

## **Die Risiken des Einsatzes von künstlicher Intelligenz im Prozess**

**Professorin an der Universität Paris Nanterre**

Ich werde zunächst einen Überblick über die Vielfalt der Realitäten geben, auf die sich die digitale Justiz bezieht (1), dann die Illusionen und Risiken des Einsatzes von KI in der Justiz aufzeigen (2) und schließlich Wege für eine normative Rahmung skizzieren (3).

## **I rischi dell'uso dell'intelligenza artificiale nei procedimenti amministrativi.**

Professore presso l'Università di Parigi Nanterre

Darò innanzitutto una panoramica della diversità delle realtà a cui si riferisce la giustizia digitale (1), prima di mostrare le illusioni e i rischi dell'uso dell'IA in campo giudiziario (2) e infine di delineare le vie per un quadro normativo (3).

1) Die digitale Justiz bezieht sich auf sehr unterschiedliche Realitäten.

Realitäten, die unterschiedlichen Funktionen entsprechen :

a) Es kann sich um leistungsfähige Suchmaschinen handeln, die bei der Argumentation (im Dienste der Anwälte/Juristen) und bei der Entscheidung (für den Richter) helfen - Suche nach Rechtsprechung, Schlussfolgerungen öffentlicher Berichterstatter... Kurz gesagt, für die juristische Dokumentationsuche. Der Aufschwung der KI im Bereich der Verarbeitung natürlicher Sprache hat die Entwicklung der semantischen Suche ermöglicht. Dies ermöglicht es, ausgehend von einer in Alltagssprache formulierten Frage auf eine Vielzahl von Quellen zuzugreifen. Viele Richter nutzen google zusätzlich zu ihren Datenbanken, aber auch Lösungen von Legaltech-Unternehmen (z. B. Doctrine). Dies kann eine wertvolle Hilfe für Richter und Anwälte sein.

1) La giustizia digitale si riferisce a realtà molto diverse

Realtà che corrispondono a funzioni diverse:

a) Possono essere potenti motori di ricerca per aiutare l'argomentazione (per gli avvocati) e il processo decisionale (per il giudice) - ricerca di giurisprudenza, conclusioni di relatori pubblici... In breve, per la ricerca di documenti giuridici. Lo sviluppo dell'intelligenza artificiale nel campo dell'elaborazione del linguaggio naturale ha reso possibile lo sviluppo della ricerca semantica. Ciò consente di accedere a una moltitudine di fonti a partire da una domanda formulata nel linguaggio quotidiano. Molti giudici utilizzano Google in aggiunta alle loro banche dati, ma anche alle soluzioni offerte da legaltech (come Doctrine, ad esempio). Questo può essere un grande aiuto per giudici e avvocati.

b) Es kann sich um eine Hilfe bei der "administrativen Bearbeitung" von Streitfällen handeln. Von der Registrierung der Beschwerden bis zur Zustellung und Veröffentlichung der Entscheidung gibt es viele Schritte, zu denen KI beitragen kann, um Richter und Kanzleien zu entlasten. So kann KI z. B. bei der Erkennung von Prozessserien (Projekt des EK), der automatischen Erstellung von Gerichtsprotokollen durch Spracherkennung und Texttranskription (Kroatien, Estland, Deutschland, Ungarn), der Übersetzung von Gerichtsurteilen (Schweden) oder der Pseudonymisierung von Gerichtsurteilen (Österreich, Finnland, Schweden, Estland, Spanien, lux, Frankreich) helfen.

b) Può essere un aiuto per il "trattamento amministrativo" dei fascicoli di contenzioso. Dalla registrazione dei ricorsi alla notifica e pubblicazione della decisione, sono molte le fasi a cui l'IA può contribuire per alleggerire magistrati e cancellerie. Ad esempio, l'IA può contribuire all'individuazione delle serie di controversie (progetto portato avanti dalla CE), alla creazione automatizzata dei verbali delle udienze mediante riconoscimento vocale IA e trascrizione del testo (Croazia, Estonia, Germania, Ungheria), alla traduzione delle decisioni giudiziarie (Svezia) o alla pseudonimizzazione delle decisioni giudiziarie (Austria, Finlandia, Svezia, Estonia, Spagna, Lussemburgo, Francia).

So haben der Staatsrat und der Kassationshof in Frankreich automatisierte Instrumente zur Pseudonymisierung - d. h. zur Entfernung von Elementen, die eine direkte Identifizierung der Parteien und der darin erwähnten Dritten ermöglichen - eingerichtet. Und zwar dank der Bildung qualifizierter interner Teams (kein Outsourcing), die aus Datenwissenschaftlern und Entwicklern bestehen und von einer Richterin und Annotatoren (Beamte der Kat. C des Gerichtshofs) gesteuert werden, die die zu verbergenden Einträge kennzeichnen, den Trainingsdatensatz zusammenstellen und die Qualitätskontrolle gewährleisten.

Il Consiglio di Stato e la Corte di Cassazione in Francia hanno quindi creato strumenti automatici di pseudonimizzazione, ovvero l'eliminazione degli elementi che consentono l'identificazione diretta delle parti e dei terzi citati nella decisione. Ciò è reso possibile dalla creazione di team interni qualificati (non esternalizzati) composti da data scientist e sviluppatori, guidati da un magistrato e da annotatori (agenti di categoria C del tribunale) che etichettano le menzioni da sopprimere, creano il set di dati di formazione e assicurano il controllo di qualità.

c) Es kann sich um KI zur Entscheidungsunterstützung handeln, mit dem Ziel, durch eine Funktion zur Empfehlung von Lösungen (Predictice für Anwälte) bei der Identifizierung, Analyse und Lösung von Fragen, die in einem Rechtsstreit gestellt werden, zu helfen (Prometea, Datajust, ChatGPT, COMPAS für Richter). Daran denken wir vor allem, wenn wir von KI in der Justiz sprechen.

Die Idee ist, Big Data und Open Data zu nutzen, um den Ausgang eines Rechtsstreits vorherzusagen. Die KI, die von Big Data gespeist wird, wird in der Lage sein, Tausende von Daten (frühere Fälle) zu verarbeiten, um ein Vorhersagemodell vorzuschlagen - eine Lösung, die mit dem in Statistiken oder "in Daten gegossenen" Rechtszustand übereinstimmt -, was ein Mensch nicht erreichen könnte. Man spricht heute eher von Jurimetrie - und nicht von prädiktiver Justiz -, um diese Möglichkeit zu charakterisieren, automatisch statistische Wiederkehrmuster in den Begründungen von Urteilen zu erkennen, um sie zu sortieren und Lösungen für den jeweiligen Fall vorzuschlagen. Auf diesem Gebiet wird am meisten experimentiert, nicht zuletzt, weil sich nicht nur Legaltechs dafür interessieren, sondern auch die Behörden.

c) può essere un supporto decisionale dell'IA, con l'obiettivo di aiutare a identificare, analizzare e risolvere le questioni sollevate in una controversia, attraverso una funzione di raccomandazione di soluzioni (Predictice per gli avvocati) (Prométéa, Datajust, ChatGPT, COMPAS per i giudici). Questo è ciò che viene in mente quando si parla di IA nel sistema giudiziario.

L'idea è quella di utilizzare i big data e gli open data per prevedere l'esito di una controversia. L'IA, alimentata dai big data, sarà in grado di elaborare migliaia di serie di dati (casi precedenti) per proporre un modello predittivo - una soluzione che sarà in linea con lo stato della legge messa in statistica, o "messa in dati" - che un essere umano non potrebbe raggiungere.

Oggi si parla più di giurimetria - piuttosto che di giustizia predittiva - per caratterizzare questa possibilità di individuare automaticamente le ricorrenze statistiche nelle motivazioni delle sentenze, per ordinarle e proporre soluzioni in base al caso.

È in questo ambito che le sperimentazioni sono più numerose, soprattutto perché è questo l'interesse di legaltech, ma anche delle autorità pubbliche.

Legaltech-Unternehmen haben ihren Markt bei Anwälten und Juristen - und nur am Rande bei Bürgern - entwickelt, denen sie verschiedene Lösungen anbieten. Predictice bietet z. B. ein "Tool, das die Wahrscheinlichkeit der Beilegung eines Rechtsstreits berechnet und die Höhe der zu erwartenden Entschädigung schätzt. Diese Daten können in Form eines anpassbaren Berichts exportiert werden.

Dieser Markt konnte in Frankreich dank der Öffnung von Gerichtsdaten, die im NRL von 2016 verankert wurde, florieren. Heute werden die meisten Entscheidungen der JA in OD gestellt. Bei der Justiz ist es langsamer, da es viel mehr Entscheidungen gibt.

Anfangs gab es kein Verbot, die Daten für die Erstellung von Profilen von Richtern und Staatsanwälten weiterzuverwenden, wie es in den USA üblich ist. Also gibt es Start-ups, die Statistiken über die Tätigkeit von Richtern erstellt haben, z. B. Supralegem, das einen Algorithmus entwickelt hat, um die Anzahl der Ablehnungen von Einsprüchen gegen OQTF (obligation de quitter le territoire français) pro Richter zu vergleichen.

Le Legaltech hanno sviluppato il loro mercato tra gli avvocati e i giuristi - molto marginalmente tra i contendenti - ai quali offrono diverse soluzioni. Ad esempio, Predictice offre uno "strumento (che) calcola la probabilità di risoluzione di una controversia e stima l'importo del risarcimento che si può ottenere. Questi dati possono essere esportati sotto forma di report personalizzabili".

Questo mercato è potuto fiorire in Francia grazie all'apertura dei dati giudiziari varata dall'ANR del 2016. Oggi, la maggior parte delle decisioni della JA viene messa in OD. Per la giustizia giudiziaria, il processo è più lento perché ci sono molte più decisioni.

All'inizio, nulla impediva di riutilizzare i dati per tracciare il profilo dei magistrati, come avviene negli Stati Uniti. Ci sono quindi start-up che hanno prodotto statistiche sull'attività dei giudici, ad esempio Supralegem, che ha messo a punto un algoritmo per confrontare il numero di rigetti di ricorsi OQTF (obbligo di lasciare il territorio francese) per giudice.

Es wurde jedoch ein Gesetz verabschiedet, das dies in Bezug auf Richter und Staatsanwälte sowie Bedienstete der Kanzlei verbietet. Anders als in den USA ist es also in Frankreich (und in der EU) verboten, ein Profiling von Richtern und Staatsanwälten durchzuführen. Mit dem Gesetz vom 23. März 2019 wurde jede Weiterverwendung der Identitätsdaten von Richtern und Staatsanwälten und Kanzleiangehörigen verboten, die zum Ziel oder zur Folge hat, ihre Berufspraktiken zu bewerten, zu analysieren, zu vergleichen oder vorherzusagen (Art. L.10 CJA und L. 111.13 COJ). Es wurden auch RIS entwickelt, um Richtern Entscheidungshilfe zu bieten. Prometea, Datajust, ChatGPT, COMPAS Eine der ältesten Verwendungen (zumindest die, über die am meisten berichtet wurde) ist der COMPAS-Algorithmus, der von der US-Strafjustiz verwendet wird.

Ma è stata approvata una legge che lo vieta in relazione a magistrati e cancellieri. Quindi, a differenza degli Stati Uniti, la profilazione dei magistrati è vietata in Francia (e nell'UE). La legge del 23 marzo 2019 ha vietato qualsiasi riutilizzo dei dati d'identità dei magistrati e dei membri del registro con lo scopo o l'effetto di valutare, analizzare, confrontare o prevedere le loro pratiche professionali (art. L.10 CJA e L. 111.13 COJ). Sono stati sviluppati anche AIS per fornire ai giudici un supporto alle decisioni. Prométéa, Datajust, ChatGPT, COMPAS. Uno degli utilizzi più vecchi (o almeno quello che ha ricevuto maggiore attenzione da parte dei media) è l'algoritmo COMPAS, utilizzato dal sistema di giustizia penale statunitense.

Dabei handelt es sich um ein AIS für die persönliche prädiktive Analyse, das das Rückfallrisiko eines Angeklagten, das Risiko gewalttätigen Verhaltens und das Risiko, nicht vor Gericht zu erscheinen, wenn keine Untersuchungshaft angeordnet wird, bewertet. Die Software, die die Praxis des evidence based-sentencing veranschaulicht, besteht also darin, dass über die Strafe und ihre Modalitäten auf der Grundlage des Scores entschieden wird, der dem Angeklagten aufgrund seines Rückfallrisikos zugewiesen wird, das anhand von 137 Eingabedaten berechnet wird. "Die Strafe wird nicht mehr nach der Schwere der Straftat entschieden, sondern nur noch an die statistische Vorhersehbarkeit der Wiederholung einer Straftat angepasst" (A. Garapon).

Si tratta di un AIS di analisi predittiva personale che valuta il rischio di recidiva, il rischio di comportamento violento e il rischio di mancata comparizione in tribunale di un imputato se non rinviato a giudizio. Il software, che illustra la pratica dell'evidence-based sentencing, consiste quindi nel decidere le pene e le modalità di condanna sulla base del punteggio assegnato all'imputato in base al suo rischio di recidiva, calcolato sulla base di 137 dati di input. "La pena non viene più decisa in base alla gravità del reato, ma solo adattata alla prevedibilità statistica della reiterazione di un reato" (A. Garapon).



In Südamerika wurde ein weiteres Tool, Prometea, in der Staatsanwaltschaft von Buenos Aires entwickelt (Lösungsempfehlungen für Streitfälle, bei denen es um die Zuweisung von Wohnraum oder Sozialleistungen geht). Die Maschine schlägt innerhalb von 20 Sekunden und in 96 % der Fälle die richtige Lösung vor und erstellt automatisch den entsprechenden Entscheidungsentwurf. Dieses AIS wurde auch in Kolumbien (PretolA) am Verfassungsgericht eingesetzt, um Tutela-Entscheidungen zu analysieren (mit denen jeder Einzelne den Schutz seiner verfassungsmäßigen Grundrechte einfordern kann, die durch das Handeln oder Nichthandeln einer Behörde bedroht sind - Entscheidungen werden automatisch an das Verfassungsgericht weitergeleitet). Das AIS analysiert in 5 Sekunden die 2700 Entscheidungen, die das Gericht jeden Tag erhält, mit einer Erfolgsquote von 95%.

In Sudamerica, un altro strumento, Prométéa, è stato sviluppato presso la Procura di Buenos Aires (per consigliare soluzioni a controversie riguardanti l'assegnazione di alloggi o l'assistenza sociale). La macchina propone la soluzione corretta in 20 secondi e nel 96% dei casi e produce automaticamente la bozza di decisione corrispondente. Questo AIS è stato utilizzato anche in Colombia (PretolA) presso la Corte costituzionale per analizzare le decisioni di tutela (che consentono a qualsiasi contendente di rivendicare la tutela dei propri diritti costituzionali fondamentali minacciati dall'azione o dall'inazione di un'autorità pubblica - decisioni che vengono automaticamente trasmesse alla Corte costituzionale). L'ASI analizza in 5 secondi le 2700 decisioni che la Corte riceve ogni giorno, con un tasso di successo del 95%.

In Frankreich: Datajust: KI, die 2020 für eine zweijährige Testphase geschaffen wurde. Erste KI, die vom Justizministerium in Frankreich eingesetzt wurde, um durch die massive Verarbeitung einer beträchtlichen Menge an Gerichtsurteilen einen Referenzrahmen für die Entschädigung von Körperverletzungen zu schaffen und nebenbei die Gerichte zu entlasten. Das Justizministerium hat sich des Themas angenommen, um nicht ein Feld zu besetzen, das von den Legaltechs bereits gut besetzt ist. Das Ministerium stützte sich auf "GI-Unternehmer" ... Das Projekt wird schließlich im Januar 2022 eingestellt, weil die Vorschläge der Maschine unzuverlässig waren. Die Datenbank, auf der der Algorithmus trainiert wurde, war verzerrt, da sie unvollständig war, da die Daten der Entscheidungen der ersten Instanz nicht darin enthalten waren.

In Francia, Datajust: AI creata nel 2020 per un esperimento di 2 anni. 1° AI utilizzata dal Ministero della Giustizia francese per creare un sistema di riferimento per il risarcimento dei danni alla persona e, in seconda battuta, per alleggerire i tribunali, basandosi sull'elaborazione massiva di una notevole quantità di sentenze. Il Ministero della Giustizia ha ripreso l'argomento per non occupare un campo già ben occupato da legaltech. Il Ministero si è affidato a "GI Contractors"... Il progetto è stato infine abbandonato nel gennaio 2022 perché le proposte della macchina non erano affidabili. Il database su cui è stato addestrato l'algoritmo era falsato perché incompleto, in quanto non includeva i dati delle decisioni di primo grado.

Der kolumbianische Richter Juan Manuel Padilla Garcia erklärt, er habe ChatGPT verwendet, um über die Frage zu entscheiden, ob eine Versicherung alle Kosten für die medizinische Behandlung eines autistischen Kindes übernehmen müsse oder nicht. Der IA entschied zugunsten der Mutter und gegen den Versicherer, da er der Ansicht war, dass nach kolumbianischem Recht alle Kosten übernommen werden müssten. Wichtig ist, dass der Richter betonte, dass der Algorithmus die Entscheidung nicht für sie getroffen habe, sondern nur eine "Entscheidungshilfe" gewesen sei.

Denn wirklich beunruhigend wäre es, wenn der Richter der Entscheidung blind folgen würde, und zwar so sehr, dass die Entscheidung in Wirklichkeit von der Maschine getroffen würde. In diesem Sinne könnte man von einem "Richter-Roboter" sprechen.

Come non citare l'uso dell'IA ChatGPT da parte di un giudice colombiano per decidere un caso a fine gennaio 2023? Il giudice colombiano Juan Manuel Padilla Garcia spiega di aver utilizzato ChatGPT per decidere se una compagnia assicurativa dovesse o meno coprire tutti i costi delle cure mediche per un bambino autistico. L'AI si è pronunciato a favore della madre e contro l'assicuratore, ritenendo che secondo la legge colombiana tutti i costi debbano essere coperti. È importante notare che il magistrato ha sottolineato che l'algoritmo non ha preso la decisione al posto della madre, ma è stato solo un "aiuto al processo decisionale".

Ciò che sarebbe davvero preoccupante è se il giudice seguisse ciecamente la decisione, al punto che la decisione sarebbe effettivamente presa dalla macchina. A questo punto si potrebbe parlare di un "giudice robot".

d) Man spricht dann von "Roboterjustiz", wenn die Entscheidung vollständig automatisiert ist. Vorteile: territoriale Einheitlichkeit der JP (Gleichheit der Bürger vor der Justiz), größere Rechtssicherheit (Verringerung oder sogar Abschaffung des juristischen Zufalls durch Neutralisierung der subjektiven Parameter, die die Entscheidung beeinflussen können), Schnelligkeit, Senkung der Justizkosten...

Bisher sind die konkreten Umsetzungen sehr begrenzt (Estland, China, Vereinigtes Königreich - Online-Kurse). Estland: Ankündigung der vollständigen Automatisierung von Urteilen in kleinen zivilrechtlichen Streitigkeiten (unter 7000 EUR) im Jahr 2019, aber offenbar noch keine konkreten Schritte.

In China experimentiert die Volksstaatsanwaltschaft Shanghai Pudong mit einem AIS, das Strafanträge auf der Grundlage einer Beschreibung der Daten des Rechtsstreits in 7 Kategorien gängiger Straftaten (Diebstahl, Betrug, Glücksspiel, rücksichtsloses Fahren ...) formuliert: Die Genauigkeit ist hoch: 97 %. Schade um die restlichen 3%!

Diese Perspektive ist meiner Meinung nach beängstigend und zum Glück entwickelt sie sich nicht allzu sehr. Die Mobilisierung von KI im Justizbereich kann in der Tat, je nach der Funktion, die man ihr zuweisen will, sehr riskant sein.

d) Si parla poi di "giustizia robotizzata" quando la decisione è interamente automatizzata. Vantaggi: uniformità territoriale della JP (uguaglianza dei cittadini davanti alla legge), maggiore certezza del diritto (riduzione o addirittura eliminazione dell'incertezza giudiziaria grazie alla neutralizzazione dei parametri soggettivi che possono influenzare la decisione), rapidità, riduzione dei costi della giustizia, ecc.

Per il momento si sta facendo molto poco (Estonia, Cina, Regno Unito - tribunali online). Estonia: annuncio della completa automazione delle sentenze nelle piccole cause civili (meno di 7.000 euro) nel 2019, ma apparentemente nulla di operativo.

In Cina, la Procura del Popolo di Shanghai Pudong sta sperimentando un AIS che formula ordini di condanna basati sulla descrizione dei dati del caso, in 7 categorie di reati comuni (furto, frode, gioco d'azzardo, guida pericolosa...): il tasso di accuratezza è elevato: 97%. Peccato per il restante 3%!

Si tratta, a mio avviso, di una prospettiva spaventosa e fortunatamente non si sta sviluppando troppo. La mobilitazione dell'IA in campo giudiziario può infatti essere molto rischiosa, a seconda della funzione che le si vuole assegnare.

## 2) KI in der Justiz: von Illusionen zu realen Risiken

a) Missverständnisse darüber, was KI leisten kann (von der Schwierigkeit, die Gerichtsentscheidung zu kodieren) (auch hier gilt: Überspringen Sie diese Folie, wenn die Vorredner diese Gedanken in Betracht gezogen haben).

- Die als "Predictive Justice" bezeichneten AIS haben in Wirklichkeit nicht die Fähigkeit, die Zukunft zu lesen. Sie beschränken sich darauf, Lösungen aus der Vergangenheit zu projizieren.

- Ein AIS "sagt nicht, was die richtige Lösung für ein Rechtsproblem ist, sondern liefert dem Nutzer die Wahrscheinlichkeit, dass Tausende von Richtern der Vergangenheit (...) die Antwort A oder die Antwort B auf eine bestimmte Frage geben". Data-Mining-Techniken können übliche Praktiken aus quantitativer Sicht wiedergeben. Dies ist ein statistischer Ansatz - man spricht von Jurimetrie. Es kann für Richter wertvoll sein, zu wissen, was eine "durchschnittliche Entscheidung" ist, um vielleicht ihr Verhalten anzupassen. Andererseits kann es aber auch zu erheblichen kognitiven Verzerrungen führen, da das "algorithmische Urteil" den Richter unter Druck setzt.

## 2) IA giudiziaria: dalle illusioni ai rischi reali

a) Idee sbagliate su ciò che l'IA può fare (a partire dalla difficoltà di codificare la decisione giudiziaria) (anche in questo caso, saltare questa diapositiva se gli oratori precedenti hanno preso in considerazione queste idee)

- La cosiddetta "giustizia predittiva" L'AIS non ha in realtà la capacità di leggere il futuro. Si limitano a proiettare soluzioni dal passato.

- Un AIS "non dice all'utente quale sia la soluzione corretta a un problema legale, ma fornisce all'utente la probabilità che migliaia di giudici del passato (...) forniscano la risposta A o la risposta B a una determinata domanda". Le tecniche di data mining possono rendere conto delle pratiche comuni da un punto di vista quantitativo. Si tratta di un approccio statistico, chiamato giurimetria. Può essere utile per i magistrati sapere qual è la "decisione media", per poter eventualmente adeguare il proprio comportamento. D'altro canto, però, può generare significativi pregiudizi cognitivi perché il "verdetto algoritmico" mette il giudice sotto pressione.

- Technische Grenzen: AIS sind noch nicht sehr leistungsfähig: z. B. das Experiment, das 2016 mit Urteilen des EGMR durchgeführt wurde. Die automatische Analyse berichtet von einem Vorhersagegrad von 80 %. Das bedeutet, dass ein Tool, das letztlich nur ein binäres Ergebnis liefert - Verletzung oder nicht Verletzung eines Artikels der Konvention - in einem von fünf Fällen falsch liegt... Beispiel auch Datajust, das aufgegeben wurde, weil das Dataset unzuverlässig war. Generell führen bestimmte Rechtsstreitigkeiten nur zu einer geringen Anzahl von Entscheidungen, daher gibt es nicht genügend Daten, um eine zuverlässige Lernbasis zu bilden. Schließlich ist die KI noch nicht leistungsfähig genug, um alle redaktionellen Feinheiten einer Entscheidung zu erfassen.

- Kurz gesagt: Im Moment wird das Potenzial der KI vielleicht etwas überschätzt. Da es sich jedoch um einen Bereich handelt, der schnell voranschreitet, ist es besser, über die Risiken - insbesondere über die Verzerrungen - nachzudenken, die oft schon gut aufgezeigt wurden, die sich abzeichnen und gegen die man sich schützen muss, damit die digitale Justiz unter Achtung der Grundrechte und der Gleichheit der Bürger gedeihen kann.

- Limiti tecnici : gli AIS non sono ancora molto efficienti: ad esempio, l'esperimento condotto sulle sentenze della CEDU nel 2016. L'analisi automatica mostra un grado di predizione dell'80%. Questo significa che uno strumento che, alla fine, si accontenta di fornire un risultato binario - violazione o meno di un articolo della convenzione - sbaglia 1 volta su 5... Un altro esempio è Datajust, che è stato abbandonato perché il dataset non era affidabile. In generale, alcuni casi danno luogo solo a un piccolo numero di decisioni, quindi non ci sono abbastanza dati per costituire una base di apprendimento affidabile. Infine, l'IA non è ancora abbastanza potente da cogliere tutte le sottigliezze editoriali di una decisione.

- In breve, per il momento il potenziale dell'IA è forse un po' sopravvalutato. Ma poiché si tratta di un campo che sta progredendo rapidamente, è meglio riflettere sui rischi - i pregiudizi in particolare - che si profilano e da cui bisogna guardarsi affinché la giustizia digitale possa prosperare nel rispetto dei diritti fondamentali e dell'uguaglianza dei cittadini.

## b) Risiken

Risiken durch Verzerrungen und Risiken durch die Undurchsichtigkeit der Funktionsweise von KI.

Zunächst zu den Biases. Sie sind zweierlei Art : Es gibt diskriminierende und kognitive Biases.

Diskriminierende Biases - Illustrationen mit COMPAS.

Der Algorithmus stützt sich auf einen Fragebogen mit 137 Kriterien, die sich auf die betreffende Person beziehen, auf ihre Vorstrafen. Es gibt keinen Hinweis auf seine ethnische Herkunft. Doch 2016 zeigte die NGO ProPublica, dass der Algorithmus trotzdem die afroamerikanische Bevölkerung diskriminierte. Konkret wurden diese Bevölkerungsgruppen mit einem höheren Wert für das Rückfallrisiko gewichtet als Weiße. Umgekehrt wurden Weiße, die tatsächlich rückfällig geworden waren, doppelt so häufig mit einem niedrigen Risiko eingestuft wie Schwarze. Auch wenn es keine Kriterien gab, die mit der ethnischen Herkunft zusammenhängen, liegt mit der Verknüpfung der Daten, darunter der Wohnort (der natürlich die Zugehörigkeit zu einer ethnischen Gruppe offenbaren kann), eine mittelbare Diskriminierung vor.

## b) Rischi

I rischi di parzialità e i rischi legati all'opacità del funzionamento dell'IA

In primo luogo, i pregiudizi. Esistono due tipi di pregiudizi : i pregiudizi discriminatori e i pregiudizi cognitivi.

Pregiudizi discriminatori - illustrazione con COMPAS

L'algorithmo si basa su un questionario con 137 criteri relativi alla persona interessata, ai suoi precedenti giudiziari. Non vi è alcun riferimento all'origine etnica. Ma nel 2016 l'ONG ProPublica ha dimostrato che l'algorithmo discriminava comunque le popolazioni afroamericane. In concreto, a queste popolazioni veniva attribuito un punteggio di rischio di recidiva più alto rispetto ai bianchi. Viceversa, i bianchi che avevano effettivamente commesso una recidiva venivano classificati a basso rischio due volte più dei neri. Anche se non c'erano criteri di etnia, con la tabulazione incrociata dei dati, compreso il luogo di residenza (che può naturalmente rivelare l'etnia), si verifica una discriminazione indiretta.

Dies wird noch dadurch verstärkt, dass das System mit verzerrten Daten aus den Gerichtsurteilen selbst gespeist wird, die "das soziale Stigma in einem Land widerspiegeln, in dem Afroamerikaner bereits zu durchschnittlich 20% längeren Haftstrafen verurteilt werden als Weiße. Diese Diskriminierung war bereits im Keim in allen Maßnahmen enthalten, die die Strafen mildern und mehr denjenigen zugute kommen, die bereits sozial beschenkt sind. Man hat größere Chancen, wieder auf freien Fuß gesetzt oder auf Bewährung entlassen zu werden, wenn man einen Arbeitsplatz, ein Zuhause und eine Familie hat. Dies läuft darauf hinaus, jemanden für seine Armut zu bestrafen und ihn für die Schwierigkeiten, die er in seinem Leben hatte, bevor er eine Straftat begangen hat, zu "überverurteilen" (Garapon). Problem der Gleichheit vor der Justiz. Daher wissenschaftlich gerechtfertigte Fortführung der sozialen Vorurteile.

Ciò è aggravato dal fatto che il sistema è alimentato da dati distorti provenienti dalle stesse decisioni dei tribunali, che riflettono "lo stigma sociale in un Paese in cui gli afroamericani sono già condannati a pene mediamente più lunghe del 20% rispetto ai bianchi". Questa discriminazione era già in atto in tutte le misure che ammorbidiscono le pene e favoriscono coloro che sono già socialmente dotati. È più probabile che si venga rilasciati o messi in libertà vigilata se si ha un lavoro, una casa e una famiglia. Questo equivale a punire qualcuno per la sua povertà e a 'condannarlo' per le difficoltà che aveva nella sua vita prima di commettere un reato" (Garapon). Problema dell'uguaglianza di fronte alla legge. Quindi rinnovamento scientificamente giustificato dei pregiudizi sociali.



**Kognitive Verzerrungen** - Es gibt zwei Arten von kognitiven Verzerrungen: Verankerungsverzerrungen, die sich darin äußern, dass ein Richter nicht oder nur schwer von dem ersten Eindruck, den ihm das System vermittelt, abrücken kann. Automatisierungsverzerrungen, die sich als die Tendenz des Menschen definieren lassen, den von der Maschine erzeugten Ergebnissen mehr zu vertrauen als seinem eigenen Urteil und blind den Empfehlungen zu folgen, die die Maschine ausspricht. Wenn der Richter den von der KI vorgeschlagenen Lösungen zu 100 % folgt, ist er dann immer noch der Urheber? Dies wirft ein Problem mit der Unparteilichkeit, dem freien Ermessen und der Unabhängigkeit von Richtern auf. Der Einsatz des Algorithmus birgt die Gefahr, dass der Richter unter Druck gesetzt wird und sich nicht mehr von seinem Gerechtigkeitssinn, bei dem die intuitive Dimension sehr wichtig ist, inspirieren lässt. Risiko eines performativen Effekts oder "Schafeffekts" der prädiktiven Justiz, der die Richter zu Konformität verleitet, aber auch die Anwälte, für die es schwieriger wird, alternative Lösungen zu vertreten, die z. B. zu einer Umkehr der Rechtsprechung führen.

Langfristig besteht auch die Gefahr, dass das Fachwissen des Richters verloren geht, da er von der Maschine enteignet wird, "mit schwer umkehrbaren Folgen" (siehe Bericht des Europäischen Rates, S. 304).

**Bias cognitivi** - Si possono identificare due tipi di bias cognitivi: il bias di ancoraggio, che è l'incapacità o la difficoltà per un giudice di allontanarsi dalla prima impressione fornita dal sistema; e il bias di automazione, che è la tendenza degli esseri umani a fidarsi dei risultati prodotti dalla macchina più che del proprio giudizio e a seguire ciecamente le raccomandazioni che essa fornisce. Se il giudice segue il 100% delle soluzioni proposte dall'IA, è ancora lui l'autore?

Ciò pone un problema di imparzialità, libertà di giudizio e indipendenza dei giudici. L'uso degli algoritmi rischia di mettere sotto pressione i giudici, impedendo loro di ispirarsi al loro senso di giustizia, in cui la dimensione intuitiva è molto importante. C'è il rischio di un effetto performativo o "effetto pecora" della giustizia predittiva, che spinge i giudici a conformarsi, ma anche gli avvocati, per i quali sarà più difficile sostenere soluzioni alternative, incoraggiando ad esempio un ribaltamento della giurisprudenza.

A lungo termine, c'è anche il rischio che il know-how del giudice vada perso, espropriato dalla macchina, "con conseguenze difficilmente reversibili" (si veda il rapporto CE, p. 304).

## **Zweitens die Risiken, die mit der Undurchsichtigkeit der Funktionsweise der KI verbunden sind.**

Die Risiken des Einsatzes von Machine Learning ("Black Boxes") - Predictive Justice darf nicht "divinatorisch" sein! Die Begründung von Gerichtsurteilen ist ein Kardinalprinzip eines Rechtsstaats. Die Bürger und die Richter selbst müssen in der Lage sein, die Hintergründe der anwendbaren Entscheidung zu verstehen.

"Erfahrene Juristen sollten verlangen, die Ergebnisse zu vergleichen, die mit mehreren Methoden erzielt wurden, die etwas beunruhigende Unterschiede offenbaren. Die großen Prinzipien der Justiz, angefangen mit dem kontradiktorischen Verfahren, müssen das letzte Wort haben" (Garapon).

Diese Wachsamkeitspunkte werden allmählich berücksichtigt. Wenn auch nur in embryonaler Form, macht das Recht auf diesen Gebieten Fortschritte.

## **In secondo luogo, i rischi legati all'opacità del funzionamento dell'IA.**

I rischi legati all'utilizzo dell'apprendimento automatico ("scatole nere") - la giustizia predittiva non può essere "divinatoria"! La motivazione delle decisioni giudiziarie è un principio cardine dello Stato di diritto. I contendenti, e gli stessi giudici, devono essere in grado di comprendere le ragioni della decisione applicabile.

"I giuristi informati dovrebbero pretendere di confrontare i risultati ottenuti con diversi metodi che rivelano disparità alquanto preoccupanti. I principi fondamentali della giustizia, a partire dal contraddittorio, devono avere l'ultima parola" (Garapon).

Questi punti di vigilanza cominciano a essere presi in considerazione. Anche se in modo embrionale, la legge sta facendo progressi in questi ambiti.

### 3) Wege zu einem normativen Rahmen

**Texte** : Zunächst sind Grundsätze für den richtigen Einsatz von KI im Justizbereich durch Normen des Soft Law entstanden.

**Leitlinien** (weiches Recht) CEPEJ-Charta des Europarats von 2018 - Europäische Ethik-Charta zur Nutzung von Künstlicher Intelligenz in Justizsystemen und ihrem Umfeld, die 5 Grundsätze hervorgehoben hat :

- Grundsatz der Achtung der Grundrechte ;
- Prinzip der Nichtdiskriminierung ; - Grundsatz der Qualität und Sicherheit (Verwendung zuverlässiger Daten, die in einer sicheren technologischen Umgebung erzeugt wurden).
- Grundsatz der Transparenz, Neutralität und intellektuellen Integrität (die Methoden der Datenverarbeitung zugänglich und verständlich machen, externe Prüfungen zulassen).
- Grundsatz der Kontrolle durch den Nutzer (dem Nutzer - in diesem Fall dem Richter - ermöglichen, ein aufgeklärter Akteur und Herr seiner Entscheidungen zu sein).

Diese Grundsätze finden allmählich eine solidere rechtliche Verankerung in europäischen Verordnungen (DSGVO und in der in Vorbereitung befindlichen EU-Verordnung über KI (Hochrisiko-KI)) und in nationalen Gesetzen. Das positive französische Recht (Texte: Loi informatique et libertés 1978, geändert 2018, Justizgesetz 2019 und Rechtsprechung) ist ein gutes Beispiel für diese Rechtsentwicklung.

Einige Grundsätze, die mit den von der CEPEJ geförderten Grundsätzen in Verbindung gebracht werden können, sind im positiven Recht latent vorhanden.

### 3) Quadri normativi

**Testi** : È innanzitutto attraverso norme di soft law che sono emersi i principi di buon uso dell'IA nella giustizia. Le linee guida (soft law) Carta CEPEJ del Consiglio d'Europa del 2018 - Carta etica europea sull'uso dell'intelligenza artificiale nei sistemi giudiziari e nel loro ambiente che ha evidenziato 5 principi :

- Principio del rispetto dei diritti fondamentali
- Principio di non discriminazione
- Principio di qualità e sicurezza (utilizzare dati affidabili, prodotti in un ambiente tecnologico sicuro)
- Principio di trasparenza, neutralità e integrità intellettuale (rendere accessibili e comprensibili le metodologie di elaborazione dei dati, consentire verifiche esterne)
- Principio del controllo dell'utente (consentire all'utente - in questo caso il giudice - di essere un attore informato e padrone delle proprie scelte).

Questi principi cominciano a trovare una base giuridica più solida nei regolamenti europei (RGPD e nel regolamento europeo sull'IA (IA ad alto rischio) attualmente in fase di elaborazione) e nelle leggi nazionali. Il diritto positivo francese (testi: Loi informatique et libertés 1978 modificata nel 2018, legge sulla giustizia del 2019 e giurisprudenza) illustra bene questa evoluzione giuridica.

Alcuni principi, che possono essere avvicinati a quelli promossi dalla CEPEJ, sono latenti nel diritto positivo.

Die Grundsätze :

Der Ausschluss von vollautomatisierten Gerichtsentscheidungen : Art. 22 DSGVO: Absatz 1. Die betroffene Person hat das Recht, nicht Gegenstand einer Entscheidung zu sein, die ausschließlich auf einer automatisierten Verarbeitung, einschließlich Profiling, beruht und Rechtswirkungen zeitigt, die sie betreffen oder sie in ähnlicher Weise erheblich beeinträchtigen.

In Absatz 2 sind jedoch Ausnahmen vorgesehen, insbesondere wenn das Recht eines Mitgliedstaats dies zulässt und es Garantien für die Grundrechte gibt... Es ist also nichts verboten !

**I principi :**

L'esclusione di decisioni giudiziarie completamente automatizzate : Art. 22 GDPR: paragrafo 1. L'interessato ha il diritto di non essere sottoposto a una decisione basata unicamente sul trattamento automatizzato, compresa la profilazione, che produca effetti giuridici nei suoi confronti o che incida significativamente su di lui.

Ma il paragrafo 2 prevede delle eccezioni, in particolare se la legge di uno Stato membro lo consente e se è circondato da garanzie per i diritti fondamentali... In altre parole, nulla è vietato !

Ma questa è l'unica cosa che viene esclusa. È quindi possibile utilizzare algoritmi che non mirano a profilare le persone.

In questo caso, si immagina che si applichino le regole applicabili all'uso dell'IA nel contesto dell'elaborazione di atti amministrativi (ma probabilmente si estrapola un po').

Einige Länder sind kategorischer. Dies ist in Frankreich der Fall :

Art 47 Loi Informatique et Libertés : Keine gerichtliche Entscheidung, die eine Beurteilung des Verhaltens einer Person beinhaltet, darf auf einer automatisierten Verarbeitung personenbezogener Daten beruhen, die dazu bestimmt ist, bestimmte Aspekte der Persönlichkeit dieser Person zu bewerten (unabhängig davon, ob die Grundlage exklusiv ist oder nicht, nebenbei bemerkt).

In Frankreich wäre COMPAS also nicht möglich, da Profiling zum Zweck einer gerichtlichen Entscheidung nicht erlaubt ist. Aber das ist das Einzige, was ausgeschlossen ist. Es ist also möglich, Algorithmen zu verwenden, die nicht darauf abzielen, ein Profil von Personen zu erstellen.

In diesem Fall kann man sich vorstellen, dass die Regeln, die für den Einsatz von KI bei der Erstellung von Verwaltungsakten gelten, hier Anwendung finden (aber wir extrapolieren wahrscheinlich ein wenig).

Alcuni Paesi sono più categorici. È il caso della Francia :

Art. 47 Loi Informatique et Libertés: nessuna decisione giudiziaria che implichi una valutazione del comportamento di una persona può essere basata su un trattamento automatizzato di dati personali destinati a valutare alcuni aspetti della personalità di tale persona (indipendentemente dal fatto che la base sia esclusiva o meno).

Quindi, in Francia, COMPAS non sarebbe possibile perché la profilazione ai fini di una decisione giudiziaria non è consentita.

Ma questa è l'unica cosa che viene esclusa. È quindi possibile utilizzare algoritmi che non mirano a profilare le persone.

In questo caso, si immagina che si applichino le regole applicabili all'uso dell'IA nel contesto dell'elaborazione di atti amministrativi (ma probabilmente si estrapola un po').

Das Recht auf Kommunikation des Algorithmus :

Wenn man das auf Verwaltungsakte anwendbare Recht überträgt, kann man davon ausgehen, dass der Einzelne das Recht hat :

1/ zu wissen, dass ein Algorithmus verwendet wurde, um bei der Entscheidung des Rechtsstreits, an dem er beteiligt ist, zu helfen (Hinweis auf die Verwendung des Algorithmus in der Gerichtsentscheidung).

2/ die "Regeln zur Definition dieser Verarbeitung sowie die Hauptmerkmale ihrer Umsetzung" zu erfahren, wenn sie dies beantragen .

Die große Einschränkung dieser Regelung sind die Rechte Dritter, in diesem Fall die Rechte am geistigen Eigentum und die Wahrung des Geschäftsgeheimnisses, wenn der Algorithmus von einem Privatunternehmen entwickelt wurde.

Fall State v. Loomis aus dem Jahr 2016 (State v. Loomis, 881 N.W.2d 749 (2016), 13. Juli 2016). Dies war der Fall bei COMPAS, das von der Firma Northpointe entwickelt wurde.

Il diritto alla comunicazione dell'algorithmo :

Trasponendo la legge applicabile agli atti amministrativi, si può ritenere che il contendente abbia il diritto di :

1/ di sapere che un algoritmo è stato utilizzato per contribuire a decidere la controversia di cui è parte (menzione dell'uso dell'algoritmo nella decisione del tribunale)

2/ di conoscere le "regole che definiscono questo trattamento e le principali caratteristiche della sua attuazione", se lo richiede.

Il limite principale di questo sistema è rappresentato dai diritti di terzi, in questo caso i diritti di proprietà intellettuale e il rispetto del segreto commerciale, quando l'algoritmo è progettato da un'azienda privata.

Stato v. Loomis del 2016 (Stato v. Loomis, 881 N.W.2d 749 (2016), 13 luglio 2016). Questo è stato il caso del COMPAS, progettato dall'azienda Northpointe.

Loomis, ein Häftling, dem ein hoher Rückfallrisikoscore zugewiesen worden war und der deshalb weiter inhaftiert blieb, hatte die Entscheidung angefochten und sich auf die Undurchsichtigkeit des Systems berufen. Die Firma, die den Algorithmus entwickelt hatte, weigerte sich, die innere Logik des Algorithmus zu enthüllen, und berief sich auf das Betriebsgeheimnis. Obwohl "die Vereinbarkeit eines solchen Instruments mit den Verteidigungsrechten und dem Recht auf ein faires Verfahren fraglich war", befand der Oberste Gerichtshof von Wisconsin, dass die Unzugänglichkeit des Algorithmus die Entscheidung nicht ungültig mache, da die Software nur eine Entscheidungshilfe sei, die letztlich auf anderen Elementen beruhe.

Angesichts des Rechts auf ein faires Verfahren (Art. 6 Abs. 1 EMRK) und des Rechts auf einen wirksamen Rechtsbehelf (Art. 13) ist anzunehmen - und zu hoffen - dass diese Lösung in Europa nicht umgesetzt werden kann.

Loomis, un detenuto a cui era stato attribuito un punteggio elevato di rischio di recidiva e che quindi era stato mantenuto in custodia, aveva impugnato la decisione adducendo l'opacità del sistema. La società che ha sviluppato l'algoritmo si è rifiutata di rivelare la sua logica interna, adducendo il segreto industriale. Sebbene "la compatibilità di un tale strumento con i diritti della difesa e il diritto a un processo equo sollevasse questioni", la Corte Suprema del Wisconsin ha ritenuto che l'inaccessibilità dell'algoritmo non invalidasse la decisione, nella misura in cui il software era solo un aiuto al processo decisionale, basato in ultima analisi su altri elementi.

Per quanto riguarda il diritto a un processo equo (art. 6§1 CEDU) e il diritto a un ricorso effettivo (art. 13), si presume - e si spera - che questa soluzione non possa essere recepita in Europa.

Das Recht auf Erklärbarkeit und menschliche Aufsicht (was den "Roboterrichter" und damit die vollautomatische Entscheidung ausschließt). Die Aufsicht muss aus zwei Perspektiven betrachtet werden: die Aufsicht über die Gerichtsentscheidung, wie sie individuell angewendet wird (denn die KI darf nur eine Entscheidungshilfe sein), und die Aufsicht über den Algorithmus selbst (von seiner Konzeption bis zu seiner Funktionsweise), der auf sehr viele Fälle angewendet werden kann und kollektive Auswirkungen haben kann. Dies wird im Vorschlag für eine europäische KI-Verordnung zur Beherrschung von hochriskanten AIS (einschließlich solcher, die im Justizbereich angewendet werden) befürwortet. Sie legt den Verantwortlichen für Hochrisiko-AIS bestimmte spezifische Pflichten auf, um die Sicherheit, Zuverlässigkeit und Nachvollziehbarkeit der Handlungen bei diesen Verarbeitungen zu gewährleisten.

Il diritto alla spiegabilità e alla supervisione umana (che esclude il "giudice robot" e quindi la decisione completamente automatizzata). La supervisione deve essere vista da un duplice punto di vista: la supervisione della decisione legale applicata individualmente (perché l'IA deve essere solo un aiuto al processo decisionale) e la supervisione dell'algoritmo stesso (dalla sua progettazione al suo funzionamento) che può essere applicato a un numero molto elevato di casi e avere effetti collettivi. Questo è ciò che la proposta di regolamento europeo sull'IA sostiene per controllare le IA ad alto rischio (comprese quelle applicate alla giustizia). Il regolamento impone ai responsabili delle AIS ad alto rischio alcuni obblighi specifici per garantire la sicurezza, l'affidabilità e la tracciabilità delle azioni relative a queste operazioni di trattamento.



So müssen diese AIS während ihres gesamten Lebenszyklus von einem Risikomanagementsystem begleitet werden, das regelmäßig getestet und aktualisiert wird. Darüber hinaus müssen sie mit zuverlässigen, qualitativ hochwertigen Daten versorgt werden, um Verzerrungen zu vermeiden, was eine angemessene Überwachung und Steuerung voraussetzt. Sie müssen genau dokumentiert werden, damit die Behörden ihre Konformität bewerten können. Die Funktionsweise der Algorithmen muss automatisch aufgezeichnet werden, um die Nachvollziehbarkeit ihrer Aktivitäten zu gewährleisten. Schließlich wird von diesen AIS erwartet, dass sie während ihres gesamten Lebenszyklus ein hohes Maß an Robustheit, Sicherheit und Genauigkeit aufweisen.

Questi AIS devono essere supportati da un sistema di gestione del rischio per tutto il loro ciclo di vita, in modo da poter essere testati e aggiornati periodicamente. Devono inoltre essere alimentati da serie di dati affidabili e di alta qualità, in modo da evitare distorsioni, il che implica che siano soggetti a monitoraggio e a un'adeguata governance. Devono essere documentati con precisione, in modo che le autorità possano valutarne la conformità. Il funzionamento degli algoritmi deve essere registrato automaticamente per garantire la tracciabilità della loro attività. Infine, ci si aspetta che questi AIS abbiano un elevato livello di robustezza, sicurezza e accuratezza per tutto il loro ciclo di vita.

Zu den Garantien für menschliches Eingreifen kommen somit Anforderungen an die Prozesskontrolle hinzu, die zwar nicht garantieren, dass die auf der Grundlage der AIS getroffenen Entscheidungen tatsächlich die materiellen Regeln (Privatsphäre, Nichtdiskriminierung, Meinungsfreiheit, freie Meinungsäußerung usw.) einhalten, es aber ermöglichen, zu bescheinigen, dass sie den Form- und Verfahrensregeln entsprechen.

Schlussfolgerung zur Schwierigkeit, die Justiz zu codieren, und zu den Risiken einer entmenschlichten Justiz.

Oltre alle garanzie di intervento umano, esistono anche requisiti di controllo del processo che, pur non garantendo che le decisioni prese sulla base delle AIS siano effettivamente conformi alle norme sostanziali (privacy, assenza di discriminazioni, libertà di opinione, espressione, ecc.), consentono di attestare la conformità alle norme di forma e di procedura.

Conclusione sulla difficoltà di codificare la giustizia e sui rischi di una giustizia disumanizzata.