ja, also beim Trainieren eine ausgefüllte Form hinzu und dann sagen das wäre ein Viereck.

Gruppe 3: Lars, Lilli und Heike

Zunächst wurde zu den Trainingsdaten bei den Vierecken ein Trapez und ein Rechteck mit dem ungefähren Seitenverhältnis 1:7 hinzugefügt, die längste Seite jeweils parallel zur unteren Blattkante, sowie ein Drache (Abb. 24).



Abb. 24: Hinzugefügte Vierecke zum neuen Datensatz: ein Trapez, ein Rechteck (Seitenverhältnis ca. 1:7) und ein Drache

Nachdem die Kinder kurz darüber nachdachten das Modell zu testen, entschieden sie sich dazu, doch zuerst noch weitere Trainingsdaten (Abb. 26) hinzuzufügen. Ein gleichschenkliges Dreieck wurde hinzugefügt (Abb. 25).



Abb. 25: Hinzugefügte Dreiecke zum neuen Datensatz: ein gleichschenkliges Dreieck

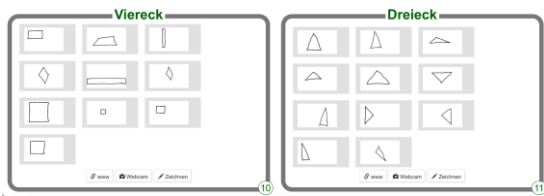


Abb. 26: Trainingsdaten

Die Lehrperson teilte das Arbeitsblatt mit einer kurzen Erläuterung aus. Die Kinder fragten, ob sie auch schon testen dürfen. Die Lehrperson bejaht und erklärt, wie das Arbeitsblatt dabei eingesetzt werden soll. Die Kinder einigen sich darauf, zuerst zu testen, bevor sie das Arbeitsblatt ausfüllen.

Heike: Soll ich jetzt mal ein Trapez machen, das haben wir ja schon eingescannt?

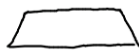


Abb. 27: 1. Test – erkannt als Viereck (98 % Sicherheit)

Das Ergebnis wurde in der Tabelle notiert (Abb. 29). Anschließend testeten die Kinder ein Rechteck und ein Quadrat, die zuverlässig erkannt wurden. Lars und Heike zeichnten abwechselnd auf dem Tablet, während Lilli die Tabelle auf dem Arbeitsblatt ausfüllte. Beim Eintragen des Ergebnisses zum Quadrat kam die Diskussion auf, ob das Quadrat ein Viereck ist:

Heike: Viereck hinschreiben.

Lilli: Aber eigentlich ist es ja ein Quadrat. Ist ja eigentlich falsch. Aber es war ja Viereck und Quadrat, deswegen ist es eigentlich richtig.

Heike: Es sind Vierecke und Dreiecke und das Quadrat ist ein Viereck.

Auch ein Viereck mit dem Seitenverhältnis 1:5 mit der kurzen Seite parallel zur Blattkante wurde nun zuverlässig (99 %) als Viereck erkannt. Das erste Dreieck in der Tabelle (Abb. 28) wurde fälschlicherweise als Viereck (73 % Sicherheit) erkannt.

Bild	Klassifikation	Zuverlässigkeitsgrad	Richtig?
	Viereck	100%	ja
	Viereck	92%	ja
	Viereck	98%	ja
	Viereck	99%	ja
	Viereck	73%	nein
	Dreieck	50%	ja
	Dreieck	83%	nein
	Dreieck	94%	ja
	Viereck	63%	ja
	Viereck	64%	ja
	Viereck	56%	nein

Abb. 28: Arbeitsblatt

Es wurde kurz diskutiert, ob man es nochmal testen solle, dann wurde sich darauf geeinigt, dass es auch falsch sein darf. Das nächste Dreieck wird zwar als solches erkannt, aber nur mit einer Zuverlässigkeit von 50%. Der Drache wurde immernoch als Dreieck erkannt (85% Sicherheit). An dieser Stelle entschied die Gruppe weitere Figuren zu den Trainingsdaten hinzuzufügen.

Heike: Aber wir haben doch schon die Raute gemacht. Warum hat er die denn dann nicht?

Es wurde sich dazu entschieden zuerst die Dreiecke zu ergänzen. Es wurde ein Dreieck der schon vorhandenen Testdaten vertikal gespiegelt und ein weiteres gleichschenkliges Dreieck wurde hinzugefügt. Bei den Vierecken wurde 'eine ganz dünne Raute' ergänzt und das Rechteck aus den Testdaten wurde ergänzt (Abb. 29).



Abb. 29: Hinzugefügte Figuren zum neuen Datensatz: zwei rechtwinklige Dreiecke, eins davon mit der längsten Seite (Hypotenuse) parallel zur rechten Blattkante, eins davon ohne parallele Seiten zur Blattkante; ein gleichseitiges Dreieck, eine 'ganz dünne' Raute und ein Rechteck (Seitenverhältnis ca. 1:5) mit der kurzen Seite parallel zur unteren Blattkante

Heike: Das ist aber eine komische Raute. Ja, dann weiß er wenigstens/

Lars: Es ist ja eigentlich gut, dass es eine komische Raute ist, es/

Heike: Ja, weil dann weiß er, dass es trotzdem eine Raute ist.

Zum Schluss wurde ein Dreieck hinzugefügt – ein rechtwinkliges, das mit keiner Seite parallel zu einer der Blattkanten steht (Abb. 31, rechts). Dann wurde wieder getestet. Das Dreieck in Zeile 8 der Tabelle (Abb. 28) wurde nun zuverlässig erkannt. Nach mehrmaligen Zeichenversuchen einer Raute wurde sich dazu entschieden, diese zu testen (Abb. 30, links).



Abb. 30: 2.-5. Test – von links: erkannt als Viereck (Sicherheit: 63 %), Viereck (64 % Sicherheit), Viereck (56 % Sicherheit), Dreieck (84 % Sicherheit)

Die Kinder widmeten sich nun dem Arbeitsblatt. Es wurde diskutiert, ob der neue Klassifizierer besser funktioniert hat.

Heike: Bei den Vierecken hat es besser funktioniert, weil guck mal hier hat er es als Dreieck gehalten, das war ja falsch und hier hat er es als Viereck gehalten.

Lilli: hm (fragend) eigentlich ist beides gleich

gewesen.

Heike: Nein wir haben doch dreimal falsch.

Lilli: Ja, stimmt, wir hatten dreimal falsch. Die Dreiecke waren schlechter, weil es war zweimal das Dreieck falsch und nur einmal das Viereck falsch.

Heike: Die Dreiecke waren schlechter.

Lars: Wir müssen ja auch auf die Prozente gucken wie sicher er sich war.

Lilli: Nein, weil so oft wars ja falsch.

Zur Beantwortung von Teilfrage c) warfen die Kinder nochmal einen Blick auf die Trainingsdaten und diskutierten, bevor der Satz in Abb. 31 notiert wurde:

Heike: Wir hatten mehr Dreiecke als Vierecke. Mehr Dreiecke wurden als Vierecke erkannt.

[...]

Lilli: Es liegt daran, dass wir mehr Dreiecke hatten.

Heike: Nein, es liegt daran, dass mehr Dreiecke falsch erkannt wurden.

Lilli: Warte mal, es hat bei den Vierecken besser

erkannt, aber die Dreiecke wurden schlechter erkannt.

Heike: Obwohl die Dreiecke mehr waren.

[...]

Lars: Was ist aber mit dem Warum?

Lilli: Ich weiß, deswegen ist es ja so witzig.

Heike: Ja weil (...) es sieht (unv.)

Lilli: Warte, es hat bei den Vierecken besser funktioniert und es liegt daran (..) keine Ahnung.

Heike: Ei, weil vielleicht die Vierecke ähm da mehr

ding gemacht mehr (..) verschiedenen gemacht

haben. Bei den Dreiecken hatten wir ja immer

Dreieck, so ein Dreieck mal so ein Dreieck ähm

Lilli: Aber dann hätte es bei den Vierecken ja

schlechter laufen müssen.

Heike: Ja, aber die Vierecke haben mehr verschiedene Formen, zum Beispiel das Trapez

Lilli: Ist ja aber eigentlich schwieriger für die KI, dann muss es ja eigentlich mehr bei den Vierecken falsch gelegen haben.

Heike: Nein.

Lilli: Oder vielleicht liegt's daran, weil die Dreiecke so schrottig gemalt worden sind. (lachen)

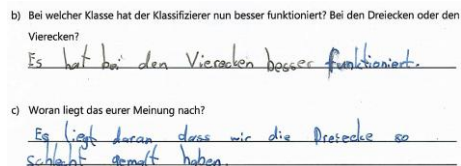


Abb. 31: Arbeitsblatt

Anschließend wurde die erste Seite des Arbeitsblattes ausgefüllt (Abb. 32).

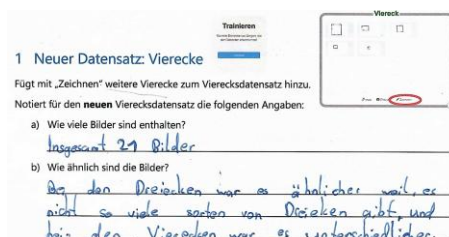


Abb. 32: Arbeitsblatt

Bei der Besprechung im Plenum beschrieben die Kinder folgende Optimierungsmaßnahmen:

Heike: Wir wollten halt gucken, dass wir die Sachen, die er jetzt falsch erkannt hat, nochmal also dort halt mehr machen. Und in der Richtung ein bisschen mehr, dass er sich dort verbessern kann, würde ich jetzt mal sagen, oder?

Lilli: So haben wir es ja eigentlich auch gemacht.

Gruppe 4: Laura, Michaela, Oskar

Zuerst wurden reihum die Trainingsdaten ergänzt, bei den Vierecken ein Drache, ein Rechteck mit dem Seitenverhältnis 1:4 mit der längsten Seite parallel zur unteren Blattkante, ein ‚kleines‘ Quadrat und ein Trapez ergänzt. Bei den Dreiecken wurde ein ‚kleines‘ gleichseitiges Dreieck und ein gleichseitiges, rechtwinkliges Dreieck, dessen Schenkel parallel zur linken und oberen Blattkante lagen (Abb. 33). Der Prozess lief weitestgehend schweigend ab. An dieser Stelle teilte die Lehrperson das Arbeitsblatt aus.

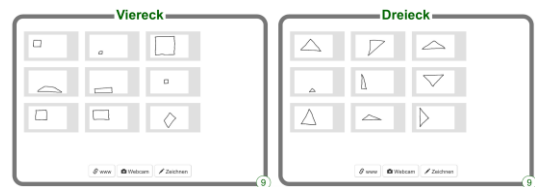


Abb. 33: Trainingsdaten

Laura fragte zweimal bei der Lehrperson nach, wie das Arbeitsblatt ausgefüllt werden solle. Die Lehrperson erklärte es zweimal und gab zu Teilfrage 1 b) den Hinweis, dass sich die Kinder die Vierecke nochmal genau anschauen sollen und dass sie vielleicht auch Namen für die Vierecke kennen. Nachdem alle Fragen geklärt waren, begannen die Kinder das Arbeitsblatt auszufüllen (Abb. 34). Nachdem die erste Seite des Arbeitsblattes ausgefüllt war, versuchten die Kinder den Klassifizierer zu testen. Dabei stellte sich heraus, dass das Tablet die Einschätzung und Sicherheit nicht anzeigte, weshalb die Gruppe ein neues Tablet bekam. Anschließend füllte die Gruppe die Tabelle aus (Abb. 35).

1 Neuer Datensatz: Vierecke

Fügt mit „Zeichnen“ weitere Vierecke zum Vierecksdatensatz hinzu.
Notiert für den neuen Vierecksdatensatz die folgenden Angaben:

a) Wie viele Bilder sind enthalten?
9

b) Wie ähnlich sind die Bilder?
viele verschiedene wie ein Trapez, rechteck und Raute

c) Inwiefern sind die Bilder unterschiedlich?
etwas fern.

2 Neuer Datensatz: Dreiecke

Fügt mit „Zeichnen“ weitere Dreiecke zum Dreiecksdatensatz hinzu.
Notiert für den neuen Dreiecksdatensatz die folgenden Angaben:

a) Wie viele Bilder sind enthalten?
9

b) Wie ähnlich sind die Bilder?
fast gleich.

c) Inwiefern sind die Bilder unterschiedlich?
kann sehr unterschiedlich.

Abb. 34: Arbeitsblatt; Die Antwort bei 1c) bezog sich vermutlich auf die Lage auf dem Blatt: dass manche neu hinzugefügte Bilder weiter weg (ferner) von der Blattkante gezeichnet wurden

Bild	Klassifikation	Zuverlässigkeitsgrad	Richtig?
	Viereck	89%	Ja
	Viereck	96%	Ja
	Dreieck	62%	nein
	Viereck	94%	Ja
	Dreieck	73%	Ja
	Viereck	69%	nein
	Dreieck	89%	nein
	Viereck	65%	Nein
	Viereck	90%	ja
	Dreieck	95%	Ja
	Dreieck	100%	Nein

Abb. 35: Arbeitsblatt

Auch hier wurde reihum getestet und geschrieben. Das Trapez wurde weiterhin als Dreieck erkannt, allerdings nur mit einer Sicherheit von 62 %. Dennoch wurde keine Diskussion dadurch ausgelöst. Ein von den Kindern als ‚Stange‘ bezeichnetes Rechteck wurde hingegen richtig als Viereck erkannt. Die Tabelle wurde weiter ausgefüllt. Während diesem Prozess wurde in der Gruppe zwar mehr gesprochen, allerdings eher über private Themen als über die Aufgabe. Auch als die Raute als Dreieck erkannt wurde, wurde keine Diskussion ausgelöst. Als schließlich eigene Figuren getestet wurden, rückte die Aufgabe etwas mehr in den Fokus der Kinder.

Oskar: Wieso hat noch nichts 100 %?

Michaela: Weil wir die so gemalt haben.

Auch hier wurde die Diskussion nicht vertieft. Das Arbeitsblatt wurde weiter ausgefüllt (Abb. 36).

b) Bei welcher Klasse hat der Klassifizierer nun besser funktioniert? Bei den Dreiecken oder den Vierecken?
Es hat gleich gut funktioniert!

c) Woran liegt das eurer Meinung nach?
Wie haben nicht mega ordentlich gezeichnet.

d) Was würdet ihr im nächsten Schritt optimieren?
Nocheinmal ausprobieren.

Abb. 36: Arbeitsblatt

Bei der Besprechung wiesen die Kinder darauf hin, dass sie fünfmal ‚nein‘ haben. Die Kinder hatten keine Idee, woran das liegen könnte.

5.2 Analyse und Deutung

Im Folgenden wird untersucht, inwiefern die beobachteten epistemischen Aktionen auf ein wachsendes Verständnis algorithmischer Verzerrung hindeuten – verstanden als das Phänomen, dass Algorithmen bestimmte Gruppen systematisch begünstigen oder benachteiligen, wenn die Trainingsdaten unausgewogen repräsentiert sind (Kordzadeh & Ghasemaghahi, 2022). Algorithmische Verzerrung dient dabei nicht nur als Lerngegenstand der Lernumgebung, sondern auch als analytischer Bewertungsmaßstab: Ein tieferes Verständnis des Begriffs zeigt sich, wenn Lernende die Ursache fehlerhafter Klassifikationen im Datensatz verorten – und nicht in menschenähnlichen Eigenschaften der KI.

5.2.1 FF1: Welche epistemischen Aktionen sind bei Lernenden bei der Erprobung der Lernumgebung erkennbar?

Die in Abschnitt 4.2 genannten Konstrukte können in den Lösungen der Lernenden gefunden werden. Insgesamt sind folgende Konstrukte bei den einzelnen Gruppen erkennbar (Tab. 2). In das Vorgehen von Gruppe 3 können alle zielführenden Konstrukte geordnet werden.

Gr.	K1	K2	K3	K4	K5	PaCC1	PaCC2
1		(x)	x			x	x
2	(x)	(x)		(x)		x	
3	x	x	x	x	(x)	(x)	
4		x	x			x	

Tab. 2: Identifizierte Konstrukte in den einzelnen Gruppen sind mit x gekennzeichnet, (x) bedeutet, dass das Konstrukt teilweise hineingedeutet werden kann

Im Folgenden werden die in Tabelle 2 erfassten Ergebnisse geordnet nach Konstrukten beschrieben.

K1 (Klasseninklusion): In Gruppe 3 stellt Heike in einer Diskussion mit Lilli richtig, dass das Quadrat ein

Viereck ist. In Gruppe 2 ist Jakob unsicher, ob die Raute ein Viereck ist, allerdings bleibt seine Frage unbeantwortet. Manche Kinder der Gruppe können dem geometrisch-analysierenden Denken des van Hiele-Modells (Franke & Reinhold, 2016) zugeordnet werden, andere dem geometrisch-abstrahierenden Denken.

K2 (Passung zwischen Trainings- und Testdatensatz): Gruppe 3 und Gruppe 4 ergänzen die Testdaten, die in Tab. 1 nicht richtig erkannt wurden, in den Trainingsdaten. Gruppe 3 geht iterativ vor und ergänzt den Datensatz wiederholt aufgrund der Testergebnisse, dabei wird auch die Passung zwischen Trainings- und Testdaten von den Kindern aufgegriffen. Gruppe 1 löscht alle Trainingsdaten. In die ergänzten Vierecke könnte ein Bezug zu den Testdaten in Tab. 1 gedeutet werden. Interessant ist, dass alle drei Gruppen zwar ein ‚langgezogenes Rechteck‘ ergänzen, allerdings mit der längsten Seite parallel zur unteren Blattkante, vermutlich, weil geometrische Figuren so für Kinder leichter erkennbar sind (Franke & Reinhold, 2016; Pielsticker, 2020). Gruppe 3 zeichnete es später dann auch in der Lage, wie es in den Testdaten (Tab. 1) enthalten war. In Gruppe 2 verdeutlicht Dennis, dass der Klassifizierer nur Dreiecke und Vierecke erkennen kann und zeigt somit, dass er verstanden hat, dass die Klassifizierung des Systems auf den eigenen Eingaben, interpretiert durch das ML-Modell, das die Kinder trainiert haben, basiert (Vartiainen et al., 2020). Ebenso Heike (Gruppe 3), die ein Trapez testen möchte, da dieses schon ‚eingescannt‘ ist.

K3 (Diversität des Datensatzes): Gruppe 3 und Gruppe 4 ergänzten verschiedenartige Figuren, wobei Gruppe 3 auch die Lage auf dem Blatt variierte, während Gruppe 4 eher die Größe der Figuren variierte. Gruppe 4 ergänzte auch ein gleichseitiges Dreieck, das im ursprünglichen Datensatz nicht enthalten war. Gruppe 1 fügte ein Trapez mit der längsten Seite parallel zur rechten Blattkante hinzu, allerdings ohne Trapeze in anderen Lagen zu ergänzen. Bei den Dreiecken ist auch ein gleichseitiges enthalten. Außerdem wurden auch ‚Randfälle‘ (oder Ausreißer) hinzugefügt, bspw. ein ausgefülltes Dreieck ‚mit Fransen‘. Diese ‚Randfälle‘ erinnern an die Ergebnisse der Studie von Burger und Shaughnessy (1986), bei der Lernende, die Aspekte wie Lage auf dem Blatt, ‚gerade‘ oder ‚mit der Spitze nach unten‘ stehen sowie das nicht bewusste Beachten relevanter (euklidischer) Eigenschaften, was z. B. auch zu ‚zackigen‘ Seitenlinien einzelner Figuren führte, dem räumlichanschauungsgebunden Denken (van Hiele-Modell; Franke & Reinhold, 2016) zugeordnet

wurden. Keine der Gruppen ergänzte ein Parallelogramm in den Trainingsdaten, obwohl es bei den Zeichnungen einzelner Kinder bei der vorhergehenden Aufgabe enthalten war. In Gruppe 3 erklärt Heike, dass der Klassifizierer – obwohl der Trainingsdatensatz der Dreiecke größer war – bei den Vierecken besser funktioniert hat, da hier mehr variiert wurde. Die fehlerhafte Funktionsweise des Systems wird also auf Verzerrungen im Zusammenhang mit dem Training des Systems zurückgeführt, wie z. B. die Über- oder Unterrepräsentation bestimmter Gruppen in den Trainingsdaten (Vartiainen et al., 2024). Heikes Deutung ist dabei im Sinne algorithmischer Verzerrung (Kordzadeh & Ghasemaghahi, 2022) als konzeptuell tragfähig einzuschätzen: Sie erkennt, dass die ungleiche Repräsentation bestimmter Figurtypen im Trainingsdatensatz – also eine Unterrepräsentation vielfältiger Vierecksformen – ursächlich für die systematisch schlechtere Klassifikationsleistung bei den Vierecken ist. Damit konstruiert Heike den Kern des Begriffs algorithmischer Verzerrung: die Datenbasis ist strukturell unausgewogen, was zu einer einseitigen Ausgabe führt. Auch in das Gespräch mit Lars zur ‚komischen Raute‘ kann dies gedeutet werden.

K4 (Iterative Optimierung): Gruppe 3 prüft immer wieder ab, ob die im Trainingsdatensatz vorhandenen Daten erkannt wurden und optimiert den Trainingsdatensatz entsprechend. Das geschieht in zwei Iterationen, das Abbruchkriterium ist vermutlich die begrenzte Zeit, da Heike als Optimierungsmaßnahme eine weitere Iteration vorschlägt. Auch Dennis (Gruppe 2) schlägt das Ergänzen einer weiteren Abbildung als Optimierungsmaßnahme vor, allerdings evtl. suggeriert durch die Lehrperson.

PaCC1: Erkennung von Figuren, die nicht im Trainingsdatensatz enthalten sind – Eine fehlerhafte Funktionsweise des Systems wird auf menschenähnliche kognitive Eigenschaften der KI zurückgeführt (KI denkt, KI glaubt) ohne Bezugnahme auf Verzerrungen in den Trainingsdaten (Vartiainen et al., 2024). Z. B. in den Gruppen 1 und 2 wurden andere Figuren gezeichnet als Dreiecke und Vierecke, in Gruppe 1 Smiley und Kreis und in Gruppe 2 Männchen und schwarze Flächen. Jakob erklärt, warum er das tut: ‚Bist du nicht auch neugierig, als was der das erkennt?‘

Ein Erklärungsansatz für dieses PaCC ist, dass menschliche Erfahrungen auf das System übertragen wurden, denn:

Menschen lernen ihr ganzes Leben lang und machen eine Vielzahl von Erfahrungen, sodass sie auch neue

Aufgaben [...] recht schnell meistern können. Ein Lernalgorithmus lernt jedoch von Grund auf. Deshalb brauchen Lernalgorithmen tendenziell viel mehr Beispiele als Menschen, um zugrunde liegende Muster zu erkennen oder Entscheidungen zu treffen. Dementsprechend wird, je mehr Daten für das Training zur Verfügung stehen, im Allgemeinen auch die Auswertung auf den Testdaten besser. In der Praxis brauchen überwachte Lernalgorithmen nicht nur ein paar Dutzend, sondern typischerweise Tausende von Beispielen, die zunächst aufwendig von Hand zugeordnet werden müssen. (Kersting et al., 2019, S. 27)

Diese Übertragung menschlicher Erfahrungen wird auch deutlich, als bspw. in Gruppe 1 versucht wird, bessere Testergebnisse durch genaues Zeichnen (mit und ohne Geodreieck), sowie in Gruppe 2 und 4, als das ‚schlechte‘ Erkennen von Zeichnungen auf die eigenen Zeichenfähigkeiten zurückgeführt wird, ohne jedoch den Trainingsdatensatz zu berücksichtigen. Das kann auch in Lillis Äußerung des ‚schrottigen‘ Zeichnens (Gruppe 3) gedeutet werden.

PaCC1 ist aus der Perspektive algorithmischer Verzerrung besonders aufschlussreich: Er zeigt, dass das konzeptuelle Verständnis des Phänomens noch nicht konstruiert wurde. Algorithmische Verzerrung setzt voraus, dass Lernende die Ursache ungleicher Ausgaben im Datensatz verorten – also in einer strukturellen Eigenschaft des Systems. Solange Kinder jedoch davon ausgehen, dass die KI wie wir Menschen ‚denkt‘ oder ‚glaubt‘, verbleibt die Ursachenzuschreibung auf einer anthropomorphen Ebene, die den datenbasierten Charakter des Phänomens verdeckt. PaCC1 markiert damit eine zentrale konzeptuelle Hürde auf dem Weg zum Verständnis algorithmischer Verzerrung.

PaCC2: 100 % oder 0 % Sicherheit erreichen – In Gruppe 1 wurde versucht eine Sicherheit von 100 % zu erreichen sowie von 0 %. Dabei wurden verschiedene Strategien angewendet, allerdings wurde der Trainingsdatensatz nicht berücksichtigt.

Auf diesem PaCC aufbauend, sollte ein Konstrukt ergänzt werden:

K5: Güte definieren (C) – Es wird ausgehandelt, wann eine Klassifikation als ‚gut‘ anerkannt wird, d. h. ab welcher Sicherheit keine Optimierung des Klassifizierers mehr notwendig ist.

Eine Aushandlung hat in keiner der Gruppen stattgefunden, allerdings wies Lars (Gruppe 3) darauf hin, dass es wichtig ist auch auf die Prozente zu schauen, um zu sehen, wie sicher sich der Klassifizierer ist.

5.2.2 FF2: Welche Hürden können rekonstruiert werden?

Um diese Forschungsfrage zu beantworten, werden (nach TDS) das Milieu, die Variablen, der didaktische Vertrag und Probleme betrachtet. Das Milieu zum Auftrag den Klassifizierer zu optimieren, indem neue Trainingsdaten – Vierecke und Dreiecke – hinzugefügt werden, bestand aus dem von der Lehrperson vortrainierten Klassifizierer, auf den die Lernenden mittels Tablets zugegriffen, eine kurze Handlungsanweisung zur Bedienung des Klassifizierers (s. Anhang 1) sowie einem Arbeitsblatt mit Aufgaben (s. Anhang 2). Zum Bearbeiten bekamen die Kinder zudem einen Stift zum Zeichnen auf dem Tablet.

Die Variablen im Milieu, die von der Lehrperson geändert werden könnten, ist z. B. die Anzahl und Art der geometrischen Figuren in den Trainingsdaten des Klassifizierers. Durch eine Änderung könnte der Fokus auf bestimmte geometrische Eigenschaften gelenkt werden (bezüglich **K2**).

Der didaktische Vertrag ist nicht leicht zu fassen, da die Kinder der Lehrperson unbekannt waren und die Kinder – die sich auch untereinander nicht kannten – an einem für sie ungewohnten Ort (im Lehr-Lern-Labor Zahlenraum der Universität des Saarlandes) unterrichtet wurden. Eine Lehrperson, die die Kinder aus der Schule kannten, war nicht anwesend. Dennoch wurden gewisse Regeln, die auch in der Schule gelten, erwartet (leise sein und zuhören; sich melden, wenn man etwas sagen möchte; usw.).

Zur Lösung der Aufgabe wurden Medienkompetenzen (Bedienkompetenzen) von den Lernenden erwartet sowie die in Abschnitt 4.2 genannten mathematischen Kompetenzen, die in der Lernumgebung vertieft wurden. Die gestellte Aufgabe ist für die Kinder sicher ungewohnt, da sie in eine andere Rolle – in die eines Lehrenden – schlüpfen mussten: sie sollten einem System (dem Klassifizierer) etwas beibringen. Das ist eine komplexe Aufgabe, da die Kinder, die geometrisch-analysierend oder -abstrahierend denken (van Hiele-Modell; Franke & Reinhold, 2016), ein System trainieren müssen, das keine menschenähnlichen kognitiven Eigenschaften hat und dessen Klassifizierung auf den eigenen Eingaben, interpretiert durch das ML-Modell, basiert (Vartiainen et al., 2020; Konstrukt **K3**). Diese Komplexität der Aufgabe sowie die durch die Lehrperson nicht gelieferte Informationen zur Zuverlässigkeitseinschätzung (Güte, **K5**) des Klassifizierers führte vermutlich zu den nur teilweise korrekten Konstrukten (**PaCC1** und **PaCC2**). Das Vorgehen bei der Datenaufbereitung und der Zusammenhang mit der

Funktionsweise des Klassifizierers wurde nicht erklärt, obwohl in ‚echten‘ Anwendungen die Vorbereitung der Daten für einen Lernalgorithmus ein arbeitsaufwendiger Prozess ist, der aus dem Erschließen geeigneter Datenquellen und dem Aufbereiten der Daten besteht (Kersting et al., 2019).

Folgende weitere Aspekte, die für die Lernenden problematisch waren, sind im Milieu aufgetreten: Ein technisches Problem ist bei dem Tablet von Gruppe 4 aufgetreten, da die Website zur Testung des Klassifizierers nicht korrekt angezeigt wurde. Da noch mehr Tablets vorhanden waren und die Trainingsdaten in einer Cloud gespeichert waren, konnte die Gruppe recht zügig mit einem neuen Tablet weiterarbeiten. Dennoch kam es zu einer Unterbrechung im Arbeitsprozess. Das Arbeitsblatt wurde zu spät ausgeteilt und (evtl. deshalb) nicht konsequent ausgefüllt. Zudem waren die Fragen ohne eine intensive Besprechung offenbar recht komplex und die Reflexionsphase, bei der beschrieben und begründet werden sollte, wie der Klassifizierer optimiert wurde (Validierungssituation), war zu kurz angelegt und Begründungen wurden nicht explizit eingefordert. Das führte vermutlich auch dazu, dass in den meisten Gruppen keine iterative Optimierung stattfand (**K4**). Ein Bezug zu den Trainingsdaten (**K2**) passierte in manchen Gruppen nicht. Das könnte daran liegen, dass man die Trainingsdaten während des Testens nicht sieht.

6. Fazit

Anhand der sieben Leitlinien von Hevner et al. (2004) kann das vorliegende DSR-Projekt evaluiert werden.

Leitlinie 1 – Design als Artefakt: das Artefakt, also die entwickelte Lernumgebung ist eine Methode (March & Smith, 1995), also eine Reihe von Schritten, um mit Kindern im Mathematikunterricht ML und algorithmische Verzerrung zu behandeln. Die Lernumgebung war im Laborsetting anwendbar, allerdings muss noch überprüft werden, inwiefern sie im regulären Unterricht eingesetzt werden kann. Optimierungsvorschläge der Lernumgebung sind unter Leitlinie 6 zu finden.

Leitlinie 2 – Problemrelevanz: Das Artefakt ist insofern relevant, als dass es an den in Vartiainen et al. (2024) identifizierten Lücken ansetzt (s. Abschnitt 1). Die Gestaltung des Artefakts beruht auf wissenschaftlichen Erkenntnissen (s. Abschnitte 2 und 3).

Leitlinie 3 – Designbewertung: Das Artefakt wurde im Rahmen einer multiplen Fallstudie mit qualitativen Erhebungs- und Auswertungsmethoden

evaluiert. Als Rahmen für die Unterstützung des didaktischen Designs wurde auf die TDS zurückgegriffen und mit AiC kombiniert, um Lernprozesse auf mikroanalytischer Ebene untersuchen zu können (s. Abschnitte 3, 4, 5).

Leitlinie 4 – Forschungsbeitrag: Der Beitrag gibt Einblicke in das Verständnis von Kindern über die geometrischen Figuren Viereck und Dreieck und die Funktionsweise eines Klassifizierers sowie algorithmische Verzerrung. Die in diesem Beitrag beschriebene Lernumgebung kann von Lehrpersonen (ggf. in angepasster Form) übernommen werden.

Leitlinie 5 – Rigorosität der Forschung: Glaubwürdigkeit („trustworthiness“) wird als wichtigstes Kriterium qualitativer Forschung angesehen (Lincoln & Guba, 1985). Vier Kriterien werden vorgeschlagen: a) Vertrauenswürdigkeit, b) Übertragbarkeit, c) Zuverlässigkeit und d) Bestätigbarkeit (Döring, 2023, S. 108). a) Es wurden Videodaten erhoben sowie die schriftlichen Schülerdokumente. Im Rahmen der a-posteriori-Analysen von AiC und TDS wurden Vorannahmen anhand von Gegenbeispielen in den Daten überdacht; b) Es wurde eine möglichst umfassende Beschreibung der Vorgehensweisen der Lernenden in Abschnitt 4.1 versucht, um eine mögliche Übertragbarkeit auf andere Kontexte zu ermöglichen; c) und d) Im Rahmen dieses Beitrags wird die Vorgehensweise gegenüber Experten detailliert erläutert und begründet.

Leitlinie 6 – Design als Suchprozess: Die Lernumgebung wurde im Juni 2024 mit Dritt- und Viertklässlern erprobt (Makrozyklus) und optimiert und mit zwei Fünftklässlergruppen im Rahmen einer Informatik-AG getestet (Mikrozyklus) und optimiert. In einer neuen Version der Lernumgebung sollte, bevor die Kinder mit dem Tool arbeiten, zunächst gemeinsam mit der Lerngruppe ausgehandelt werden, wann eine Ausgabe des Klassifizierers ‚gut‘ ist (**K5**). Zudem sollte besprochen werden, wie eine KI lernt. Damit können die beiden PaCCs adressiert werden. Evtl. könnte es hier auch hilfreich sein, nur eine Art von Vieleck zu betrachten, sodass die zwei Klassen z. B. ‚Viereck‘ und ‚kein Viereck‘ wären. Zudem wäre es sinnvoll, mindestens zweistufig vorzugehen mit zwei Gruppenarbeitsphasen und zwei Reflexionsphasen. In der ersten Reflexionsphase könnten Schwierigkeiten und Fragen der Kinder gemeinsam geklärt werden, wie die, die in Gruppe 3 aufgetaucht sind, oder Fragen zur Klasseninklusion. Zudem wird so die iterative Optimierung unterstützt (**K4**), die von manchen Gruppen evtl. aus Zeitgründen nicht durchgeführt wurde (s. z. B. Gruppe 4, Abb. 36).

Durch dieses zweistufige Vorgehen kann die Situation wiederverwendet werden und Fragen liefern, die den Prozess wieder in Gang setzen. Auch wenn es materialintensiv ist, sollte ein zweites Tablet pro Gruppe zur Verfügung gestellt werden, sodass Trainingsdaten und Testergebnisse gleichzeitig betrachtet werden können (K2). Die Idee hinter den beiden Tablets muss den Kindern erklärt werden, da dieses In-Beziehung-Setzen nicht selbstverständlich ist.

Leitlinie 7 – Kommunikation der Forschung: Durch den vorliegenden Beitrag wird die Konzeption der Lernumgebung mit wissenschaftlichem Publikum geteilt. Für Parktiker:innen wurde ein Beitrag in einer Praxiszeitschrift erstellt (Platz, 2025).

Abschließend lässt sich der Frage nachgehen, inwiefern die Lernumgebung ihr übergeordnetes Ziel – das Verständnis algorithmischer Verzerrung bei Grundschulkindern zu fördern – erreicht hat. Die Ergebnisse zeigen ein differenziertes Bild: Einzelne Lernende, insbesondere in Gruppe 3, konstruierten zentrale Aspekte des Begriffs. Sie erkannten, dass eine unausgewogene Repräsentation von Figurtypen im Trainingsdatensatz zu einer systematisch einseitigen Klassifikationsleistung führt – und näherten sich damit dem Kern algorithmischer Verzerrung im Sinne von Kordzadeh und Ghasemaghahi (2022). Demgegenüber zeigt PaCC1, dass ein erheblicher Teil der Lernenden die Ursache fehlerhafter Klassifikationen nicht im Datensatz, sondern in menschenähnlichen Eigenschaften der KI verortete. Diese anthropomorphe Deutung stellt eine konzeptuelle Hürde dar, die in einer überarbeiteten Version der Lernumgebung gezielt adressiert werden sollte – etwa durch eine gemeinsame Aushandlung zu Beginn, was ein ML-Modell kann und was es nicht kann (vgl. Leitlinie 6).

Die Integration des ‚Lernens über KI‘ in den Mathematikunterricht der Grundschule könnte über das Thema ML und algorithmische Verzerrung möglich sein, bringt allerdings gewisse Hürden mit sich. Dennoch lohnt es sich, um Kindern einen dauerhaft souveränen und kritischen Umgang mit der sich ständig weiterentwickelnden KI zu ermöglichen und gleichzeitig über mathematische Themen ins Gespräch zu kommen.

Danksagung

Ich danke den gutachtenden Personen für die hilfreichen, konstruktiven Anmerkungen sowie den beteiligten Kindern für ihre hervorragende Mitarbeit.

Zusatzmaterial

Handlungsanweisung zur Bedienung des Klassifizierers (Anhang 1) und Arbeitsblatt (Anhang 2).

Literatur

- Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., ... & Zimmermann, T. (2019). Software engineering for machine learning: A case study. *Proceedings of the 41st International Conference on Software Engineering: Software Engineering in Practice*, 291–300.
- Artigue, M., Haspekian, M. & Corblin-Lenfant, A. (2014). Introduction to the theory of didactical situations (TDS). In A. Bikner-Ahsbahr & S. Prediger (Hrsg.), *Networking of theories as a research practice in mathematics education* (S. 47–65). Cham: Springer.
- Bakker, A. (2018). *Design research in education: A practical guide for early career researchers*. Routledge.
- Bender, E. M. & Hanna, A. (2025). *The AI Con: How to fight big tech's hype and create the future we want*. Random House.
- Bessot, A. (2024). *Einleitung zur Theorie der Situationen*. HAL open science.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Botsch, B. (2023). *Maschinelles Lernen-Grundlagen und Anwendungen*. Springer-Verlag.
- Buolamwini, J. A. (2017). *Gender shades: intersectional phenotypic and demographic evaluation of face datasets and gender classifiers*. Master Thesis, MIT.
- Burger, W. F. & Shaughnessy, J. M. (1986). Characterizing the van Hiele levels of development in geometry. *Journal for research in mathematics education*, 17(1), 31–48.
- Brousseau, G. (2008). Research in mathematics education. In M. Niss (Hrsg.), *Proceedings of ICME-10* (S. 244–254). Denmark: IMFUFA, Department of Sciences, Systems and Models, Roskilde University.
- Currie, P. & Pegg, J. (1998) Investigating students understanding of the relationships among quadrilaterals. In: C. Kanes, M. Goos and E. Warren (Eds), *Teaching Mathematics in New Times, Proceedings of the Annual MERGA-Conference* (S. 177–184).
- de Villiers, M. (1994) The role and function of a hierarchical classification of quadrilaterals, *For the Learning of Mathematics*, 14(1), 11–18.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87.
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. Springer.
- Drechsler, A. & Hevner, A. 2016. A four-cycle model of IS design science research: capturing the dynamic nature of IS artifact design. In: J. Parsons, T. Tuunanen, J. R. Venable, M. Helfert, B. Donnellan & J. Kenneally (Hrsg.), *Breakthroughs and Emerging Insights from Ongoing Design Science Projects: Research-in-progress papers and poster presentations from the 11th DESRIST 2016*. St. John, Canada, 23–25 May (S. 1–8). DESRIST.
- Dreyfus, T. & Kidron, I. (2014). Introduction to abstraction in context (AiC). In A. Bikner-Ahsbahr & S. Prediger (Hrsg.), *Networking of theories as a research practice in mathematics education* (S. 85–96). Springer.

- Dreyfus, T., Herschkowitz, R. & Schwarz, B. (2015). The nested epistemic actions model for abstraction in context: Theory as methodological tool and methodological tool as theory. In A. Bikner-Ahsbahs, C. Knipping & N. Presmeg (Hrsg.), *Approaches to Qualitative Research in Mathematics Education* (S. 185–217). Springer.
- Dwivedi, U., Gandhi, J., Parikh, R., Coenraad, M., Bonsignore, E. & Kacorri, H. (2021). Exploring machine teaching with children. In: *IEEE Symposium on VL/HCC* (S. 1–11). IEEE.
- Eickhoff-Schachtebeck, A. & Strecker, K. (2025). *starkeSeiten. Künstliche Intelligenz und Maschinelles Lernen*. Klett.
- Erez, M. & Yerushalmy, M. (2006). “If you can turn a rectangle into a square, you can turn a square into a rectangle”: young students’ experience the dragging tool, *International Journal of Computers for Mathematical Learning*, 11(3), 271–299.
- Flach, P. (2019). *Machine learning: the art and science of algorithms that make sense of data*. Cambridge University Press.
- Franke, M. & Reinhold, S. (2016). *Didaktik der Geometrie in der Grundschule*. Elsevier, Spektrum, Akad. Verlag.
- Fujita, T. and Jones, K. (2007). Learners’ understanding of the definitions and hierarchical classification of quadrilaterals: towards a theoretical framing, *Research in Mathematics Education*, 9(1&2), 3–20.
- Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M. & Bouchaechera, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37.
- Gravemeijer, K. & Prediger, S. (2019). Topic-specific design research: An introduction. In G. Kaiser & N. Presmeg (Hrsg.), *Compendium for early career researchers in mathematics education* (S. 33–57). Springer.
- Gresse von Wangenheim, C., Hauck, J. C., Pacheco, F. S. & Bertouceli Bueno, M. F. (2021). Visual tools for teaching machine learning in K-12: A ten-year systematic mapping. *Education and Information Technologies*, 26(5), 5733–5778.
- Hastie, T., Tibshirani, R. & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.
- Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly*, 28(1), 75–105.
- Hoi, S. C., Sahoo, D., Lu, J. & Zhao, P. (2021). Online learning: A comprehensive survey. *Neurocomputing*, 459, 249–289.
- Kafai, Y. B., Fields, D. A. & Searle, K. A. (2019). Understanding media literacy and DIY creativity in youth digital productions. *The international encyclopedia of media literacy*, 1-10.
- Kersting, K., Lampert, C. & Rothkopf, C. (2019). *Wie Maschinen lernen*. Springer.
- Kidron, I., Artigue, M., Bosch, M., Dreyfus, T. & Haspekian, M. (2014). Context, milieu, and media-milieus dialectic: A case study on networking of AiC, TDS, and ATD. In A. Bikner-Ahsbahs & S. Prediger (Hrsg.), *Networking of theories as a research practice in mathematics education* (S. 153–177). Springer.
- KMK (2016). *Strategie der Kultusministerkonferenz „Bildung in der digitalen Welt“* (Beschluss der Kultusministerkonferenz vom 08.12.2016).
- KMK (2022). *Bildungsstandards für das Fach Mathematik Primarbereich* (Beschluss der Kultusministerkonferenz vom 15.10.2004, i.d.F. vom 23.06.2022).
- KMK (2024). *Handlungsempfehlung für die Bildungsverwaltung zum Umgang mit Künstlicher Intelligenz in schulischen Bildungsprozessen*. (Beschluss der Kultusministerkonferenz vom 10.10.2024).
- Kordzadeh, N. & Ghasemaghaei, M. (2022). Algorithmic bias: review, synthesis, and future research directions. *European Journal of Information Systems*, 31(3), 388–409.
- Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lane, D. (2024). *How ML for Kids handles numbers models now*. <https://dalelane.co.uk/blog/?p=5119>
- Lincoln, Y. S. & Guba, E. G. (1985). *Naturalistic inquiry*. Sage.
- Mangiante-Orsola, C., Perrin-Glorian, M. J. & Strømskag, H. (2018). Theory of didactical situations as a tool to understand and develop mathematics teaching practices. *Annales de Didactique et de Sciences Cognitives. Revue internationale de didactique des mathématiques*, 145–174.
- Machine Learning for Kids (2017). <https://machinelearningforkids.co.uk>
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision support systems*, 15(4), 251–266
- Miao, F. & Shiohira, K. (2024). *AI competency framework for students*. UNESCO.
- Monaghan, F. (2000). What difference does it make? Children views of the difference between some quadrilaterals, *Educational Studies in Mathematics*, 42(2), 179–196.
- Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- Osterhaus, C., Brandone, A. C., Vosniadou, S. & Nicolopoulou, A. (2021). The emergence and development of scientific thinking during the early years: Basic processes and supportive contexts. *Frontiers in Psychology*, 12.
- Papert, S. (1980). *Mindstorms: Children, computers, and powerful ideas*. Basic Books.
- Payne, B. H. (2019). *An Ethics of Artificial Intelligence Curriculum for Middle School Students*. MIT Media Lab.
- Peffer, K., Tuunanen, T., Rothenberger, M. A. & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45–77.
- Pickreign, J. (2007). Rectangle and rhombi: how well do pre-service teachers know them? *Issues in the Undergraduate Mathematics Preparation of School Teachers*, 1.
- Pielsticker, F. (2020). *Mathematische Wissensentwicklungsprozesse von Schülerinnen und Schülern. Fallstudien zu empirisch-orientiertem Mathematikunterricht am Beispiel der 3D-Druck-Technologie*. Springer Spektrum
- Platz, M. (2025). KI, Internetsuche und algorithmische Verzerrung. *Mathematik differenziert 4-2025*, 22–31.
- Prediger, S., Link, M., Hinz, R.; Hußmann, S., Thiele, J. & Ralle, B. (2012). Lehr-Lernprozesse initiieren und erforschen – Fachdidaktische Entwicklungsforschung im Dortmunder Modell. *MNU* 65(8), 452–457.
- Provost, F. & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.
- Quionero-Candela, J., Sugiyama, M., Schwaighofer, A. & Lawrence, N. D. (2009). *Dataset shift in machine learning*. MIT Press.

- Ron, G., Dreyfus, T. & Hershkowitz, R. (2010). Partially correct constructs illuminate students' inconsistent answers. *Educational Studies in Mathematics*, 75(1), 65–87.
- Rich, E. (1983). *Artificial Intelligence*. McGraw-Hill.
- Russell, S. & Norvig, P. (2020). *Artificial intelligence: a modern approach* (4. Aufl.). Pearson.
- Sanusi, I. T., Oyelere, S. S., Vartiainen, H., Suhonen, J. & Tukiainen, M. (2023). A systematic review of teaching and learning machine learning in K-12 education. *Education and Information Technologies*, 28(5), 5967–5997.
- Schnell, S. (2013). *Muster und Variabilität erkunden: Konstruktionsprozesse kontextspezifischer Vorstellungen zum Phänomen Zufall*. Springer.
- Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., ... & Dennison, D. (2015). Hidden technical debt in machine learning systems. *Advances in Neural Information Processing Systems*, 28.
- Settles, B. (2009). *Active learning literature survey* (No. 1648). University of Wisconsin-Madison Department of Computer Sciences.
- Shamir, G. & Levin, I. (2022). Teaching machine learning in elementary school. *International Journal of Child-Computer Interaction*, 31, 100415.
- Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *Journal of Data Warehousing*, 5(4), 13–22.
- Simon, H.A. (1970). *The Sciences of the Artificial*, MIT-Press.
- SWK (2024). *Large Language Models und ihre Potenziale im Bildungssystem. Impulspapier der Ständigen Wissenschaftlichen Kommission der Kultusministerkonferenz*.
- Tan, Q. & Tang, X. (2025). Unveiling AI literacy in K-12 education: A systematic literature review of empirical research. *Interactive Learning Environments*, 1–17.
- Touretzky, D., Gardner-McCune, C., Martin, F. & Seehorn, D. (2019). Envisioning AI for K-12: What should every child know about AI?. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 33, No. 01, S. 9795-9799).
- Vartiainen, H., Tedre, M. & Valtonen, T. (2020). Learning machine learning with very young children: Who is teaching whom?. *International journal of child-computer interaction*, 25.
- Vartiainen, H., Kahila, J., Tedre, M., López-Pernas, S. & Pope, N. (2024). Enhancing children's understanding of algorithmic biases in and with text-to-image generative AI. *new media & society*.
- Wilke, F. (2018). *Künstliche Intelligenz diskriminiert (noch)*. ZEIT ONLINE.
- Wittmann, E. C. (1995). Mathematics education as a 'design science'. *Educational studies in Mathematics*, 29(4), 355-374.
- Wiyatno, R. R., Xu, A., Dia, O., & De Berker, A. (2019). Adversarial examples in modern machine learning: A review. arXiv preprint arXiv:1911.05268.
- Yin, R. K. (2018). *Case study research and applications*. Sage.
- Zweig, K. (2019). *Ein Algorithmus hat kein Taktgefühl: Wo künstliche Intelligenz sich irrt, warum uns das betrifft und was wir dagegen tun können*. Heyne Verlag.

Anschrift der Verfasserin

Melanie Platz
 Universität des Saarlandes
 Lehrstuhl für Didaktik der Primarstufe – Schwerpunkt Mathematik
 Campus E2 4
 66123 Saarbrücken
melanie.platz@uni-saarland.de