Künstliche Intelligenz

Bedrohung oder Perspektive?

Göttingen, 13.9.2018

Dr. Andreas Mayert
Sozialwissenschaftliches Institut der EKD (SI)

Gliederung

I. Was ist und kann eine KI

II. Aktuelle Anwendungsbeispiele

III. Technologische Limitationen: Was kann eine KI nicht

IV. Ethische Fragestellungen und Gefahren des Einsatzes von KI

I. Was ist und kann eine KI?

I. Was ist und kann eine KI?

- Es existiert keine einheitliche Definition von Künstlicher Intelligenz
- Sehr weite Definitionen schließen jede Form digitaler Technologien ein, die in bestimmten Bereichen Fähigkeiten besitzen, die über menschliche Fähigkeiten hinausgehen, also z.B. auch Taschenrechner
- Wenn heute über Künstliche Intelligenz diskutiert wird, sind allerdings jene Technologien gemeint, die in den letzten Jahren erhebliche Fortschritte in Bereichen wie Bilderkennung, Texterkennung, natürliche Spracherkennung u.s.w. ermöglicht haben

I. Was ist und kann eine KI?

Diese Fortschritte beruhen zum Teil auf Theorien, die bereits sehr alt sind:

- **„Neuronale Netzwerke“**: Das erste neuronale Netzwerk („Perceptron“) wurde in den **1950er Jahren** entwickelt – die dahinter stehenden Ideen blieben aber aufgrund der rudimentären Computertechnik für Jahrzehnte ohne praktische Relevanz

- „**Maschinenlernen“:** Die Fortschritte in diesem Bereich beruhen auf einem bereits **1986** entwickelten, wegweisenden theoretischen Ansatz (sog. „Backpropagation“ bzw. Fehlerrückwärtssuche), der erstmals zeigte, wie neuronale Netzwerke lernfähig gemacht werden können

- „**Deep Learning**“: 2012 wurde mit der nun ausreichenden Rechnerleistung und unter Verwendung nun vorhandener sehr großer Datensätze praktisch gezeigt, dass sog. tiefe neuronale Netzwerke, die unter Verwendung von „Backpropagation“ trainiert werden, revolutionäre Fortschritte im Bereich Bilderkennung ermöglichen

Wie funktioniert eine solche KI?

Beginnen wir zunächst mit einem Wort, das heute in aller Munde ist: **Algorithmus**. Was ist das eigentlich?

Einfaches Beispiel: Kreditvergabe („Credit Scoring“)

Annahme: Das Risiko, einen Kredit an eine Person zu vergeben, hängt nur von zwei Variablen ab: Vom Einkommen und von den Ersparnissen dieser Person.

Ziel: Auf Grundlage einer Vielzahl von bislang vergebenen Krediten und beobachteten Kreditausfällen berechnen, wann eine Person ein hohes oder ein geringes „Kreditrisiko“ ist.

I. Was ist und kann eine KI?

Wie funktioniert eine solche KI?

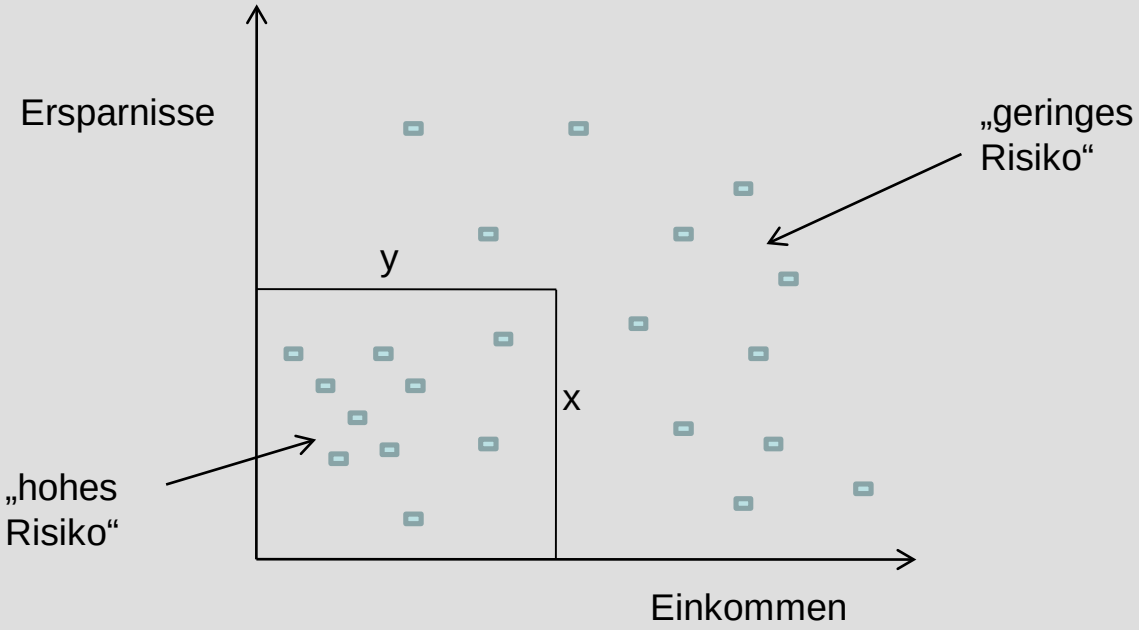
Algorithmus: Im Grunde ist ein Algorithmus eine simple Sache: eine Regel, die verwendet wird, um die Behandlung von Daten zu automatisieren. Wenn Zustand x vorliegt (z.B. ein ausreichendes Einkommen), dann empfehle a (z.B.: Kreditvergabe), wenn nicht, dann empfehle b (z.B. keine Kreditvergabe).

Ein Algorithmus besteht dann aus fixen Befehlen (Wenn-Dann-Regeln) und variablen Parametern (Einkommen, Ersparnisse):

WENN einkommen > x UND ersparnisse > y DANN geringes Kreditrisiko SONST hohes Kreditrisiko

I. Was ist und kann eine KI?

Wie funktioniert eine solche KI?



Ziel: Die KI „lernt“ die Parameterwerte für x und y und passt sie bei neuen Informationen automatisch an

Wie funktioniert eine solche KI?

Basis: Neuronale Netzwerke

- Inspiriert von kognitiven Vorgängen im menschlichen Gehirn
- Das menschliche Gehirn verfügt über ungefähr 100 Milliarden Nervenzellen (**Neuronen**), die über ungefähr 100 Billionen **Synapsen** miteinander verbunden sind.
- Die Neuronen bilden dabei **„Schichten“ (Layer)**, wobei jede Schicht bei kognitiven Prozessen andere, aber miteinander verbundene Aufgaben erledigt
- Es ist noch immer wenig darüber bekannt, wie der menschliche Verstand funktioniert. Nachweisen lässt sich, dass bei kognitiven Vorgängen bestimmte **Neuronen „angeregt“** werden und Informationen an eine Vielzahl weiterer Neuronen über Synapsen weiterleiten, die nun ebenfalls angeregt werden – oder nicht. Die Synapsen bestimmen über die Stärke der Weiterleitung
- Ergebnis (Output) der ablaufenden Prozesse ist z.B. das Erkennen eines Gegenstandes, das Verstehen eines Satzes oder eine gedankliche Schlussfolgerung.

I. Was ist und kann eine KI?

Wie funktioniert eine KI?

Neuronale Netzwerke

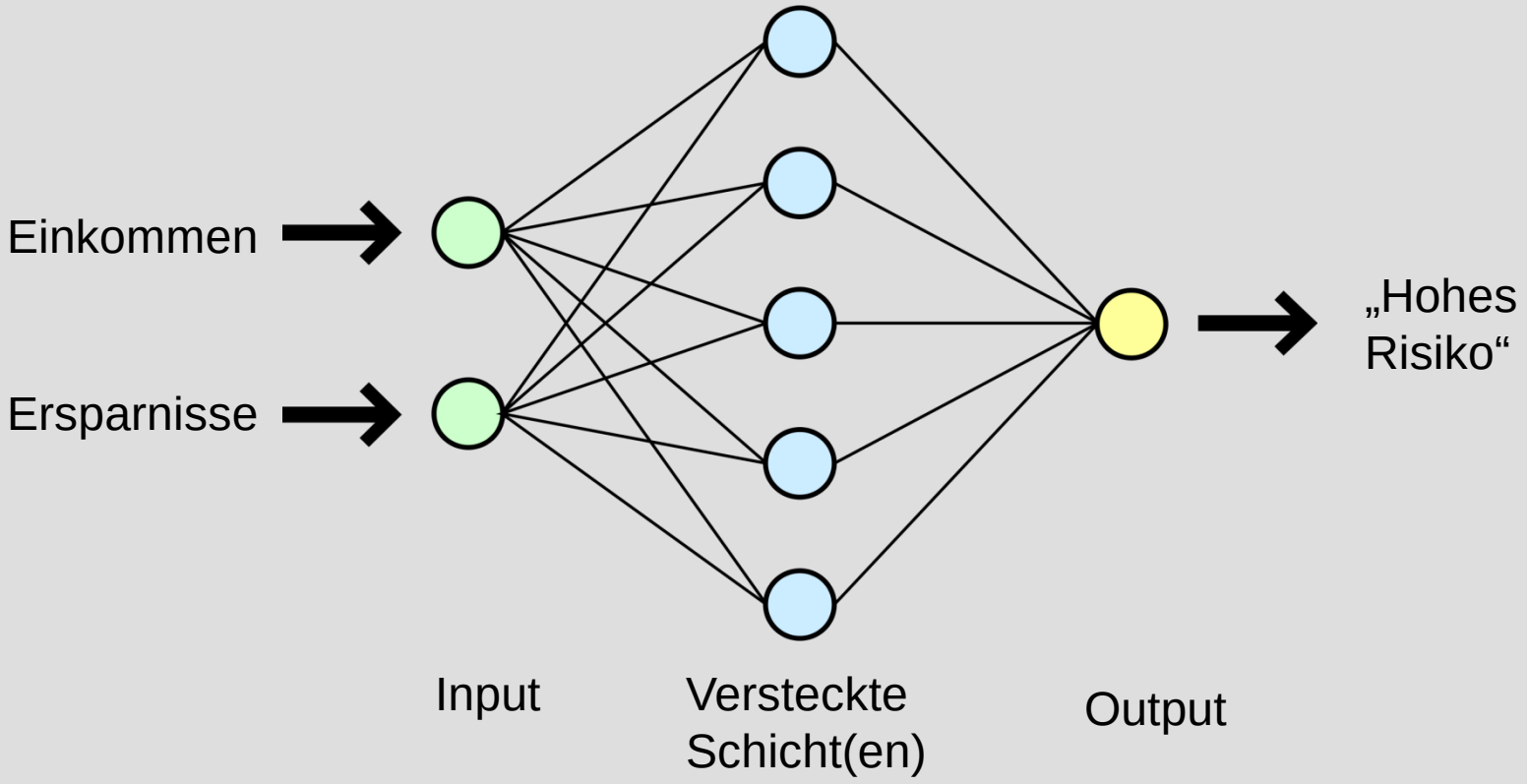
- **Simulierte neuronale Netzwerke** bestehen im Vergleich dazu aus relativ wenigen Neuronen in Form kleiner Recheneinheiten und vergleichsweise ebenso wenigen Synapsen in Form von codierten Verbindungen zwischen diesen Recheneinheiten.

- Sie bestehen zudem nur aus sehr wenigen Schichten. Von „**Deep Learning**“ spricht man, wenn das neuronale Netzwerk über mindestens drei Schichten verfügt.

- Die Anordnung entspricht einem Sandwich: In die unterste Schicht fließen die Informationen (Input), die oberste Schicht liefert das Ergebnis (Output). Die dazwischen liegenden Schichten – auch **versteckte Schichten** genannt – verarbeiten die Informationen im besten Fall so, dass sie zu einem korrekten Ergebnis führen.

I. Was ist und kann eine KI?

Wie funktioniert eine solche KI?



I. Was ist und kann eine KI?

Wie funktioniert eine KI?

Wie „lernt“ die KI?

- Eine KI wird anhand einer Vielzahl von bekannten Input-Output-Beziehungen „**trainiert**“

- **Beispiel**: In einem ersten Schritt werden der KI Informationen über einen Kreditnehmer mit geringem Risiko gegeben. Die Neuronen der ersten Schicht nehmen entsprechende Werte für die Variablen Einkommen und Ersparnisse an.

- Diese Informationen werden über Synapsen, die den Informationen ein bestimmtes Gewicht zuweisen, an die erste versteckte Schicht weitergegeben

- Die Neuronen der ersten Schicht werden angeregt oder nicht. Werden sie angeregt, nehmen sie einen von Null verschiedenen Zahlenwert an, der wiederum über die Synapsen gewichtet an die nächstniedrige Schicht weitergegeben wird, u.s.w.

Wie lernt die KI?

- Die Neuronen der letzten versteckten Schicht liefern schließlich über Synapsen gewichtete Werte an die Ergebnis-Schicht. Da es sich um ein „geringes Risiko“ handelt, sollte das Neuron der letzten Schicht den Wert Null annehmen, das letze Neuron also nicht angeregt werden.

- Nimmt das Neuron der letzten Schicht hingegen einen von Null verschiedenen Zahlenwert an, liegt ein Fehler vor.

- Um zu erkennen, was zu dem Fehler geführt hat, wird rückwärts – d.h. ausgehend vom letzten Neuron – überprüft, wie stark die Synapsen, die zu diesem Neuron führen, zum Fehler beigetragen haben.

- Wenn das getan worden ist, werden alle Verbindungen betrachtet, die in die Neuronen der nächstniedrigeren Schicht führen und ihr Beitrag zum Fehler ermittelt, u.s.w. (Deshalb die Bezeichnung **Backpropagation = Fehlerrückführung**)

- Schließlich ist bekannt, wie viel jede einzelne Verbindung zum Gesamtfehler beigetragen hat und in einem letzten Schritt werden die Gewichtungen der Synapsen so verändert, dass der Fehler insgesamt am besten reduziert wird.

I. Was ist und kann eine KI?

Wie lernt die KI?

- Das Ganze wiederholt man nun mit einer sehr großen Zahl an Beispielen von „guten“ und „schlechten“ Risiken

- Ergebnis ist (hoffentlich), dass die KI am Ende des Trainings mit sehr hoher Trefferquote eigenständig zwischen guten und schlechten Risiken unterscheiden kann, auch wenn die nun eingehenden Informationen nicht Teil des Trainings waren. **Die KI hat „gelernt“ zu klassifizieren**.

- Wird die KI weiterhin mit eingehenden Informationen über Kreditausfälle und Kreditrückzahlungen gefüttert, passt sie ihre Parameter und damit auch ihre Entscheidungen fortlaufend an. Da diese Informationen Folge ihrer eigenen Entscheidungen sind, kann man im übertragenen Sinn, auch sagen, dass die KI „**dazulernt**“

- Das ist natürlich ein sehr simples Beispiel, aber die Vorgänge sind in verschiedenen Anwendungsbereichen ähnlich

II. Aktuelle Anwendungsbeispiele

II. Anwendungsbeispiele

Bilderkennung

- Angenommen, dass Ziel der Bilderkennung besteht darin, dass die KI erkennt, ob sich auf einem Bild irgendwo eine Katze befindet.

- Als Information wird nun eine Bilddatei in das Netzwerk eingespeist, bspw. in der Auflösung 100 x 100. Eine solche Bilddatei enthält 10.000 Pixel und jedem Pixel wird je nach Helligkeit ein Zahlenwert zugeordnet.

- Die **Input-Schicht** enthält insofern **10.000 Neuronen**

- Der weitere Vorgang entspricht, wenn auch wesentlich komplexer, dem bereits geschilderten: Die Neuronen der ersten versteckten Schicht werden angeregt oder nicht und senden über die Synapsen gewichtete Zahlenwerte an die nächste Schicht etc. Am Ende steht das Ergebnis-Neuron, das einen von Null verschiedenen Zahlenwert annehmen kann (Katze vorhanden) oder den Wert Null (Keine Katze). Ist das Ergebnis falsch, erfolgt die Fehlerrückführung.

- Da das Problem komplexer ist, ist auch das Training der KI aufwendiger. **Um Bilderkennung zu optimieren, muss eine KI im Regelfall mit Millionen von Bildern „trainiert“ werden.**

II. Anwendungsbeispiele

Bilderkennung

- Das Erstaunliche an diesem Verfahren ist aber nicht nur, dass eine KI am Ende des Trainings Katzen sehr genau erkennen kann, sondern dass sie im Laufe des Trainings beginnt, so ähnlich zu „sehen“ wie unser visuelles System.

- Das heißt, die erste Schicht lernt im Laufe des Trainings z.B. Kanten zu erkennen, in dem Sinne, dass ihre Neuronen bei Kanten angeregt werden und nicht angeregt werden, wenn es keine gibt; die nächste Schicht lernt Kantengruppen wie z.B. Ecken zu erkennen; die nächste Schicht darüber beginnt, Formen zu sehen und die nächste Schicht ist in der Lage, bestimmte Eigenschaften dieser Formen zu erkennen (z.B.: Größe der Katze)

- D.h.: **Das Netz organisiert sich selbst in hierarchische Schichten, ohne jemals explizit so programmiert worden zu sein**. Es tut das offensichtlich, weil es der effizienteste Weg der Bilderkennung ist.

- Die großen Potentiale der Bilderkennung liegen selbstredend nicht im Erkennen von Katzen. Ein wichtiges Anwendungsfeld ist die medizinische Diagnostik, bspw. in der **Radiologie**. Andere denkbare Anwendungsfelder sind **Überwachung** und **Drohnen**.

II. Anwendungsbeispiele

Texterkennung

- Ähnliche Ergebnisse lassen sich auch in der Texterkennung erzielen
- **Beispiel**: Man „füttert“ ein neuronales Netzwerk mit den gesamten Wikipedia-Einträgen
- Man trainiert die KI darauf, Muster in diesen Einträgen zu erkennen, z.B. das bestimmte Wörter häufig nahe beieinander liegen
- Nach ausreichendem Training ist die KI dann z.B. in der Lage zu erkennen, dass nicht nur Frankreich und Paris in Zusammenhang miteinander stehen, sondern dass dieser Zusammenhang der gleiche ist wie der zwischen Deutschland und Berlin – obwohl der KI nie das Konzept einer „Hauptstadt“ beigebracht worden ist
- Sie versteht das Konzept „Hauptstadt“ natürlich auch weiterhin nicht – sie erkennt nur, dass es einen Zusammenhang zwischen den verschiedenen Begriffen geben muss bzw. dass sie nicht zufällig in Nähe zueinander auftauchen

II. Anwendungsbeispiele

Übersetzung

- Übersetzungs-Software existiert seit langer Zeit – und ihre Ergebnisse waren bis vor wenigen Jahren eher schwach, insbesondere im Bereich Grammatik.

- Die Ergebnisse von Übersetzungen auf Grundlage einer lernenden KI sind dagegen ein Quantensprung, wie jeder z.B. anhand des **Google-Übersetzungsprogramms** feststellen kann.

- Der Vorteil des Einsatzes einer lernenden KI liegt wiederum im Training dieser KI. **Beispiel: Deutsch-Englisch:** Die Trainingsdaten, mit denen eine KI gefüttert werden kann, sind beinahe unerschöpflich. Es liegen Millionen von Büchern, Dokumenten etc. in digitaler Form vor, die vom Englischen ins Deutsche oder umgekehrt übersetzt worden sind.

- Entsprechend kann eine lernende KI auf eine ebenso beinahe unerschöpfliche Zahl von Übersetzungen einzelner Sätze oder Satzbestandteile zurückgreifen, die – da eine Übersetzung nicht zufällig ist – Muster aufweisen

- **Die Übersetzung erfolgt daher deutlich häufiger grammatikalisch korrekt, obwohl die KI selbst überhaupt keine Grammatikregeln beherrscht**. Sie greift lediglich auf die Fülle von bereits erfolgten Übersetzungen gleicher oder sehr ähnliche Sätze/Satzbestandteile zurück – und sie lernt fortlaufend dazu, da die Fülle übersetzter Texte täglich zunimmt.

II. Anwendungsbeispiele

Erstellen eigener Texte

- Technologisch verwandt mit Übersetzungsfähigkeiten ist die – zurzeit noch limitierte – Fähigkeit von KI-Systemen, eigene Texte zu erstellen.

- Die Verwandtschaft mit der Übersetzungsfähigkeit besteht darin, dass eine KI anhand von Beispieltexten darauf trainiert werden kann, bestimmte Ereignisse mit einem generierten Text zu verbinden.

- **Beispiel Börsenberichterstattung**: Anders als bei der Börsenanalyse (z.B. Kommentierung des Inhalts eines Geschäftsberichts, Analyse des Einflusses von unternehmensbezogenen oder politischen Entwicklungen auf Börsenkurse) geht es bei der Börsenberichterstattung lediglich um die zumeist **phrasenhafte** Information über die Entwicklung bestimmter Kurse, über Tagesgewinner und –verlierer etc. Diese Berichterstattung wird heute zum Teil schon von KI‘s übernommen.

- **Beispiel Sportberichterstattung**: Auch hier lassen sich Analysen, Meinungsartikel und umfangreiche Beschreibungen von der reinen Berichterstattung über Ergebnisse unterscheiden, die wie Börsennachrichten aus stets wiederkehrenden Phrasen bestehen. KI‘s werden daher zum Beispiel bei Livetickern verwendet.

- Heute noch mehr als Spielerei werden KI‘s zum Teil auch mit **literarischen Texten** trainiert (z.B. Gedichte einer bestimmten Epoche/eines bestimmten Stils), und liefern dann stilistisch korrekte, aber inhaltlich wertlose **Imitationen**.

II. Anwendungsbeispiele

Autonomes Fahren

- Auch die gängigen KI‘s, die heute im Bereich des autonomen Fahrens genutzt werden, beruhen auf lernenden neuronalen Netzwerken. Ihre Aufgabe besteht darin, die über verschiedene Sensoren (GPS, Kamera, Radar, Lidar) eingehenden Daten zu verarbeiten und daraus folgend richtige Entscheidungen im Straßenverkehr zu treffen.

- Trainiert werden sie mit den Aufnahmen oder Simulationen von Millionen von Verkehrssituationen und den entsprechenden Entscheidungen der menschlichen Fahrer.

- So lernen sie Muster dieser Entscheidungen zu erkennen und Entscheidungen in entsprechenden – nicht simulierten – Situationen auf Grundlage eingehender Sensorwerte selbst zu treffen

- Ein **Problem** des Einsatzes trainierter neuronaler Netzwerke im Straßenverkehr besteht allerdings darin, dass zwar während 99% der gefahrenen Kilometer nur Routineaufgaben zu erledigen sind, für die entsprechend viele Trainingsdaten vorliegen. Während 1% der gefahrenen Kilometer treten aber besondere, zum Teil einmalige Situationen auf, die der Natur der Sache nach nicht trainiert werden können. Vollständig autonomes Fahren scheitert bislang u.a. an einer technischen Lösung für diese Sondersituationen.

II. Anwendungsbeispiele

Kundenanalyse

- **Herkömmliche Kundenanalyseprogramme** arbeiten – ähnlich wie beim Beispiel des Credit Scorings – mit **bekannten Informationen** über den Kunden. Wer zum Beispiel an den bekannten Loyalitäts-Bonus-Systemen im Einzelhandel teilnimmt (Deutschland-Card, etc.), der gibt zur Anmeldung bestimmte persönliche Informationen preis, mit deren Hilfe z.B. bestimmte Kundensegmente identifiziert werden können, um das Angebot zu optimieren.

- Bestimmte sehr persönliche Informationen werden im Regelfall nicht abgefragt, sind aber für verschiedene Informations-Nutzer u.U. von großem Interesse.

- Um an solche Informationen heranzukommen, **können neuronale Netzwerke auch quasi rückwärts arbeiten**. Statt auf Grundlage von bekannten Kundeneigenschaften auf ein bestimmtes Verhalten oder bestimmte Interessen zu schließen, schließen sie vom Verhalten bzw. von den Interessen auf sog. **„versteckte“ Eigenschaften** eines Kunden

II. Anwendungsbeispiele

Kundenanalyse

- **Eher harmloses Beispiel**: Filmempfehlungen bei Netflix oder YouTube

- Eine Filmempfehlungs-KI, wie sie früher üblich war, sammelte die gesehenen Filme von Kunden und sortierte relativ grob. Wer zumeist Horrorfilme gesehen hatte, bekam als Empfehlung weitere Horrorfilme angezeigt

- Das Portfolio der gesehenen Filme, lässt aber – wenn diese nicht völlig zufällig ausgewählt wurden – auch Rückschlüsse auf versteckte Eigenschaften des Nutzers zu, bspw. Bildungsstand, das Vorhandensein von Kindern und ihr ungefähres Alter, Migrationshintergrund, politische Orientierung,…

- Für den Kunden können solche Rückschlüsse durchaus vorteilhaft sein, wenn tatsächlich bessere Empfehlungen abgegeben werden

- Harmlos ist dieses Beispiel allerdings nur dann, wenn entsprechenden Daten nicht weitergegeben werden

II. Anwendungsbeispiele

Kundenanalyse

- **Nicht harmloses Beispiel**: Soziale Netzwerke

- Die große Anzahl von Informationen, die Nutzer bei Facebook oder anderen sozialen Netzwerken hinterlassen, können selbst dann problematisch sein, wenn Nutzer bestimmte sehr persönliche Eigenschaften nicht explizit preisgeben.

- So kann eine KI auf Grundlage von Freunden, eingestellten Fotos, geteilten Videos, Likes, verfassten Texten und darin geäußerten Interessen und Meinungen u.s.w. sehr genau **versteckte Eigenschaften erkennen**: Politische Orientierung, sexuelle Orientierung, Gesundheitszustand, finanzielle Verhältnisse, u.s.w.

- Für soziale Netzwerke sind diese Informationen doppelt interessant, denn sie finanzieren sich (a) durch Werbung und (b) durch Informationshandel. Ist ersteres noch harmlos, gilt dies für den Verkauf „versteckter“ persönlicher Eigenschaften offensichtlich nicht.

III. Technologische Limitationen Künstlicher Intelligenz

III. Technologische Limitationen

- Die großen Fortschritte, die in der Bild-, Text- und Spracherkennung, bei der Übersetzung von Texten und in vielen anderen Bereichen innerhalb weniger Jahre erzielt worden sind, führen häufig zum **Trugschluss, dass eine Künstliche Intelligenz mittlerweile tatsächlich versteht, was sie tut oder wahrnimmt**.

- Dass viele Begriffe, die im Rahmen der KI-Forschung verwendet werden, Analogien zu menschlichem Denken aufweisen (Neuronale Netzwerke, Lernen und „tiefes“ Lernen, etc.) führt zum zusätzlichen **Trugschluss, dass die Weiterentwicklung heutiger KI-Systeme bald zu einer KI führt, die generell – d.h. nicht nur in einzelnen Bereichen – ein menschliches Intelligenzniveau erreicht bzw. übertrifft** („Singularität“).

- Tatsächlich sind die heutigen KI-Systeme aus verschiedenen Gründen von diesen Hoffnungen/Befürchtungen noch Lichtjahre entfernt.

III. Technologische Limitationen

Unterschiede zur menschlichen Intelligenz

- Das häufigste Missverständnis über künstliche Intelligenz beginnt mit dem üblichen Missverständnis über natürliche Intelligenz. Dieses **Missverständnis ist, dass Intelligenz nur eine einzelne Dimension besitzt**.

- Am einen Ende ist die niedrige Intelligenz von z.B. einem kleinen Tier, am anderen Ende ist die hohe Intelligenz von einem Genie - fast so, als ob Intelligenz ein Schallpegel in Dezibel wäre. Natürlich ist es dann einfach, sich eine Entwicklung vorzustellen, bei der die Lautstärke der Intelligenz immer weiter wächst, schließlich unsere eigene hohe Intelligenz übersteigt und zu einer künstlichen Superintelligenz wird.

III. Technologische Limitationen

Unterschiede zur menschlichen Intelligenz

- **Intelligenz hat aber nicht nur eine einzige Dimension**. Wir besitzen verschiedene Formen von Kognition, die für unterschiedliche Formen des Denkens stehen: Deduktion, Induktion, symbolisches Denken, emotionale Intelligenz, räumliche Logik, Kurzzeitgedächtnis und Langzeitgedächtnis, u.s.w.

- **Künstlicher Intelligenzen sind hingegen tatsächlich nur eindimensional**. Sie sind mit hochspezialisierten Organismen zu vergleichen, die eine Sache übermenschlich gut können, aber ansonsten sehr wenig.

- Sie können uns daher in einigen Dimensionen übertreffen, aber keine KI auf Basis der heutigen Techniken (Neuronale Netzwerke, Tiefes Lernen) wird auch nur ansatzweise alles besser können als wir.

- Es ist ähnlich wie mit unseren physischen Fähigkeiten. Während alle Maschinen als Klasse einen Teil der physischen Möglichkeiten eines einzelnen Menschen (Geschwindigkeit des Laufens, Hebekraft, Präzisionsschneiden usw.) übertreffen können, gibt es keine *einzelne* Maschine, die einen durchschnittlichen Menschen auch nur in einem Bruchteil seiner gesamten Fähigkeiten übertreffen kann.

Technologische Limitationen

1. Die gegenwärtigen KI-Systeme sind viel zu Datenhungrig

- Menschen können abstrakte Zusammenhänge in kurzer Zeit lernen
- Beispiel Sprache: Angenommen, wir lernen ein neues Verb kennen und zusätzlich seine Bedeutung, z.B. „klamsen“ = „Auf dem Hintern sitzen“
- Mit dieser Information sind wir sofort in der Lage, das neue Verb in vielen verschiedenen Zusammenhängen zu verwenden und mit anderen Wörtern zu verbinden bzw. solche Verwendungen/Verbindungen zu verstehen:
- „*Sie klamsen jetzt schon seit geraumer Zeit in diesem Raum*“, „*Durch das Herumklamsen der Mitarbeitenden geht viel Arbeitszeit verloren*“, „*Die Stühle sind so hart, ich habe mich bald wundgeklamst*“
- Wenn wir lernen, was klamsen ist, brauchen wir nicht Millionen von Trainingsbeispielen. Wir besitzen die Fähigkeit, ein neues Verb sofort mit allem bisher sprachlich Gelerntem in Verbindung zu bringen.

Technologische Limitationen

1. Die gegenwärtigen KI-Systeme sind viel zu Datenhungrig

- KI‘s fehlt derzeit ein Mechanismus zum Lernen von Abstraktionen durch explizite verbale Definitionen. Sie funktionieren am besten, wenn sie mit Hunderttausenden, Millionen oder sogar Milliarden von Trainingsbeispielen gefüttert werden.

- Menschen sind offensichtlich weitaus effizienter im Lernen abstrakter Regeln als eine KI.

Technologische Limitationen

2. Die gegenwärtigen KI-Systeme sind trotz der Bezeichnung „Deep Learning“ im Grunde das Gegenteil: Flach

- Obwohl „Deep Learning“ bzw. „Tiefes Lernen“ zu einigen erstaunlichen Dingen fähig ist, muss man sich bewusst sein, dass sich das Wort "tief" allein auf eine technische Eigenschaft (die Anzahl der versteckten Schichten in neuronalen Netzen) bezieht, nicht auf eine konzeptionelle Eigenschaft.

- Wenn wir im landläufigen Sinne tiefes Denken beschreiben sollten, würden wir es z.B. mit einem gedanklichen Versunkensein in philosophischen oder ethischen Fragestellungen gleichsetzen.

- Ein solches „Nachdenken“ ist einer KI fremd. Sie versteht auch nicht, was abstrakte Konzepte wie Gerechtigkeit, Menschenrechte oder Ähnliches bedeuten. Wie auch? Es sind soziale Konstruktionen. Menschenrechte existieren nur, weil wir an sie glauben. Aber eine KI glaubt nicht, sie findet Muster in Datensätzen und zieht eindeutige (aber nicht unbedingt korrekte) Schlussfolgerungen.

Technologische Limitationen

3. Die gegenwärtigen KI-Systeme scheitern häufig an Problemen, die mit gesundem Menschenverstand sehr einfach lösbar sind

- Probleme, die weniger mit Mustererkennung zu tun haben als mit dem gesunden Menschenverstand liegen zumeist außerhalb des Rahmens dessen, wofür eine KI brauchbar ist.

- Beispiel: „*Wer ist größer, Prinz William oder seine dreijährige Tochter Charlotte?*“ Jeder durchschnittliche Mensch kann diese Frage ohne Notwendigkeit eines speziellen „Trainings“ beantworten und auch ohne je von den beiden Personen gehört zu haben.

- Die bereits angesprochene, mit allen Wikipedia-Artikeln gefütterte KI, die scheinbar das Konzept einer Hauptstadt erkannt hat, scheitert an solchen Fragen. Der Begriff der Körpergröße findet sich offenbar nicht häufig genug in der Nähe der beiden Personen. Die KI ist möglicherweise zu der Schlussfolgerung fähig, dass Erwachsene im Regelfall größer sind als Dreijährige – aber der Begriff Erwachsener taucht in der Frage nicht auf – und da nicht nur Erwachsene Töchter haben können, kann sie auch nicht auf einem Umweg zur Lösung des Problems gelangen.

Technologische Limitationen

4. Eine KI erkennt nur Verbindungen zwischen Daten, kann aber nicht auf Kausalität schließen

- **Beispiel**: Eine KI kann leicht erlernen, dass in der Bevölkerung Deutschlands die Körpergröße von Menschen und ihr Wortschatz in Zusammenhang stehen. Sie versteht aber nicht, dass dieser Zusammenhang nur deshalb besteht, weil Kinder i.A. kleiner sind und einen geringeren Wortschatz besitzen als Erwachsene.

- Da eine KI keine Kausalität versteht, könnte sie zudem nicht schlussfolgern, ob Kinder im Laufe ihres Wachstums mehr Wörter lernen, oder ob das Lernen von Wörtern Kinder wachsen lässt.

III. Technologische Limitationen

Der Weg zur Überwindung dieser Limitationen ist lang

- Auch wenn die heutigen Künstlichen Intelligenzen weit entfernt davon sind, dem menschlichen Verstand in all seinen Dimensionen auch nur nahe zu kommen, heißt das nicht unbedingt, dass eine KI niemals dazu in der Lage sein wird

- Im Laufe der Zeit wurden verschiedenste Tests vorgeschlagen, deren Ziel darin besteht, die Nähe einer KI zum menschlichen (oder überhaupt einem) Verstand festzustellen. Der bekannteste ist der **Turing-Test**

- Bei diesem Test führen Menschen eine Unterhaltung mit einer KI – also im Grunde mit einem Chatbot, der ähnlich wie wie Siri oder Alexa funktioniert. Merken sie nicht, dass sie sich mit einer KI und nicht mit einem anderen Menschen unterhalten, gilt der Turing-Test als bestanden

- Mittlerweile existieren lernende Chatbots, die mit tatsächlichen menschlichen Konversationen trainiert werden, um „echt“ wirkende Antworten zu geben.

Technologische Limitationen

- Es dürfte nur eine Frage der Zeit sein, bis die schiere Menge an Konversationsdaten dazu führt, dass diese Chatbots Menschen tatsächlich täuschen können, auch wenn es heute noch nicht möglich ist, mit Chatbots wie Siri oder Alexa tatsächlich eine gehaltvolle Unterhaltung zu führen.

- Doch selbst wenn das möglich wäre, besäßen diese Chatbots weiterhin keinerlei Verstand. Ähnlich wie lernende Übersetzungsprogramme, greifen sie lediglich auf den Fundus geführter Unterhaltungen zurück, wenn sie antworten. Sie erkennen Muster in Unterhaltungen und wählen als eine Antwort auf eine Äußerung ihres Gegenübers eine Formulierung aus, für die sie einen Zusammenhang mit der Äußerung erkannt haben.

- **Der Turing-Test ist daher nicht geeignet, auf den tatsächlichen Verstand einer KI zu schließen**.

- Im Laufe der Zeit wurden daher andere Testverfahren entwickelt. Sie zeigen, wie weit die heutigen KI-Systeme noch von einem tatsächlichen Verstand entfernt sind.

Technologische Limitationen

Beispiel 1: **Gesunder-Menschenverstand-Test**

Bei diesem Test werden einer KI Sätze mit einer nicht eindeutigen Aussage vorgelegt und anschließend eine Verständnisfrage gestellte:

„*Die Mitglieder des Stadtrates verweigern den Demonstranten eine Genehmigung durch die Innenstadt zu marschieren, weil sie Gewalt befürchten*.“

Frage: Wer befürchtet Gewalt?

„*Die Mitglieder des Stadtrates verweigern den Demonstranten eine Genehmigung durch die Innenstadt zu marschieren, weil sie eine gewaltbereite Gruppe sind*“

Frage: Wer ist eine gewaltbereite Gruppe?

Technologische Limitationen

Mit gesundem Menschenverstand sind beide Fragen leicht zu beantworten, weil wir um die üblichen Gegebenheiten in der realen Welt wissen und der Begriff „Genehmigung“ darauf hindeutet, dass sich der Grund der Verweigerung auf die Demonstranten bezieht.

Für eine KI ist eine Beantwortung weit schwieriger, es sei denn sie ist mit ähnlichen Situationen gefüttert worden, was beim Common-Sense-Test jedoch verboten ist – schließlich will man das Verständnis der KI für Situationen in der realen Welt testen, nicht ihre Fähigkeit zur Mustererkennung.

2016 wurde der Test erstmals mit einer Vielzahl von KI‘s durchgeführt, wobei insgesamt 60 Aussage-Fragen-Paare in den Test eingingen. Als Bestanden galt der Test bei einer Trefferquote von 90%.

Ergebnis: Der Sieger erreichte eine Trefferquote von 58%, was kaum von dem Wert abweicht, die durch reines Raten erzielt werden kann (50%).

Technologische Limitationen

Beispiel 2: KI-IQ-Test

Eine KI muss im Laufe dieses Tests eine Vielzahl von sehr verschiedenen Problemstellungen lösen, so dass der Test gängigen IQ-Test-Verfahren bei Menschen ähnelt. Aufgaben sind z.B.:

– Bilderkennung mit wachsendem Schwierigkeitsgrad
– Simultanübersetzung eines gesprochenen französischen Textes ins Englische
– Den Inhalt einer abgespielten Sprachdatei zusammenfassen
– Die Story eines gesehenen Films wiedergeben
– Ein Diagramm auf Grundlage einer verbalen Beschreibung zeichnen

Ergebnis: Einige KI-Systeme zeigten gute Ergebnisse in Bilderkennung und Simultanübersetzung. Die schwierigeren Aufgaben konnten erwartungsgemäß nicht erfüllt werden. Wann sie erfüllbar sein könnten, wird von den Initiatoren des Tests als „**zurzeit noch im Bereich von Science Fiction**“ beschrieben.

IV. Ethische Fragestellungen und Gefahren des Einsatzes von KI

Autonomes Fahren

- Auch wenn vollautonomes Fahren die Zahl von Verkehrsunfällen drastisch senken kann, sind sie nicht völlig ausgeschlossen

- Dies gilt umso mehr, weil vollautonome Fahrzeuge, wenn sie denn irgendwann am Straßenverkehr teilnehmen, mit herkömmlichen LKW und PKW sowie mit Radfahrern und Fußgängern konfrontiert werden, deren Verhalten bekanntlicherweise keinem eindeutigen Muster folgt und ab und zu völlig irrational sein kann.

- Kommt die KI des autonomen Fahrzeugs nun in einer bestimmten gefährlichen Fahrsituation zu der Schlussfolgerung, dass es **unvermeidlich bei jeder ihrer Entscheidungsvarianten einen Unfall mit Personenschaden geben wird, nach welchen Maßstäben soll sie sich dann verhalten?**

- Soll sie primär ihre Fahrzeuginsassen schützen? Soll sie die Zahl der Opfer minimieren, auch wenn ihre Insassen dann zu Opfern werden? Soll sie zwischen potentielle Opfern unterscheiden, indem sie z.B. Kinder bevorzugt schützt? Soll sie, um ihr jeweiliges Ziel zu erreichen, auch ungesetzlich handeln können, z.B. indem sie zur Minimierung der Opferzahl absichtlich auf den Bürgersteig oder in den Gegenverkehr fährt?

Autonomes Fahren

- Diese Fragen sind **aus ethischer Sicht nicht leicht zu beantworten**

- Sie sind zudem – neben noch vorhandenen technologischen Problemen – ein Grund dafür, dass vollautonome Fahrzeuge bislang außer zu Testzwecken nicht eingesetzt werden. Denn **neben den ethischen Fragen bestehen auch Haftungsfragen**.

- **Beispiel**: Ein Automobilhersteller haftet heute dafür, wenn ein Produktfehler zu einem Unfall geführt hat. Wurde der Produktfehler nicht wider besseren Wissens im verkauften PKW belassen, lässt sich zugunsten des Herstellers annehmen, dass er den Unfall und den dadurch entstandenen Schaden nicht mit Vorsatz in Kauf genommen hat.

- Ist aber ein **vollautonomes Fahrzeug** an einen Unfall beteiligt und **wählt aufgrund einer ihm vorgegebenen Norm „Minimierung der Opferzahl“ bestimmte Opfer aus**, so werden diese mit **Vorsatz** zu Unfallopfern. Möglicherweise handelt es sich dabei um ansonsten völlig Unbeteiligte, d.h. sie tragen keine Mitschuld am Unfall. Die Haftungsfrage spielt dann in einer ganz anderen Liga, denn nun geht es zusätzlich um strafrechtliche Normen wie **Totschlag oder vorsätzliche Körperverletzung**.

Diskriminierung durch KI-Entscheidungen

- In vielen Bereichen, bspw. bei der Einstellung eines Bewerbers, der Vermietung eines Hauses, dem Verkauf von Versicherungen oder der Gewährung von Krediten ist Diskriminierung aufgrund verschiedener Merkmale (z.B. Migrationshintergrund, Geschlecht, sexuelle Orientierung) explizit verboten.

- Wir wissen, dass bei solchen Entscheidungen, wenn Menschen sie treffen, dennoch häufig diskriminiert wird, das ändert allerdings nichts am Verbot.

- Werden solche Entscheidungen **an eine KI delegiert, könnte Diskriminierung im Idealfall gesenkt werden**, wenn rechtlich ausgeschlossen wird, dass Informationen über die entsprechenden Merkmale in die KI eingespeist werden. Sie wären dann keine **„bekannten Eigenschaften“**.

- Nun haben wir aber gelernt, dass eine Stärke von KI-Systemen darin besteht, **„versteckte Eingenschaften“** über Personen herauszufinden.

Diskriminierung durch KI-Entscheidungen

- **Beispiel**: Bei einer **Hausvermietungsentscheidung**, die eine größere Maklerfirma einer KI überträgt, wird nicht explizit nach dem Migrationshintergrund des Bewerbers gefragt und der KI wird auch kein Name des Bewerbers mitgeteilt, auf dessen Grundlage sie einen Migrationshintergrund annehmen könnte.

- Gehen nun aber weitere völlig legitime Informationen in die KI ein, wie z.B. der bisherige Wohnort, die Art der Beschäftigung, der Grund des angestrebten Umzuges, die Art der Kreditsicherheiten, die Zahl der Kinder etc., kann eine KI zwar nicht auf den Migrationshintergrund des Bewerbers schließen.

- Ist aber z.B. der Migrantenanteil in bestimmten Wohnorten, Berufsgruppen und Familienkonstellationen hoch und sind bisherige Mieter aus gleichen Wohnorten, Berufsgruppen und Familienkonstellationen öfter bei der Miete rückständig, wird die KI die Schlussfolgerung ziehen, dass der Bewerber abgelehnt werden sollte.

- Die KI hat somit statt des nicht abgefragten Migrationshintergrundes selbst einen Ablehnungsgrund entwickelt, der von einer Ablehnung aufgrund des Migrationshintergrundes im Ergebnis nicht unterscheidbar ist.

Diskriminierung durch KI-Entscheidungen

- **Es handelt sich insofern um eine eigentlich verbotene Diskriminierung**

- **Aber wie soll sie nachgewiesen werden?** Das Maklerbüro kann darauf verweisen, alles getan zu haben, um eine solche Diskriminierung zu verhindern. Und die KI wird ihre Entscheidung nicht mit dem Migrationshintergrund begründen – sie hat einfach nur Muster in den Datensätzen erkannt und ein hohes Risiko eines Mietausfalls berechnet. Wie kann der Bewerber dann nachweisen, dass er aufgrund seines Migrationshintergrundes diskriminiert worden ist?

- Das geschilderte Problem ist nicht nur deshalb hochrelevant, weil es zu Diskriminierung führt. Seine besondere Relevanz besteht darin, dass es der auch ohne KI zu beobachtenden Diskriminierung einen weiteren Faktor hinzugefügt hat: **Die Diskriminierung ist nicht nachzuweisen, weil die KI ja tatsächlich nicht absichtlich aufgrund des Migrationshintergrundes diskriminiert**, sondern durch die Erkennung von Mustern.

Diskriminierung durch KI-Entscheidungen

- Es gibt verschiedene **Vorschläge, dieses Problem zu lösen**. So könnte ein „**Recht auf Erklärung**“ eingeführt werden, d.h. dem Bewerber muss mitgeteilt werden, aus welchen Gründen an ihn keine Wohnung vermietet wird. Aus den geschilderten Gründen ist das aber schwer handhabbar.

- Ein anderer Ansatz setzt auf eine **Kontrolle der Entscheidungen einer KI**. Im vorliegenden Beispiel müsste das Maklerbüro zum Beispiel eine Statistik ihrer Vermietungsentscheidungen vorlegen. Wenn sich daraus ableiten lässt, dass der Einsatz der KI zu Entscheidungen führt, die bei ansonsten ähnlichem sozioökonomischen Status Bewerber mit Migrationshintergrund signifikant häufiger ablehnt als Bewerber ohne Migrationshintergrund, müsste eine Kontrollbehörde den Einsatz der KI aufgrund nachgewiesener Diskriminierung verbieten.

Schutz der Privatsphäre

- Während beim Thema Schutz der Privatsphäre zumeist auf die soziale Medien und ihre Möglichkeiten des Datensammelns verwiesen wird, möchte ich ein Unternehmen in den Mittelpunkt rücken, das kaum bekannt ist, dessen Praktiken aber auf die größte Gefahr für den Schutz der Privatsphäre hinweisen: **Axciom**

- Dieses Unternehmen ist dem Bereich der sog. **Datenbroker** zuzurechnen, Unternehmen, dessen Geschäftszweck im Sammeln, Analysieren und Verkauf von personenbezogenen Daten besteht. Axciom ist in Besitz des weltweit größten kommerziellen personenbezogenen Datensatzes und operiert auch in Deutschland.

- Das Unternehmen sammelt, kombiniert, analysiert und verkauft sensible personenbezogene Daten aus einer Reihe von Quellen, einschließlich **öffentlicher Aufzeichnungen, Umfragen und Fragebögen, Einkäufen im Einzelhandel, Web-Browsing-Cookies und Social-Media-Postings**. Da das Unternehmen auch im Bereich Geomarketing aktiv ist, ist anzunehmen, dass es auch im **Besitz von Bewegungsdaten** der Nutzer von Smartphones oder sonstiger Produkten ist, die immer online sind, und auch diese Daten mit anderen verknüpft.

Schutz der Privatsphäre

- Das Analyseteam des Unternehmens hat ein KI-Modell mit entwickelt, das **bis zu 10.000 mögliche Eigenschaften von Personen identifizieren kann**. „Mögliche Eigenschaften“, weil bestimmte Eigenschaften in Form von „Scores“ angegeben werden, die als Wahrscheinlichkeit interpretiert werden können, dass eine Person die betreffende Eigenschaft besitzt.

- Um einen Eindruck zu gewinnen, wie intim diese Eigenschaften zum Teil sind: Es existiert z.B. ein Score für das Vorliegen einer Erektilen Dysfunktionsstörung.

- **Die Gefahr** für unsere Privatsphäre besteht daher nicht vorwiegend darin, auf einer einzelnen Seite wie Facebook persönliche Informationen zu hinterlassen. Sie **besteht darin, dass wir an sehr vielen Orten Informationen hinterlassen und KI-Systeme in der Lage sind, sie so zu verknüpfen, dass auf einzelne Personen und eine Unmenge von persönlichen Eigenschaften geschlossen werden kann**.

- Weil diese Datenmenge mit jedem Tag wächst, werden diese Schlussfolgerungen im Zeitverlauf immer genauer und inhaltlich immer umfangreicher werden.

Schutz der Privatsphäre

- **Was geschieht mit diesen Daten?** Axciom verweist darauf, dass Käufer sie nur zu Werbezwecken einsetzen. Andererseits war Axciom in den Skandal um Cambridge Analytica verwickelt: Die Daten wurden also auch zur zielgerichteten Wählerbeeinflussung genutzt

- Eine andere Zweckverwendung lässt sich daher auch bei anderen Käufern der Daten nicht ausschließen. Warum sollten z.B. Unternehmen die kostspieligen Daten nicht nutzen, um ihre Belegschaft oder Job-Bewerber zu durchleuchten? Was würde ein autoritärer Staat mit solchen Daten anfangen?

- Ich sehe hier **zurzeit die größte Gefahr, die von KI-Systemen in Verbindung mit einem wachsenden Datenvolumen über beinahe jeden Bürger ausgeht**, zumal unsere Datenschutzgesetze den technologischen Möglichkeiten weit hinterherhinken.

Vielen Dank für ihr Kommen, ihre Geduld und ihre Aufmerksamkeit!