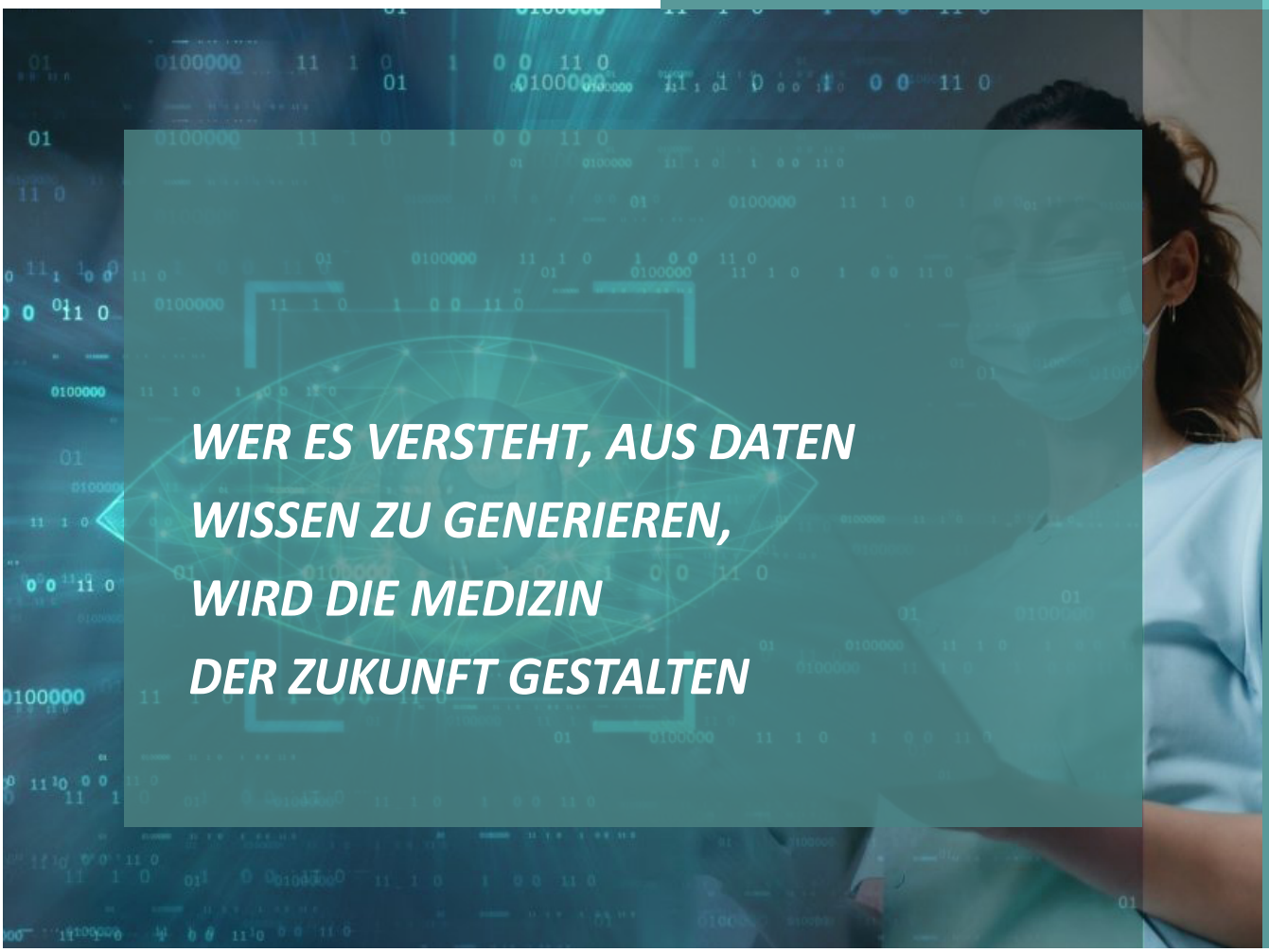
Xplain Data
Discover Causality

VISION PAPER



WER ES VERSTEHT, AUS DATEN WISSEN ZU GENERIEREN, WIRD DIE MEDIZIN DER ZUKUNFT GESTALTEN

VISION4HEALTH – DAS POTENTIAL VON CAUSAL DISCOVERY FÜR DIE MEDIZIN DER ZUKUNFT

Xplain Data GmbH
Grünlandstr. 27
85604 Zorneding
www.xplain-data.com

„EMPIRISCHES WISSEN“ BEDEUTET HEUTE: DATEN

Die Medizin gilt als „empirische“, d.h. auf Erfahrung und Beobachtung beruhende Wissenschaft. Man beobachtet Zusammenhänge und nutzt diese, auch wenn die zugrunde liegenden Prozesse nicht im Detail bekannt sind. Pharmazeutische Unternehmen untersuchen die Wirkung von Arzneimitteln empirisch anhand von Beobachtungen in Interventions- und Kontrollgruppen.

„Erfahrung“ spielt auch in der täglichen Praxis eine zentrale Rolle:

Ein erfahrener (Haus-) Arzt ist besonders gefragt – und in komplexen Fällen konsultieren umsichtige Ärzte ihre Kollegen oder Fachexperten, um von deren Erfahrung zu profitieren.

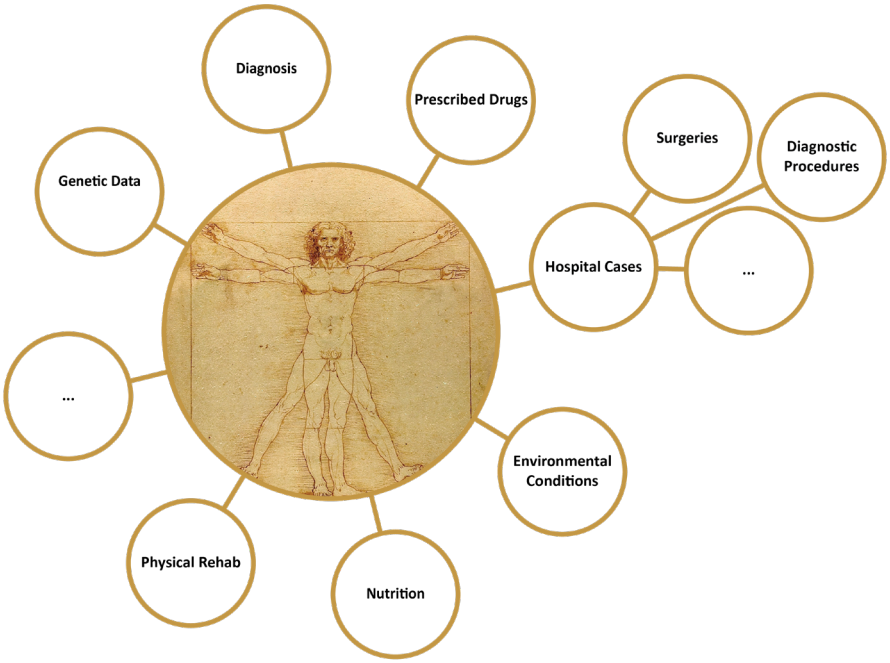
Angenommen, unser Arzt hätte bereits 100 Millionen Patienten behandelt und einen immensen Erfahrungsschatz gesammelt.

Gerade bei schwierigen oder seltenen Krankheitsbildern würden wir uns hier besonders gut aufgehoben fühlen.

Natürlich kann ein einzelner Arzt nicht die Erfahrung aus der Behandlung von 100 Millionen Patienten haben. Aber er könnte das Wissen daraus im unmittelbaren Zugriff haben.

Dieses Wissen ist in den Daten der Patienten implizit enthalten. Nehmen wir an, wir hätten zu 100 Millionen Patienten umfassende medizinische Informationen: alle Symptome, Diagnosen, verabreichte Medikamente, Ernährungsgewohnheiten, sportliche Aktivitäten, beruflicher Hintergrund, soziales Umfeld, genetische Daten, Verlaufsdaten von Laborwerten, MRT- und CT-Scans sowie proteomische Informationen. Hinzu könnten Informationen über medizinische Eingriffe und deren Langzeitfolgen kommen – kurz gesagt: 100 Millionen medizinische Lebensläufe in umfassender Form.

Solche Daten würden ein enormes Reservoir an medizinischem Know-how darstellen - selbst ein Bruchteil davon wäre implizit schon ein enormes Wissen.



DOCH TROTZ GROSSEN POTENZIALS: DATEN ALLEIN SIND NOCH KEIN WISSEN

Die Transformation von Daten in Wissen - in Form allgemeiner Zusammenhänge oder konkreter Einzelfall-Empfehlungen - bleibt eine Herausforderung, bei der wir noch am Anfang stehen.

Wissen aus Daten zu generieren, bedeutet insbesondere, die in den Daten enthaltenen kausalen Zusammenhänge zu verstehen, z.B. die Faktoren, die kausal zur Entwicklung von Multipler Sklerose (MS) oder MS-Schüben führen. MS ist nur eine von vielen Krankheiten, deren genaue Ursachen noch unbekannt sind. Das Verständnis kausaler Ursachen hilft in vielerlei Hinsicht. Es hilft uns, Risikopatienten zu identifizieren. Es hilft, neue Medikamente zu entwickeln, die die kausalen Ketten einer Erkrankung unterbrechen. Und es hilft, intelligente Einzelfall-Interventionen durchzuführen, um den Krankheitsverlauf kausal in die gewünschte Richtung zu beeinflussen.

WISSEN ÜBER KAUSALITÄT IST DIE GRUNDLAGE FÜR INTELLIGENTES HANDELN

Um ein Ziel zu erreichen, müssen wir die Faktoren kennen, die in einem kausalen Zusammenhang zu diesem Ziel stehen. Erst damit können wir intelligente Strategien in Bezug auf ein Ziel entwickeln.

Unter dem Begriff „Causal AI“ entsteht ein neues Feld der Künstlichen Intelligenz, um Ursache-Wirkungs-Zusammenhänge auf der Basis von Beobachtungsdaten zu verstehen. Der Begriff „Causal AI“ wurde erstmals 2022 im *Gartner Hype Cycle for Artificial Intelligence* aufgelistet.

Nur mit sehr umfangreichen und vielfältigen Daten können neue Ursache-Wirkungs-Zusammenhänge aufgedeckt werden. Fehlen Informationen über kausale Faktoren, können bestenfalls indirekte Hinweise gewonnen werden. Darüber hinaus müssen alle Kontextinformationen (sog. Confounder) in die Analyse einbezogen werden. Wir brauchen also ein umfassendes Bild eines Patienten, ermöglicht durch umfangreiche Patientendaten.
Wer über solche Daten verfügt sowie über Algorithmen und Know-how, um aus diesen Daten kausales Wissen zu generieren, wird die Zukunft der Medizin entscheidend mitgestalten.

DIE HERAUSFORDERUNG: DIE KOMPLEXITÄT DER DATEN

Wenn wir von umfassenden Patienteninformationen sprechen, denken wir an eine elektronische Patientenakte. Eine Patientenakte hat hunderte von Tabellen, um die vielen unterschiedlichen Informationen zu speichern. Dieses Datenschema ist äußerst komplex, und gleichzeitig kann das Volumen der Daten sehr umfangreich werden. Bei 100 Millionen Patienten über einen Zeitraum von 10 Jahren ergeben sich ca. 10 bis 30 Milliarden Arzneimittelverordnungen und 50 Milliarden Diagnosen - insgesamt mehrere 100 Milliarden gespeicherte Ereignisse.

DIESE KOMPLEXITÄT PRALLT AUF DIE LIMITS DER HEUTIGEN KI

Wie kann aus solch umfangreichen und komplexen Daten tatsächlich verwertbares empirisches Wissen, insbesondere Kausalwissen, gewonnen werden? Der Weg dorthin ist noch weit. Viele von uns haben bereits etwas Statistik betrieben, basierend auf Daten in einer Tabelle. Tatsächlich gehen viele Verfahren des maschinellen Lernens und der künstlichen Intelligenz davon aus, dass die Daten in einer flachen Tabelle oder ähnlich einfachen Strukturen vorliegen – weit entfernt von der Komplexität der Daten in einer elektronischen Patientenakte.

DIE WELT: IM KLEINEN FLACH, IM GROSSEN RUND

Um neue Horizonte zu erschließen, benötigen wir Technologien, die mit der Komplexität und dem Volumen von Real-World-Daten umgehen können.

Auf dem Weg dorthin hat Xplain Data neue patentgeschützte Verfahren entwickelt. Ohne vorherige Annahmen durchforsten diese Verfahren umfangreiche und komplexe Daten nach potenziellen Ursachen in Bezug auf ein Ziel. Numerisch aufwendige Algorithmen analysieren rückblickend von dem zu verstehenden Ziel-Ereignis (z.B. eine MS-Diagnose) viele Millionen von Faktoren hinsichtlich eines potenziellen Ursache-Wirkungs-Zusammenhangs. Diese Suche geht über banale Korrelationen hinaus. Denn gleichzeitig mit der Vielzahl möglicher Einflussfaktoren werden auch ebenso viele „Kontextvariablen“ (Confounder) berücksichtigt, um zu relevanten Hypothesen für Kausalitäten zu gelangen und ganze Kausalketten aufzudecken.

XPLAIN DATA OBJECTANALYTICS: DAS GROSSE GANZE IM BLICK

Um umfangreiche und komplexe Daten aus so vielen verschiedenen Blickwinkeln durchleuchten zu können, hat Xplain Data mit ObjectAnalytics eine spezifische Form der Datenhaltung entwickelt. Im Gegensatz zu klassischen relationalen Datenbanken, die einen Patienten in seine Entitäten zerlegen und unterschiedliche Teile in unterschiedlichen Tabellen speichern, ermöglicht es Xplain Data ObjectAnalytics, auf alle Informationen eines analysierten Objektes als Ganzes zuzugreifen. Ohne teure Joins zwischen Tabellen können alle zu einem Patienten gehörenden Informationen in Relation zueinander analysiert werden.

Darauf aufbauend führen hochentwickelte Machine Learning-Algorithmen eine kausale Suche durch. Trotz hoher Geschwindigkeit kann eine solche Suche je nach Datenmenge und Hardware mehrere Stunden bis Tage dauern. Um die Leistung vieler Prozessoren zu nutzen, sind diese Algorithmen deshalb für den Einsatz auf großen Rechnersystemen konzipiert.

Diese Verfahren werden bereits in verschiedenen Bereichen eingesetzt. Beispielsweise haben sich Causal Discovery-Algorithmen in der produzierenden Industrie bereits bewährt, um ursächliche Faktoren für Qualitätsmängel oder Maschinenausfälle zu identifizieren. Auch im Gesundheitswesen wird die ObjectAnalytics-Technologie bereits erfolgreich eingesetzt, zum Teil auf sehr großen Datenmengen. Für die Causal Discovery-Verfahren gibt es hier erste Evaluierungen, aber noch keinen systematischen Einsatz.

Was wäre, wenn ...

- die Daten von 30 Millionen Patienten mit allen Ereignissen der letzten 10 Jahre – geschätzt 50 Milliarden Diagnosen, verschriebene Arzneimittel, Laborwerte ...,
- eine Rechnerfarm mit 10 Servern à 128 Cores, auf denen modernste KI-basierte Causal-Discovery-Verfahren implementiert sind,
- ein Team von medizinischen Experten unterschiedlicher Fachrichtung
- unterstützt von einem Data Science-Team zur Verfügung ständen?

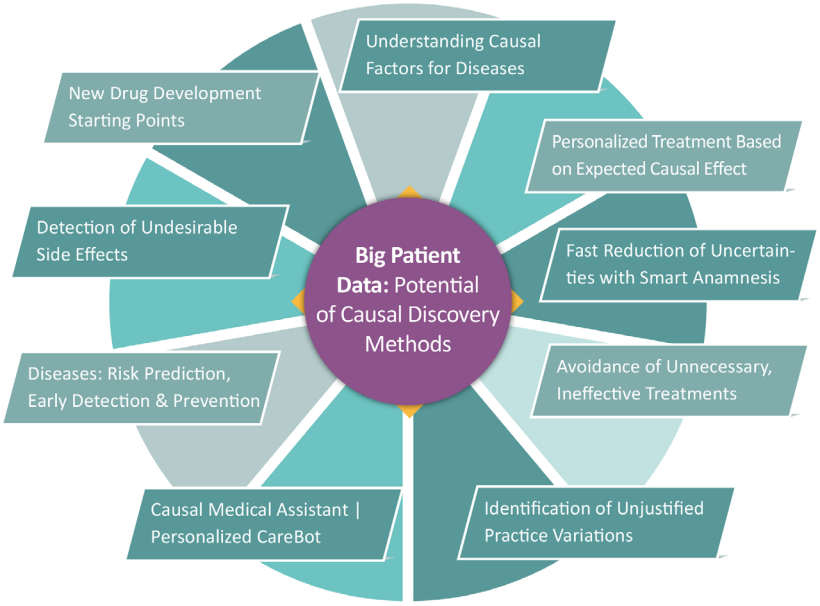
WIR ERWARTEN EINEN QUANTENSPRUNG IN DER MEDIZINFORSCHUNG!

Diese Kombination aus Daten, KI-Verfahren und medizinischer Expertise hat das Potenzial bahnbrechende Entwicklungen in der medizinischen Forschung und der Patientenversorgung in Gang zu setzen.

Es würde nicht lange dauern, bis erste Erkenntnisse die heutige Welt der Medizin überraschen. Es besteht die Chance auf sehr konkrete erste Ergebnisse – z.B. zum Verständnis der Multiplen Sklerose oder anderer Autoimmunerkrankungen.

Aus dem Wissen kausaler Ketten ergeben sich Ansatzpunkte für neue Therapien, die gezielt in die kausalen Ketten eingreifen, die zu einer Erkrankung führen. Causal AI kann ein neues Paradigma „von Daten zu Therapien“ darstellen, das sich auf viele Therapie-Bereiche anwenden lässt.

Über die Entwicklung neuer Therapien hinaus besteht die Chance, Wissen aus Daten unmittelbar an den Point-of-Care zurückzubringen, d.h. jedem einzelnen Patienten und den betreuenden Ärzten an die Seite zu stellen.



Kausale Modelle erlauben Vorhersagen über mögliche Diagnosen, wodurch insbesondere seltene Krankheiten früher erkannt werden können. Sie ermöglichen die Entwicklung von Strategien für intelligente Anamneseprozesse, um Unsicherheiten über bestehende Erkrankungen schnell zu minimieren. Darüber hinaus können sie eine intelligente Auswahl von Therapien unterstützen, basierend auf den erwarteten Ergebnissen oder der erwarteten kausalen Wirkung der Therapie.

Umfangreiche Daten in Kombination mit neuen Causal AI-Algorithmen rücken einen „Causal Medical Assistant“ in greifbare Nähe.

Daten und intelligente Algorithmen sind die Zukunft der Medizin

Wer über umfassende Patientendaten verfügt und die Fähigkeiten entwickelt, aus diesen Daten kausales Wissen abzuleiten, wird die Zukunft der Medizin gestalten.

Die entscheidenden Fragen lauten:

WER WIRD IN DIESEM NEUEN FELD INNOVATIONSFÜHRER SEIN UND WORAUF WARTEN WIR NOCH?

Xplain Data
Discover Causality

Kontakt:
www.xplain-data.com