KI-Kolumne

Diversität in Daten?

Von Paola Lopez

Oft wird berichtet, dass es Schieflagen, Bias, in Datensets gibt.[1] Manche Aspekte mancher Gruppen von Menschen werden weniger sichtbar, übertrieben sichtbar oder aber falsch abgebildet. Die KI-Modelle, die auf Grundlage verzerrter Daten gebaut werden,[2] liefern entsprechend problematische Outputs. Ein bekanntes Beispiel ist mangelnde Diversität in den Trainingsdaten von automatisierter Gesichtserkennung.

Das ist mühsam, wenn man per Frontkamera sein Smartphone entsperren möchte, die Kamera das eigene Gesicht aber nicht als Gesicht erkennt – und es kann lebensbedrohlich sein, wenn die Polizei in Ermittlungsverfahren Matching-Algorithmen einsetzt, um ein Gesicht auf dem Bild einer Überwachungskamera mit dem zentralen Register der Führerscheinbilder abzugleichen, und dabei Fehler geschehen.[3] Menschen, die laut KI-Matching ein hinreichend ähnliches Gesicht haben wie der oder die Schuldige,[4] können ohne eigenes Zutun unvermittelt in die Fahndungsmaschinerie geraten. Aufgrund des Bias in den Daten trifft das überproportional viele Menschen mit dunklerem Hautton.

Ich habe darüber geschrieben, dass Sprachmodelle wie ChatGPT zu stereotypen Outputs neigen.[5] Ein Chatbot, den das österreichische Arbeitsmarktservice (AMS) auf seinen Seiten anbietet, hat das exemplarisch gezeigt: Öffentlich zugänglich, soll der »Berufsinfomat« Informationen und Tipps zur beruflichen Orientierung geben.[6] Er soll Ausbildungswege zum Wunschberuf anzeigen und, wenn es keinen Wunschberuf gibt, dabei helfen, einen zu finden. Der Chatbot basiert auf ChatGPT und wurde zusätzlich mit AMS-internem Infomaterial angereichert, um die entsprechenden Informationen in gut verträglicher dialogischer Form den

1 Vgl. Ninareh Mehrabi u.a., *A Survey on Bias and Fairness in Machine Learning*. In: *ACM Computing Surveys*, Nr. 54/6, Juli 2021.

2 Paola Lopez, *Artificial Intelligence und die normative Kraft des Faktischen*. In: *Merkur*, Nr. 863, April 2021.

3 Kashmir Hill, *Wrongfully Accused by an Algorithm*. In: *New York Times* vom 24. Juni 2020.

4 Bei KI-Matchings gibt es de facto nie eine hundertprozentige Übereinstimmung der destillierten Gesichtszüge des Input-Gesichts mit jenen in Bildern aus dem Vergleichsregister. Die gäbe es höchstens dann, wenn das exakt gleiche Bild, das im Vergleichsregister vorliegt, als Input für das Matching geliefert werden würde. Es entscheiden im Voraus bei der Erstellung des KI-Systems festgelegte Schwellenwerte darüber, ab wann ein Input-Gesicht einem Gesicht im Register ähnlich genug ist, um als Ergebnis zu gelten.

5 Paola Lopez, *ChatGPT und der Unterschied zwischen Form und Inhalt*. In: *Merkur*, Nr. 891, August 2023.

6 www.ams.at/arbeitsuchende/aus-und-weiterbildung/berufsinformationen/berufsinformation/berufsinfomat#wien

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH

Nutzerinnen und Nutzern zur Verfügung zu stellen. Am Tag nach der Veröffentlichung tauchten die ersten Screenshots von Outputs des »Berufsinfomat« auf. Einem achtzehnjährigen Mann mit sehr gutem Schulabschluss wurde eine Karriere im IT-Bereich nahegelegt, einer achtzehnjährigen Frau mit sehr gutem Schulabschluss ein Studium der Gender Studies.[7] Es ist wie ein schlechter Scherz. Und es ist das, was die zuständige staatliche Stelle mit staatlichem Auftrag der Öffentlichkeit als Service bereitstellt. Interessant ist, dass offenbar niemand auch nur den basalsten Bias-Test im Vorfeld durchgeführt hat. Stattdessen veröffentlichte man den Chatbot, gab sich am nächsten Tag gekränkt ob des »Hohns«,[8] arbeitete dann aber die unbezahlte Arbeit von Privatpersonen ein und verbesserte das Produkt ein klein wenig durch ex post eingefügte *quick fixes*.[9]

Als Lösung für die Problematik solcher KI-basierten Prognosen, Klassifizierungen, Matchings, Rankings, Chatbots, Bildergeneratoren, etc. werden oft diversere Datensets vorgeschlagen. Wenn in Datensets mehr Diversität vorhanden ist, so das Argument, dann »sieht« das darauf aufbauende KI-Modell auch die volle Heterogenität von Gesichtern, Menschen, Erwerbsbiografien und so weiter und kann sie entsprechend abbilden. Mit »Diversität« meint man dabei nicht irgendeine Verschiedenartigkeit unter Menschen, sondern Kategorien, die man mit gesellschaftlichen Ungleichheiten in Verbindung sieht. Das sind oft Kategorien, die im Antidiskriminierungsrecht als geschützte Merkmale aufscheinen und die sich vergleichsweise gut datafizieren – also in Daten gießen – lassen. Im Folgenden möchte ich diesen Lösungsvorschlag in vier Punkten verkomplizieren.

Erstens: Die gezielte Beschaffung von »diverseren Daten« kann ausbeuterisch sein. Im Jahr 2019 präsentierte Google sein neues Smartphone Pixel 4, zusammen mit einem Werbeclip: In diesem Video entsperrt eine schwarze Frau mit dunklem Hautton in einem dunklen Raum ihr Smartphone per Gesichtserkennung. Der Videoclip ist als direkte Antwort auf die damals schon existierende Bias-Debatte zu verstehen,[10] und Google berief sich ausdrücklich auf die bessere Inklusivität des Verfahrens.[11] Wenn man als Hersteller ein

7 Vgl. Chris Köver, *AMS erntet Hohn mit neuem KI-Chatbot*. In: *Netzpolitik* vom 5. Januar 2024 (netzpolitik.org/2024/diskriminierung-ams-erntet-hohn-mit-neuem-ki-chatbot/).

8 Vgl. ein Posting auf (ehemals) Twitter des Vorstandsvorsitzenden des AMS am 5. Januar 2024 (twitter.com/JohannesKopf/status/1743190500516020644).

9 Privatpersonen investieren immer wieder ihre Zeit und Expertise und zeigen Unzulänglichkeiten von algorithmischen Systemen auf. Institutionen oder Unternehmen übernehmen diese Expertise, arbeiten sie in ihre Produkte ein und proklamieren, sie seien offen für Kritik. Für ein anderes Beispiel eines solchen größtenteils unbezahlten »Crowdsourcing« in großem Maßstab vgl. Paola Lopez, *Power and Resistance in the Twitter Bias Discourse*. In: *Algorithmic Regimes*, Amsterdam University Press, i. E. 2024.

10 Vgl. beispielsweise Joy Buolamwini / Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*. In: *Proceedings of Machine Learning Research*, Nr. 81, 2018.

11 Vgl. Dieter Bohn, *Google's »field research« offered people $5 to scan their faces for the Pixel 4*. In: *The Verge* vom 29. Juli 2019. (www.theverge.com/2019/7/29/8934804/google-pixel-4-face-scanning-data-collection).

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH

KI-System baut, das anhand vieler Daten das Erkennen von Gesichtern lernen soll, und feststellt, dass die Funktionalität für dunklere Hauttöne nicht zufriedenstellend ist, dann braucht man Trainingsdaten von bisher nicht hinreichend repräsentierten Hauttönen. Am besten möglichst schnell, möglichst unaufwändig und möglichst kosteneffizient, denn dieses Unterfangen ist immer noch in die Logiken eines Unternehmens eingebettet.

Google hatte die Datengewinnung in diesem Fall outgesourct und einen Subunternehmer beauftragt, die Daten mittels »field research« zu beschaffen. Im Nachhinein war zu hören, die »field researcher« hätten zweifelhafte Methoden angewendet, um an Daten von Gesichtern mit dunklen Hauttönen zu gelangen. So seien sie nach Atlanta gereist, durchaus gezielt, da hier besonders viele schwarze Wohnungslose leben, und hätten diesen fünf US-Dollar für das digitale Scannen ihres Gesichts angeboten. Und zwar unter dem Vorwand, sie sollten ein »selfie game« testen. Dabei aber seien die Gesichter heimlich gescannt worden.[12]

Es liegt nahe, anzunehmen, dass die Researcher sich von Wohnungslosen weniger Rückfragen und mehr Zusagen erwarteten. Man sieht: Wenn ein Unternehmen Bias in Daten »reparieren« möchte, kann das seinerseits schnell in ausbeuterische Praktiken umschlagen, die genau jene treffen, die bereits benachteiligt sind und deren Benachteiligung das Unternehmen – auf den ersten Blick – zu reduzieren versucht. Entlang der Logik eines Diversitätskapitalismus erweitert Google die Kundschaft – jene, die es sich leisten können – für das eigene Produkt und bemüht dafür die Rhetorik von Diversität und Repräsentation auf Kosten anderer, die sich fast nichts leisten können.

Ruha Benjamin, die zu den Verschränkungen von Technologien mit rassistischen Strukturen arbeitet,[13] sieht diesen Fall als eine Fortführung von historischen Ausbeutungsverhältnissen zwischen wissenschaftlicher Datengewinnung und vulnerablen Gruppen. Sie nennt Beispiele aus der US-amerikanischen Geschichte, in denen die Wissenschaft Informationen und Daten von Menschen unter teilweise furchtbaren Bedingungen gewann: Zum einen berichtet sie von Versklavten, die im 19. Jahrhundert für medizinische Experimente benutzt wurden, zum anderen von Inhaftierten in Vollzugsanstalten, die in den 1970ern für dermatologische Tests rekrutiert wurden. Benjamin: »Google's experiment builds on a long tradition and it might seem not as severe in comparison but it sets a precedent in which we sort of turn a blind eye when it goes after people that are already vulnerable and who are likely to be harmed even when these technologies are developed well.«[14] Der Ruf nach mehr Diversität in Trainingsdaten ist also nicht das Ende der Diskussion über Datenpraktiken, sondern ihr Beginn.

12 Vgl. Sean Hollister, *Google contractors reportedly targeted homeless people for Pixel 4 facial recognition*. In: *The Verge* vom 3. Oktober 2019 (www.theverge.com/2019/10/2/20896181/google-contractor-reportedly-targeted-homeless-people-for-pixel-4-facial-recognition).

13 Vgl. Ruha Benjamin, *Race after Technology: Abolitionist Tools for the New Jim Code*. Medford / Mass.: Polity 2019.

14 Vgl. die Podcast-Folge von *Reset* mit Ruha Benjamin vom 17. Oktober 2019.

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH

Zweitens: Nicht alle wollen immer algorithmisch akkurat »gesehen« werden. Der Versuch, Bias in Daten zu beseitigen, führt schnell zu Daten-Maximalismus. Auf den ersten Blick leuchtet das ein: Man muss so viele Daten inkludieren, dass alle Differenzen möglichst akkurat und möglichst granular vertreten sind. Nur so »sehen« KI-Systeme jede Nische des menschlichen Lebens und funktionieren bestmöglich[15] – und das bestmögliche Funktionieren liegt im Interesse aller. So ungefähr lautet das Argument. Das sind aber voraussetzungsvolle Annahmen, denn Repräsentation ist ambivalent. Was datafiziert wird, ist sichtbar, und gerade bei algorithmischen Systemen gibt es das Potential der Überwachung, das in der Sammlung, Aufbereitung, Speicherung und im Transfer von Daten natürlich grundsätzlich angelegt ist.

Außerdem haben Kategorien eine normierende Wirkung. Um als man selbst in einem System aufscheinen zu können, muss man sich den Datenkategorien dieses Systems fügen. Das gilt so prinzipiell, dass es noch auf das Antidiskriminierungsrecht durchschlägt. Auch Rechtsinstrumente beruhen auf Kategorien, die erst einmal vorgeben, wie oder wer man zu sein hat, um ein Recht durchsetzen zu können. Elisabeth Holzleithner schreibt dazu: »Law shapes the categories that are the object of self-identification and the identification of others. By holding what it means to be discriminated against on the ground of a certain category, the law lends contours to the meaning of the category as such.«[16] Kategorien geben zwar nichts letztgültig vor – aber sie konturieren. Die Zugehörigkeit zu einer Kategorie kann Beschränkung bedeuten oder Befreiung, die Kategorie kann verfehlen oder überschießen. Identifikation ist komplex und kann sich mit der Zeit auch ändern. Daten aber, mögen sie noch so filigran ausdifferenziert sein, fixieren einen Moment wie ein Schnappschuss.

Es ist nicht klar, ob es für eine Person immer das bestmögliche Szenario ist, von einem KI-System möglichst adäquat gesehen zu werden. Interessen variieren, und vor allem auch innerhalb einer »marginalisierten Gruppe« können die Interessen der Einzelnen sehr unterschiedlich sein. Gruppen von Menschen, auch kleine und spezifische, sind nie homogen. Manche Menschen möchten in manchen Konstellationen bestimmte Aspekte ihres Lebens durch Daten adäquat erfasst sehen, manche Menschen nicht.

»Richtige« algorithmische Ergebnisse sind nicht immer das, was Menschen in einem bestimmten Kontext auch möchten. Wenn einer Kreditnehmerin – kor-

15 In der Datenschutz-Grundverordnung ist demgegenüber das Prinzip der Datenminimierung festgelegt. Der Konflikt zwischen Datenminimierung und algorithmischer Performativität ist aber nicht immer so geradlinig: Mitunter wird der Performativitätsgewinn, der aus zusätzlichen Daten stammt, überschätzt. Vgl. Divya Shanmugam u.a., *Learning to Limit Data Collection via Scaling Laws: A Computational Interpretation for the Legal Principle of Data Minimization*. In: *2022 ACM Conference on Fairness, Accountability, and Transparency*. Seoul: ACM 2022.

16 Elisabeth Holzleithner, *Law and Social Justice*. In: Kathy Davis/Helma Lutz (Hrsg.), *The Routledge International Handbook of Intersectionality Studies*. London: Routledge 2023.

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH

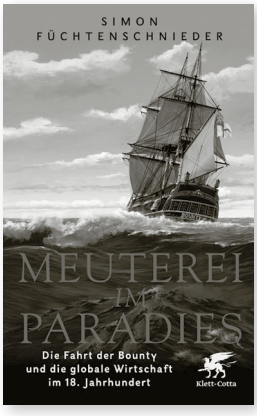
rekterweise – durch ein algorithmisches System eine ungenügende finanzielle Bonität prognostiziert wird, dann liegt das nicht unbedingt in ihrem Interesse, sondern in dem der Kreditgeber. Hier wie zumeist sind KI-Systeme eingebettet in Anwendungskontexte, in denen Interessen quer zueinander liegen. Es ist insgesamt also nur begrenzt sinnvoll, »im Namen von benachteiligten Gruppen« bestimmte algorithmische Systeme in ihrer Funktionalität verbessern zu wollen.

Größer gedacht, kann es mit politischen oder aktivistischen Zielsetzungen, an der Abschaffung eines Systems zu arbeiten, gut zusammengehen, wenn das KI-System schlecht funktioniert. Im Rahmen der Studie *Automating Public Services: Learning from Cancelled Systems* (2022) untersuchte etwa das Data Justice Lab weltweit 61 algorithmische Systeme im staatlichen Sektor, die wieder abgeschafft wurden.[17] Der häufigste Grund für die Abschaffung ist laut der Studie mangelnde Effektivität: Das System hat nicht getan, was es tun sollte. An zweiter Stelle stehen Proteste und Kritik aus der Bevölkerung.

Der Ruf nach mehr Diversität in Datensätzen ist eine Antwort auf die »Wie«-Frage des Einsatzes von KI-Systemen und begünstigt damit eine Diskursverschiebung: Man stellt sich die »Ob«-Frage nicht, sondern arbeitet daran, ein nicht ausreichend funktionales KI-System zu verbessern. Man kann auch ganz grundsätzlich fragen: Ist ein bestimmtes KI-System überhaupt eine gute Idee? Was sind unsere Anforderungen an eine Situation?

17 carnegieuktrust.org.uk/publications/automating-public-services-learning-from-cancelled-systems/

www.klett-cotta.de



Simon Füchtenschnieder
Meuterei im Paradies
Die Fahrt der Bounty und die globale Wirtschaft im 18. Jahrhundert
304 Seiten, gebunden mit Schutzumschlag
ISBN 978-3-608-98773-7
€ 25,– (D) / € 25,70 (A)

Mitreißend schildert Simon Füchtenschnieder die Seefahrt der Bounty, die Meuterei und das Schicksal der Besatzungsmitglieder. Erstmals bettet er diesen sagenhaften Stoff in den seefahrthistorischen und wirtschaftlichen Kontext seiner Zeit ein: Die Geschichte der Bounty war Wegbereiter einer frühen Globalisierung, die unsere Welt bis heute prägt.

Klett-Cotta

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH

Genügt ein KI-System diesen Anforderungen? Welche Probleme löst der Einsatz eines KI-Systems? Und wichtiger: Welche Probleme kreiert er?

Drittens: Es gibt keine richtigen Daten. Man kann nicht nur das allgemeine Interesse an akkurater algorithmischer Darstellung hinterfragen, sondern auch akkurate algorithmische Darstellung an sich. Die Repräsentation eines Phänomens in Daten ist eine sehr spezifische Art der Darstellung. Daten sind, sobald sie maschinell verarbeitet werden sollen, immer quantitativ. Auch Bilddaten oder Textdaten, die a priori keine Zahlen sind, werden in Zahlenform maschinenlesbar gemacht. Es stellt sich also die Frage, was überhaupt quantifizierbar ist und quantifizierbar gemacht werden kann.

Heterogene, unaufgeräumte Lebensphänomene erscheinen in quantifizierter Darstellung messbar, vergleichbar und damit ein Stück weit kontrollierbar. Geht man davon aus, dass Daten Abstraktionen von Lebensphänomenen sind, dann meint man implizit auch, dass es das Echte gibt und das Abstrahierte und dass das Abstrahierte eine notwendige Verkürzung, also eine handliche Version des Echten ist. Das Quantifizieren und in Daten Gießen ist aber ein produktiver Akt, und sozialwissenschaftliche Forschung analysiert schon lange, wie Daten aktiv hergestellt statt nur »gemessen« werden.[18] Auch Kategorisierungen an sich sind produktiv.[19] Die Raster, in denen Daten produziert werden, sind weder zufällig noch naturgegeben, sondern Ausdruck von Denkordnungen. Das Nachdenken und Sprechen über die Realität strukturiert diese erst.

Es gibt nie richtige Daten über das, was uns interessiert. Es gibt nur Annahmen darüber, dass Daten etwas Bestimmtes beschreiben. Was uns tatsächlich interessiert, ist meist komplex: Beispielsweise gibt es keine richtigen Daten über Krankheiten. Es gibt nur Daten über Arztbesuche, über Krankenhausaufenthalte, über Behandlungen, über gezielte medizinische Forschung. Es gibt auch keine richtigen Daten über Betrug bei Sozialleistungen – etwas, das viele Länder gerade interessiert. Es gibt keine Möglichkeit, herauszufinden und festzuhalten, wer auf welche Weise betrügt. Was es gibt, sind Daten über jene Menschen, die geprüft und überführt wurden. Diese sind natürlich immer auch geformt von den jeweiligen Überprüfungspraktiken: Unter dem Begriff »toeslagenaffaire« wurde etwa bekannt, dass der niederländische *Belastingdienst* ein datenbasiertes System zur Risikoschätzung von Betrug beim Bezug von Kindergeld einsetzte, das eine nichtniederländische Staatsangehörigkeit explizit als Risikofaktor nutzte.[20] Die Steuerbehörde musste wegen Verstößen gegen die Datenschutz-Grundverordnung – das Verarbeiten der Nationalität war in diesem Kontext unzulässig – ein Bußgeld in

18 Vgl. Annemarie Mol, *The Body Multiple: Ontology in Medical Practice*. Durham: Duke University Press 2002.

19 Geoffrey C. Bowker / Susan Leigh Star, *Sorting Things Out. Classification and Its Consequences*. Cambridge / Mass.: MIT Press 2008.

20 Vgl. Amnesty International, *Xenophobic machines: Discrimination through unregulated use of algorithms in the Dutch childcare benefits scandal* vom 25. Oktober 2021 (www.amnesty.org/en/documents/eur35/4686/2021/en/).

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH

Millionenhöhe zahlen.[21] Hier gab es wie so oft diskriminierungsbezogene Bedenken, die datenschutzrechtlich ausgehandelt wurden.

Der Wunsch nach tatsächlich repräsentativen Daten, die die ganze Heterogenität der Menschheit akkurat abbilden, enthält immer eine implizite Vorstellung davon, wie und in welcher Weise genau die Welt wirklich heterogen ist. Diese Frage ist alt und kann nicht letztgültig gelöst werden. Das ist an sich nichts Schlechtes und auch kein Grund zur Beunruhigung, sondern eine ganz alltägliche Unzulänglichkeit, mit der sich jede wissenschaftliche Disziplin seit jeher gut arrangiert. Nur vergessen sollte man sie nicht.

Viertens: Auch richtige Daten (die es nicht gibt) können Ungleichheiten verfestigen. Die reine Funktionalität und Genauigkeit von datenbasierten algorithmischen Systemen sagt wenig darüber aus, inwiefern durch ihren Einsatz Ungleichheiten potentiell verschärft werden können. Zu betrachten ist, was das System an seiner Einsatzstelle leisten soll und was mit den Ergebnissen tatsächlich geschieht.

Es gibt zwei konzeptuell unterschiedliche Arten von Bias.[22] Wenn es zu benachteiligenden Ergebnissen kommt, kann das daran liegen, dass die Datenbasis und damit die Ergebnisse systematisch von der Wirklichkeit abweichen, und zwar in einer Weise, die benachteiligte Gruppen unsichtbar oder übermäßig sichtbar macht: Die Daten sind falsch. Verschiedene KI-Systeme zur Gesichtserkennung sind hier das klassische Beispiel.

Die zweite Variante ist, dass gesellschaftliche Schieflagen in Daten erfasst werden, und zwar in adäquater Weise. Die Daten an sich sind nicht »falsch«. Allein die unhinterfragte Verstärkung führt zu Problemen. Angenommen, es wäre erwiesen, dass 95 Prozent aller Frauen nach der Haftentlassung brutale Straftaten begehen. Wir als Gesellschaft haben uns darauf geeinigt, mit Blick auf bestimmte Kategorien nicht zu diskriminieren, und eine davon ist Geschlecht. Wir könnten sie also trotzdem nicht systematisch anders behandeln oder beurteilen. Das hängt zusammen mit der Frage, warum »racial profiling« auch in jenen Fällen falsch ist, in denen man dadurch eine Schuldige findet. Wir haben Anforderungen an Entscheidungsprozesse, die nicht durch »Wissen« ausgehöhlt werden dürfen, das am Outcome orientiert ist.

21 Autoriteit Persoonsgegevens [niederländische Datenschutzbehörde], *Tax Administration fined for discriminatory and unlawful data processing* vom 7. Dezember 2021 (autoriteitpersoonsgegevens.nl/en/current/tax-administration-fined-for-discriminatory-and-unlawful-data-processing).

22 Vgl. Paola Lopez, *Bias Does Not Equal Bias: A Socio-Technical Typology of Bias in Data-Based Algorithmic Systems*. In: *Internet Policy Review*, Nr. 10/4, 2021.

Lizenziert für merkur-zeitschrift.de am 13.12.2025 um 23:05 Uhr

© MERKUR, J. G. Cotta'sche Buchhandlung Nachfolger GmbH