

Bürokratische Entscheidungsfabriken

Wegen des nahegelegenen Frankfurter Flughafens fallen beim Amtsgericht Frankfurt 15.000 Fälle pro Jahr an, bei denen Fluggäste gegen Fluggesellschaften klagen. Mehr als die Hälfte aller Zivilfälle in Frankfurt sind solche Fluggastfälle. Die dafür nötigen Ressourcen fehlen an anderer Stelle. Weil die Rechtslage eigentlich klar ist, sollten viele dieser Fälle gar nicht erst vor Gericht landen (*no pun intended*). Bei einem Flugausfall muss die Airline meist eine Entschädigung zahlen. Aber manche Fluggesellschaften spekulieren darauf, dass Gäste ihre Rechte nicht kennen, einen Rechtsstreit scheuen oder einfach entscheiden, dass die Entschädigung den Stress nicht wert ist. Die Summe aller nicht gezahlten Entschädigungen ist so groß, dass es mittlerweile Online-Portale gibt, die gutes Geld damit verdienen, dass man seinen Fall an sie abtritt. Gegen Erfolgsprovision setzen diese Online-Portale die Fluggastrechte für einen durch. Ganz ohne Stress. Das rechnet sich für die Online-Portale, weil die Fälle sich leicht standardisieren und automatisiert verarbeitet lassen. Das ist gut für die Kunden, die dadurch leichter zu ihrem Recht kommen. Dadurch steigt aber die Anzahl der Klagen bei den Gerichten weiter.¹

Richterinnen und Richter können sich sicher auch spannendere Fälle vorstellen. Um sie von dieser nervigen Aufgabe zu entlasten, schlägt der Deutsche Richterbund vor, KI zu nutzen. In Zukunft könnten also in Standardfällen Computersysteme für Anwälte automatisch Klagen einreichen, die dann Computersysteme für Richter automatisch bearbeiten. Glücklicherweise sind die deutschen Richter nicht so naiv zu glauben, dass sich alle ihre Probleme so lösen lassen. Stattdessen schlagen sie ein ganzes Maßnahmenpaket vor, in dem – neben einer

1 Dursun & Stradinger (2022) berichten über die Klagewelle durch Flugverspätungen und die Überlastung der Gerichte.

besseren Personalausstattung und Änderungen der Prozessordnung – Automatisierung und KI nur ein Teil der Lösung sind. Ein Urteil bilden wollen die Richter sich noch selber. KI soll nur unterstützen und die Effizienz erhöhen.²

Im Fluggastrecht treten viele ähnliche Fälle auf. Diese Fälle zu standardisieren und automatisch abzuarbeiten, ist technisch relativ leicht zu bewerkstelligen. Standardisierung führt zu mehr Effizienz in der Verwaltung. Ein Gericht, das überlastet ist und seinen Aufgaben nicht nachkommen kann, arbeitet vielleicht weniger sorgfältig oder muss Menschen länger auf ihr Recht warten lassen. Standardisierung kann auch zu mehr Gerechtigkeit führen, wenn dadurch sichergestellt ist, dass in ähnlichen Fällen ähnliche Urteile ergehen. Aber Gerechtigkeit bedeutet nicht nur, dass Gleiches gleich behandelt wird, sondern auch, dass unterschiedliche Fälle unterschiedlich beurteilt werden. Ein Richter muss auch dem Einzelfall gerecht werden. Eine zu starke Standardisierung der Abläufe mit dem Ziel der Effizienzsteigerung kann aber dazu führen, dass der Einzelfall nicht mehr als Einzelfall gesehen wird. Das ist beim Fluggastrecht vielleicht kein Problem, aber bei einem Sorgerechtsstreit wäre das völlig inakzeptabel.

Dieser Konflikt zwischen einer effizienten Verarbeitung von Standardfällen und einer sorgfältigen Prüfung von Einzelfällen findet sich in jeder Verwaltung, sei es bei einer Behörde oder in einem Konzern. Verwaltungen in großen Organisationen sind Entscheidungsfabriken. Für jede Verwaltung sind Einzelfälle lästig, weil sie nicht durch die gut geölte Bürokratiemaschine automatisch verarbeitet werden können. Sie verursachen extra Aufwand und manchmal auch Ärger. Da die Arbeitsbelastung ohnehin schon zu hoch ist und jedes Jahr noch mehr Fälle bearbeitet werden müssen, ist es nur allzu verständlich, dass Verwaltungen ihre Effizienz durch Standardisierung und Automatisierung erhöhen wollen. Wenn Ihr Fall in das Standardraster passt, freuen Sie sich, dass Ihr Fall schnell automatisch beschieden wurde. Aber wehe, wenn nicht... dann hängen Sie in der Warteschleife der Telefonhotline. Zwar könnten die durch Effizienzsteigerungen gesparten Res-

2 Kempfle (2023) stellt die Positionen des Richterbundes zu Massenverfahren dar. Der Einsatz von KI in der Justiz wird in der KI-Verordnung der EU als Hoch-Risiko-Anwendung eingestuft (Annex III, 8(a)), sodass man auch deshalb nicht erwarten muss, dass die deutsche Justiz übereilt ein KI-System einführt und mögliche Folgen dabei nicht bedenkt.

sourcen für die komplizierteren Einzelfälle genutzt werden, aber wäre es nicht noch effizienter, wenn diese Fälle, die im Moment noch nicht automatisch von Computern bearbeitet werden können, in Zukunft von KI-Systemen bearbeitet werden?

Wie es zum Kindergeldskandal kam

Chermaine Leysner hatte als Studentin vom niederländischen Staat eine finanzielle Unterstützung für die Betreuung ihrer drei Kinder bekommen. Dieses Geld, mehr als 100.000 Euro, forderte der Staat Jahre später von ihr zurück. Neun Jahre, viele Überstunden, eine Depression und eine Scheidung später stellt sich heraus, dass das »nur« ein bedauerlicher Fehler war. Zehntausenden Menschen in den Niederlanden erging es ähnlich. Ihnen wurden Leistungen für die Kinderbetreuung vorenthalten oder sie mussten wie Chermaine Leysner sogar Geld zurückzahlen. Die Rückforderungen waren teilweise so hoch, dass Familien deshalb verarmten. Häuser wurden verkauft und Insolvenzverfahren eingeleitet. Viele Eltern mussten ihre Kinder in Pflegefamilien geben. Andere gaben ihren Arbeitsplatz auf, um sich um die Kinder kümmern zu können. Es kam sogar zu einigen Selbstmorden. Erst nach vielen Beschwerden, Klagen, Presseberichten und Untersuchungen gestand die Regierung der Niederlande dieses Unrecht ein und trat 2021 deshalb zurück. Der Skandal nahm seinen Anfang damit, dass eine effiziente Verwaltung Betrugsfälle automatisch erkennen wollte.³

In den Jahren 2004 und 2005 wurden in den Niederlanden die rechtlichen Grundlagen für ein effizientes System zur Verteilung des Kinderbetreuungsgeldes geschaffen. Eltern zahlen zum Beispiel Geld für einen Kindergartenplatz an einen privaten Träger und können dafür eine Unterstützung vom Staat beantragen. Ein Ziel der Reform war, diese Unterstützung möglichst schnell auszubezahlen, damit Eltern das Geld flexibel zur Kinderbetreuung einsetzen können, wie sie es gerade brauchen. Daher bekamen Eltern einen Vorschuss und die

3 Der Toeslagen-Skandal (»Toeslagen« ist Niederländisch für Zulagen, in diesem Fall zur Kinderbetreuung) und das Beispiel von Chermaine Leysner sind bei Goujard & Manancourt (2022) beschrieben. Eine Beschreibung der Folgen findet sich auch bei Peeters & Widlak (2023). Meine folgende Zusammenfassung des Skandals basiert auf der Analyse von Peeters & Widlak (2023), die eine große Zahl an Medienberichten und offiziellen Untersuchungen ausgewertet haben.

Abrechnung passierte erst im Nachhinein über das Finanzamt. Alle Prozesse, von der Antragstellung über die Antragsbearbeitung bis zur Auszahlung, wurden hoch automatisiert. Daten des Finanzamtes, der Meldeämter und anderer Behörden wurden dafür mit Daten über Kindergartenplätze abgeglichen. Da das Finanzamt nun Zugriff auf mehr Daten hatte, die verarbeitet werden wollten, richtete es eine technische Abteilung zur Datenanalyse ein, die sich um den Austausch der Daten und deren Analyse kümmerte. Das heißt nicht, dass die Entscheidungen über die Anträge vollständig automatisch getroffen wurden, aber die Sachbearbeiter sollten bestmöglich unterstützt werden, um effizient am Fließband Entscheidungen zu treffen.

Bald wurde klar, dass dieses System anfällig für Betrug war. Rechnungen wurden gefälscht und Eigenanteile nicht gezahlt. Nachdem der Staat seinen Bürgerinnen und Bürgern durch die Vorschüsse viel Vertrauen entgegengebracht hatte, sollte nun ein hartes Durchgreifen für die nötige Abschreckung sorgen. Die Sachbearbeiter und -bearbeiterinnen im Finanzamt taten, was von ihnen erwartet wurde und genehmigten nur noch Anträge, die vollständig und korrekt waren. Damit kein möglicher Betrugsfall übersehen wurde, führte jede kleine Unregelmäßigkeit dazu, dass ein Antrag als verdächtig gekennzeichnet wurde. So sollten alle Betrugsfälle aufgedeckt werden.

Für diesen neuen Zweck nutzte man ganz selbstverständlich die zuvor etablierte technische Infrastruktur. Die technische Abteilung zur Datenanalyse machte sich an die Arbeit und teilte die Fälle automatisch in Risikogruppen ein. Auf Grundlage dieser Einschätzung des Betrugsrisikos setzten Sachbearbeiter ihre Ressourcen zur Kontrolle entsprechend ein. Ich weiß nicht genau, welche statistischen Methoden eingesetzt wurden und wie groß die Rolle von maschinellem Lernen war. Obwohl in Berichten über den Skandal immer von selbstlernenden Algorithmen die Rede ist, vermute ich, dass die Methoden so einfach waren, dass es irreführend wäre, hier von KI zu sprechen. Sicher ist hingegen, dass bei dem Versuch, Betrugsversuche automatisch zu erkennen, verschiedene Daten über die Antragsteller aus unterschiedlichen Quellen genutzt wurden. Man nutzte alte Anträge, die genehmigt oder als möglicher Betrugsfall gekennzeichnet wurden, um die Merkmale zu finden, die am besten vorhersagen, ob es sich um einen verdächtigen Antrag handelt oder nicht. Ganz genau so, wie neuronale Netze zur Objekterkennung die Merkmale lernen, die am besten vorhersagen, ob auf einem Bild eine Katze zu sehen ist oder nicht.

Während jeder Datenanalyst und jede -analystin beurteilen kann, ob auf einem Bild eine Katze zu sehen ist, wussten die Datenanalysten im Finanzamt nicht, wie genau die Einschätzung als Verdachtsfall zustande kam. Da sie selber keine Sachbearbeiter waren, ist anzunehmen, dass sie die Feinheiten der Abläufe in der Behörde nicht kannten und wahrscheinlich deshalb auch nicht vollständig verstanden, was die Daten, die sie analysierten, bedeuteten und wie ihre Analysen eingesetzt würden. Insbesondere kannten sie nicht den Grund für die Einschätzung eines Falles als möglicher Betrug, da in den Daten nicht zwischen unvollständigen Angaben, einfachen Fehlern und echtem Betrug unterschieden wurde. Man sollte meinen, diese Unterscheidung wäre wichtig und bei einem echten Betrugsverdacht würde die Behörde anders reagieren, als wenn die Angaben einfach nur unvollständig sind. Weil es aber in der Verwaltungssoftware keine Möglichkeit gab, die finanzielle Unterstützung zur Prüfung auszusetzen, wurde die Unterstützung einfach in allen Fällen ganz beendet. Zusammen mit der Vorgabe, dass Betrug unbedingt verhindert werden sollte, führte das dazu, dass schon im Verdachtsfall jede weitere Zahlung gestoppt wurde und Rückzahlungen eingefordert wurden. Ein späterer Bericht fand heraus, dass es in 94 Prozent der als möglicher Betrugsfall gekennzeichneten Fälle keinen konkreten Anhaltspunkt für Betrug gab. Mit katastrophalen Folgen für die fälschlich beschuldigten Menschen, die offenbar keiner in der Behörde bedacht hatte.

Aufgrund des hohen Automatisierungsgrades hatten die Sachbearbeiter nur wenig Kontakt mit den Bürgern, über deren Anträge sie entschieden. Für die Datenanalysten war jeder Fall ohnehin nur eine Reihe von Zahlen und Merkmalen in einer Tabelle. Die Beamten in der Beschwerdestelle hatten hingegen einen direkten Kontakt zu den Bürgern, die fälschlicherweise unter Betrugsverdacht standen. Diese Beamtinnen und Beamten sahen das Leid und die Verzweiflung, die die Behörde bei Menschen wie Chermaine Leysner verursachte. Sie äußerten ihre Kritik an den Entscheidungsprozessen laut und deutlich. Doch anstatt die Entscheidungsprozesse zu überdenken, reagierte die Leitung der Behörde damit, die Beschwerdeprozesse effizienter zu gestalten.

Im zuständigen Ministerium nahm man zur Kenntnis, dass es eine außergewöhnlich hohe Anzahl an Rückforderungen gab, war sich aber der dramatischen Folgen für die betroffenen Menschen nicht bewusst. Dass das ganze System einen erheblichen Verwaltungsaufwand bei

den Bürgern verursachte, die ihre Kinderbetreuungskosten auf Heller und Pfennig mit dem Finanzamt abrechnen mussten, wurde hingegen durchaus als Problem gesehen. Es wurde diskutiert, ob man statt das Geld an die Eltern auszuzahlen, nicht besser die Kindertagesstätten direkt finanzieren sollte. Diese Idee wurde aber verworfen, weil die Umstellung der IT zu schwierig erschien.

Diskriminierung ist jetzt automatisch

Selbst wenn das niederländische Finanzamt, welches das Kinderbetreuungsgeld verwaltet, weniger unverständliche Formulare haben sollte als deutsche Behörden, kann man sich leicht vorstellen, dass es trotzdem eine Herausforderung sein kann, alle Formulare richtig auszufüllen. Das betrifft insbesondere Menschen, die nicht ihre Eltern um Rat fragen können, die die Sprache nicht so gut beherrschen oder denen Zeit und Fähigkeiten fehlen, sich richtig zu informieren. Da die Datenanalyse zur Betrugserkennung nicht sauber zwischen Betrugsverdacht und fehlerhaften oder unvollständigen Anträgen unterschied, waren alle Menschen, die einfach nur Probleme damit hatten, die bürokratischen Hürden zu nehmen, automatisch verdächtig.

So ist es keine Überraschung, dass die automatische Betrugserkennung Antragsstellern mit geringen Einkommen oder einer doppelten Staatsbürgerschaft ein besonders hohes Betrugsrisiko zuschrieb. Daneben stellte sich im Zuge der Aufarbeitung des Skandals heraus, dass es schon lange vor der Reform des Kinderbetreuungsgeldes schwarze Listen in den Finanzämtern gab, die insbesondere türkische und marokkanische Staatsbürger unter Betrugsverdacht stellten. Diese schwarzen Listen flossen auch in die Bewertung des Betrugsrisikos für das Kinderbetreuungsgeld ein. Die niederländische Datenschutzbehörde, die die Einhaltung der Europäischen Datenschutzgrundverordnung überwacht, beurteilt es als diskriminierend und rechtswidrig, die (doppelte) Staatsbürgerschaft zu nutzen, um das Betrugsrisiko vorherzusagen. Die Finanzverwaltung musste deshalb eine Strafe von 2,75 Millionen Euro zahlen und die Regierung unter der Führung von Mark Rutte trat 2021 wegen des Skandals um das Kinderbetreuungsgeld zu-

rück. Aber nach den folgenden Neuwahlen und zähen Koalitionsverhandlungen war er mit der gleichen Koalition wieder im Amt.⁴

Das Versprechen, dass ein höherer Grad an Automatisierung neben einer Effizienzsteigerung auch zu mehr Objektivität und damit Gerechtigkeit bei der Bearbeitung von Anträgen führt, wurde beim niederländischen Kinderbetreuungsgeld nicht eingelöst. Im Gegenteil: Bestehende Diskriminierung wurde verstärkt!

Man muss keine Absicht unterstellen, um zu erklären, wie das passieren kann. Es gilt das DIDO-Prinzip (zur Erinnerung: DIDO steht für ›discrimination in, discrimination out‹). Nehmen wir folgendes Beispiel: Eine große Firma will ihre Personalauswahl effizienter gestalten und deshalb automatisieren. Dazu analysiert sie über einen längeren Zeitraum, welche Merkmale einer Bewerbung am besten vorhersagen, ob jemand eingestellt wird. Wenn Frauen im bestehenden Bewerbungsprozess gegenüber Männern benachteiligt werden, wird der Algorithmus lernen, dass das Geschlecht für die Einstellung relevant ist. Vielleicht soll auch der spätere Erfolg im Unternehmen vorhergesagt werden, und wenn es in der Firma bisher hauptsächlich Männer auf der Führungsetage gab, dann wird der Algorithmus lernen, dass das Geschlecht Erfolg vorhersagt. Ein Empfehlungsalgorithmus, der basierend auf diesen Daten Bewerbungen vorsortiert, wird Männer bevorzugen.

Gibt es schon ohne Algorithmen bei der Personalauswahl Diskriminierung, verschwindet diese nicht, nur weil ein Computer den Prozess unterstützt und dadurch scheinbar objektiver macht. Daten bilden bestehende Diskriminierung ab und Algorithmen, die auf solchen Daten beruhen, reproduzieren diese Diskriminierung. Man kann versuchen, mehr Geschlechtergerechtigkeit dadurch zu erreichen, dass ein Algorithmus das Merkmal Geschlecht nicht nutzen kann. So macht man das inzwischen, auch wenn kein Algorithmus beteiligt ist, etwa beim Vorspielen für Stellen in renommierten Orchestern, wo das Geschlecht keine Rolle spielen sollte. Wichtig ist nur, wie gut eine Person ihr Musikinstrument beherrscht. In vielen Auswahlverfahren für Orchester sind Kandidatinnen und Kandidaten deshalb beim Vorspielen nicht zu sehen, sondern nur zu hören.⁵

4 Siehe nochmal Goujard & Manancourt (2022).

5 Basierend auf der Studie von Goldin & Rouse (2000).

Die Lösung ist nicht immer so einfach. In einer Bewerbung für eine Informatikstelle mag kein Geschlecht stehen, aber Lücken im Lebenslauf könnten trotzdem auf Babypausen hindeuten. Männer, die Informatik studieren, wählen vielleicht eher Maschinenbau als Nebenfach und Frauen eher Psychologie. In einem Lebenslauf finden sich viele solche Hinweise auf das Geschlecht und nicht alle sind so offensichtlich wie das Abiturzeugnis einer katholischen Mädchenschule. Ein Algorithmus, der das Geschlecht eines Bewerbers nicht kennt, kann all diese Hinweise trotzdem indirekt nutzen.⁶ Außerdem: Wenn einer Frau, weil sie eine Frau ist, in der Vergangenheit weniger Chancen gegeben wurden als einem Mann, ist sie in einem Bewerbungsverfahren selbst dann noch benachteiligt, wenn oberflächlich betrachtet, ihr Geschlecht bei der Auswahl gar keine Rolle spielt. Leider gibt es für diese Probleme keine einfachen, rein technischen Lösungen.⁷

Digital first, Bedenken second

Die FDP hat im Bundestagswahlkampf 2017 neben einem Schwarz-Weiß-Bild von Christian Lindner, der gebannt auf sein Handy schaut, ihren Wahlslogan in leuchtenden Farben und Großbuchstaben auf Plakate gedruckt: »DIGITAL FIRST. BEDENKEN SECOND.« Bei dem schleppenden Tempo der Digitalisierung in Deutschland ist nur allzu verständlich, dass manch einer ungeduldig wird. In der Softwareentwicklung, insbesondere bei Start-ups, gilt oft ein ähnliches Motto: »Move fast and break things.« Einfach mal machen, statt lange zu überlegen. Man kann sowieso nicht alles im Voraus planen. Dementsprechend löst man die Probleme am besten erst, sobald sie wirklich auftreten. Auf dem Softwaremarkt ist das eine ausgezeichnete Strategie, um schnell ein Produkt zu erstellen, das man in der Praxis erprobt und nach und nach anpasst. In der Forschung oder bei der Entwicklung von Prototypen führt dieser Ansatz dazu, dass man schnell lernt, was funktioniert und was nicht.

Bei der Entwicklung von Software ist es zu einem gewissen Grad akzeptabel geworden, dass das Produkt ausgeliefert wird und Prob-

⁶ Siehe nochmal Dastin (2018).

⁷ Buyl & De Bie (2024) geben einen Überblick darüber, warum es wesentlich schwieriger ist, als man denken könnte, Fairness technisch zu erreichen.

leme erst im laufenden Betrieb beim Kunden durch Updates behoben werden. Dass moderne Smart-TVs und Smartphones häufig Fehlermeldungen produzieren, regt nur noch Leute auf, die sich daran erinnern können, wie verlässlich analoge Fernseher und Telefone zuletzt waren. Alle anderen scheinen sich daran gewöhnt zu haben, dass Software im Alltag nicht immer so problemlos zu benutzen ist, wie man sich das wünschen könnte. Software hat oft Fehler, aber Probleme entstehen auch dadurch, dass Software umständlich zu bedienen ist oder nicht so funktioniert, wie die Nutzer es erwarten. Für das autonome Fahren wünscht man sich allerdings, dass vorhersehbare Probleme schon bei der Entwicklung behoben werden und Unvorhersehbares im Probebetrieb auffällt, bevor ein KI-System an den Kunden geliefert wird und dort womöglich Unfälle verursacht.⁸

Bei traditioneller Software gibt es lange etablierte Entwicklungs- und Prüfverfahren, um Fehler möglichst zu vermeiden. Bei sicherheitskritischer Software (anders als bei Smart-TVs) werden diese auch konsequent eingesetzt. Besonders wichtig ist dabei Eingaben und erwartetes Verhalten durch Standardisierung genau zu spezifizieren, damit man im Betrieb keine Überraschungen erlebt. KI soll aber immer genau in den Fällen helfen, die schwer zu spezifizieren sind und bisher noch menschliches Urteilsvermögen erfordern. Deshalb ist nicht leicht zu testen, ob die KI-Software auch wirklich das macht, was sie soll. Für KI-Systeme gibt es deshalb leider noch keine etablierten Standards.⁹ Während es in manchen Anwendungsfällen nur nervt, wenn eine Software sich nicht wie erwartet verhält, kann es in anderen Fällen ernste Folgen haben. Sind die Risiken größer, sollten auch die Bedenken größer sein. Mehr Bedenken hätten wahrscheinlich auch im niederländischen Finanzministerium Schlimmeres verhindert.

Als die ersten Probleme sichtbar wurden, hatten die Verantwortlichen in den Niederlanden zu Recht die Sorge, dass eine Umstellung der Verwaltung des Kinderbetreuungsgeldes große Änderungen in der IT erfordern würde. Große IT-Projekte gehen oft schief. Vom Drama um die Technik zur Mauterhebung auf der Autobahn bis zur elektro-

8 Das ist aber nicht immer der Fall. Tesla hat z.B. das Problem, dass sich Fahrer zu sehr auf den »Autopiloten« verlassen. Dieses Problem ist aber nicht leicht durch ein reines Softwareupdate zu fixen, ohne auch die Auswirkungen der Änderungen auf das menschliche Verhalten genau zu untersuchen (Duncan & Thadani, 2024; Krisher & The Associated Press, 2024).

9 Aber siehe DIN & DKE (2022).

nischen Patientenakte – die Medien berichten gerne und ausführlich darüber, wenn große IT-Projekte des Staates scheitern. Aber jeder, der einmal die Einführung einer neuen Software in einem Unternehmen mitbekommen hat, kann berichten, dass es auch in der Privatwirtschaft selten wirklich glattläuft. Laut einer einschlägigen Statistik sind IT-Projekte im Durchschnitt fast 75 Prozent teurer als geplant. Betrachtet man das schlechteste Fünftel aller IT-Projekte, waren diese im Durchschnitt sogar mehr als fünfmal so teuer. Offensichtlich wird bei vielen IT-Projekten nicht gut geplant. Fehlende Sorgfalt bei der Planung, übermäßig optimistische Schätzungen und unklare und widersprüchliche Ziele gehören zu den Problemen, die nicht nur IT-Projekte plagen, aber große IT-Projekte scheinen im Vergleich zu anderen Großprojekten besonders schlecht zu laufen. Die Erfahrung zeigt, dass am Ende viele Projekte ungleich komplizierter waren als ursprünglich gedacht.¹⁰ Da auch die Entwicklung und Einführung von KI-Software ein IT-Projekt ist, gibt es keinen Grund anzunehmen, dass es bei KI-Projekten besser laufen wird.

Die Einführung von neuer Software ist keine rein technische Aufgabe. Fast immer gehen damit Veränderungen in den Abläufen einer Organisation einher. Das gilt für Behörden genauso wie für Unternehmen. Probleme treten zum Beispiel auf, wenn Abläufe im Entwicklungsprozess missverstanden werden und nachträglich in der Software angepasst werden müssen. Andere Abläufe wurden vielleicht vergessen und besonders komplexe Abläufe aus Kostengründen zunächst nicht abgebildet. Einzelfälle, die nicht ins Raster passen, machen nach der Einführung der neuen Software noch mehr Probleme als vorher, weil bei der Entwicklung niemand an sie gedacht hat. Alte Daten müssen verfügbar bleiben und die Umstellung findet im laufenden Betrieb statt. Vielleicht haben sich auch rechtliche Rahmenbedingungen seit der Erteilung des Entwicklungsauftrages geändert. Es gibt viele Gründe, warum ein IT-Projekt scheitern kann. Viele Komplikationen entstehen aber dadurch, dass sich die Software in die Arbeitsprozesse der Menschen einpassen muss und die Menschen können oder wollen ihre Prozesse nicht immer an die Software anpassen.

Es gibt ohne Zweifel eine Vielzahl an langweiligen Aufgaben in Verwaltungen, die Sachbearbeiter und -bearbeiterinnen gerne automa-

10 Siehe Flyvbjerg & Gardner (2023). Die genannte Statistik findet sich auf S. 193 und die Inflation ist hierbei nicht mal eingerechnet.

tisieren würden. Deutschland ist nach wie vor ein Entwicklungsland, was die Digitalisierung der Behörden angeht. Ich kann aus erster Hand bestätigen, dass es an den Universitäten nur schleppend mit der Digitalisierung vorangeht und noch viele Papierakten gepflegt werden. Das liegt weniger daran, dass es zu viele Bedenken-träger gibt. Vielmehr fehlen teilweise noch wichtige rechtliche Grundlagen und da der Aufwand riesig ist, fehlen auch Zeit und Geld – oftmals auch Expertise. Bevor wir über den großflächigen Einsatz von KI in Behörden nachdenken, müssen wir in Deutschland erst bei der Digitalisierung der Verwaltung aufholen. Der aktuelle KI-Hype führt vielleicht dazu, dass es schneller vorangeht, aber wir sollten dabei darauf achten, dass nicht solche Katastrophen wie beim Kindergeldskandal in den Niederlanden passieren.

Auch in Deutschland ist die Versuchung groß, menschliches Urteilsvermögen im Namen der Effizienz durch Computer zu ersetzen. Es ist eine Sache, Verwaltungen durch den Einsatz von IT effizienter gestalten zu wollen, es ist aber eine gänzlich andere Sache, wenn Entscheidungen nicht mehr von Menschen, sondern von Computern getroffen werden. Was beim Fluggastrecht unproblematisch erscheinen mag, ist bei der Einstufung eines Antrags als Betrugsfall inakzeptabel. Moderne KI-Methoden können zwar noch nicht alle Einzelfälle, die in Verwaltungen anfallen, verlässlich automatisch bearbeiten, aber selbst wenn sie es könnten, besteht die große Gefahr, dass durch den Einsatz von KI bestehende Diskriminierung verstärkt wird. Wichtige Entscheidungen, die vielleicht große Auswirkungen auf die betroffenen Menschen haben können, dürfen nicht vollautomatisch von Maschinen getroffen werden. Sollen Computer aber menschliche Entscheidungen lediglich unterstützen, so wie die KI-Verordnung der EU es bei folgenreichen Entscheidungen vorsieht, dann stellt sich die Frage, wie genau Mensch und Maschine bei der Entscheidungsfindung zusammenarbeiten sollen. Auf diese Frage gibt noch keine befriedigende Antwort und sie wird wahrscheinlich für jede zukünftige KI-Anwendung in der Verwaltung separat beantwortet werden müssen.

