

## ERA DIGITALĂ ȘI JUSTIȚIA (II). OBSTACOLE GENERALE ÎN CALEA DEZVOLTĂRII IA ÎN JUSTIȚIE. CONDIȚIILE PENTRU A AVEA INSTRUMENTE DE IA EFICIENTE ÎN JUSTIȚIE



*Conf. univ. dr., av. Andreea CIUREA*  
*Facultatea de Drept*  
*Universitatea „Transilvania” din Brașov*

### **Abstract**

*In order to be an effective aid to justice, AI must be subject to permanent constraints during its creation and implementation process. IT specialists and jurists have identified certain risks and obstacles inherent in the use of AI in justice system, as well as the conditions for a healthy and efficient use of new technologies. The development of these technologies in the judiciary system must not lead to injustice or unfairness.*

*A cross-understanding of data and issues is essential, through a constant dialogue between those who can conceive and those who can use AI in justice. This interaction will be facilitated if professionals in both fields are aware of these issues from the early stages of training and if some of them specialize in them.*

**Keywords:** *risks, new technologies, cross-understanding.*

### **1. Introducere**

Pentru a reprezenta un ajutor eficient în materie de justiție, IA trebuie să fie supusă unor constrângeri permanente, pe tot parcursul procesului său de creare și implementare. Specialiștii IT și juriștii au identificat anumite riscuri și obstacole inerente folosirii IA în justiție<sup>1</sup>, precum și condițiile necesare unei utilizări sănătoase și eficiente a noilor tehnologii, astfel încât dezvoltarea acestora în mediul judiciar să nu conducă la inechitate sau nedreptate.

---

<sup>1</sup> L. Pécaut-Rivolier, S. Robin, *Justice et intelligence artificielle, préparer demain* – ep. II, publicat pe Dalloz Actualité, la 15.04.2020.

IA poate ajuta în mod semnificativ accesul la informație, algoritmi pot să prezică soluții mai bine decât un om, iar soluția matematică are, de regulă, un efect intimidant, căruia este greu să nu i te supui. Dar la ce bun? din moment ce noi nu știm cum se face predicția și de ce o dorim... căci motivarea este cea care conferă legitimitatea soluției. Așadar, prezicerea nu poate deveni un scop în sine.

Este fundamentală o înțelegere încrucișată a datelor și a problemelor, printr-un dialog constant între cei care pot concepe și cei care pot folosi IA în justiție. Această interacțiune va fi facilitată dacă profesioniștii din ambele domenii sunt conștienți de aceste probleme din faza incipientă de pregătire și dacă o parte dintre ei se specializează în ele.

## 2. Obstacole generale în calea dezvoltării IA în justiție

Riscurile folosirii IA în justiție sunt generate de alegerea bazei de învățare, de evoluția algoritmilor, de particularitățile și complexitatea activității judiciare. Învățarea automată, care caracterizează aplicațiile IA, nu este conștientă de complexitatea unor subiecte socio-umane (interpretarea legii, a fenomenelor sociale) și poate crea mai multe probleme decât soluții.

*Aceste riscuri și limitări sunt interdependente, se susțin și se potențează unele pe altele.*

### a) Riscul „supraînvățării”

Majoritatea algoritmilor de învățare se bazează pe baze de date în care se află înregistrat un număr mare de răspunsuri umane la cazuri specifice. Formula de predicție produsă de un algoritm este o funcție *de învățare*. Prin urmare, formula de predicție depinde de baza de învățare, ceea ce generează riscul „supraînvățării” sau *generalizarea* formulei de predicție. Pentru informaticieni nu este dificil să creeze o funcție care să nu facă nici o eroare asupra tuturor cazurilor care constituie baza de date: este suficient să învețe „pe de rost” toate aceste cazuri. Însă, o astfel de formulă produce o predicție nepotrivită pentru un caz care nu se află în baza de date.

O practică obișnuită, numită „validare încrucișată”, constă în separarea bazei de date în două: cazurile din prima parte utilizate pentru a stabili formula de predicție și cazurile din partea a doua fiind folosite pentru a evalua performanța formulei în cazuri noi. Această abordare face posibil controlul capacității de „generalizare” a formulei de predicție.

### b) Riscul „opacității” formulei de predicție

De regulă, „predictorii” furnizați de IA funcționează ca niște *cutii negre* atât pentru utilizatorul-jurist, cât și pentru justițiabil: complexitatea formulei de predicție împiedică înțelegerea acesteia și anticiparea unei soluții doar la simpla

vedere a formulei. Justițiabilul nu are capacitatea de a cunoaște, de fapt, regula care îi este aplicată. Și aici se naște întrebarea: este suficientă încrederea în obiectivitatea mașinii pentru a justifica ignoranța în care se găsește destinatarul deciziei? O soluție ar consta în face public predictorul, astfel încât fiecare participant procesual să poată dobândi cel puțin o cunoaștere empirică a previziunilor oferite de mașină. Însă, chiar dacă inițial este posibil să înțelegi o formulă învățată pe o primă bază de date, ulterior – pe măsură ce devin disponibile noi decizii sau opinii – comportamentul algoritmului va fi modificat, astfel încât dobândirea unei înțelegeri în timp real va fi dificilă.

### c) Incapacitatea de a da sens hotărârilor judiciare

Unul din principalele obstacole în calea utilizării pe scară largă a IA în probleme de justiție este *incapacitatea de a da sens hotărârilor judiciare*. Formalismul matematic este performant pentru sarcini foarte specializate, dar *nu poate dezvălui complexitatea raționamentului juridic<sup>2</sup>, nu poate explica deciziile instanțelor și comportamentul judecătorilor<sup>3</sup>.*

Această problematică este riguros analizată și în Carta Etică a CEPEJ (Commission européenne pour l'efficacité de la justice)<sup>4</sup>.

Astfel, specialiștii ne atrag atenția că obiectivul principal al proiectanților de algoritmi nu este să înțeleagă funcționarea modelelor construite automat de computer, ci să se asigure că aparatul reușește să se apropie treptat de rezultatele așteptate. Utilizate în medii închise, cu date cuantificabile, rezultatele pot fi impresionante. Însă, formalismul matematic este dificil de aplicat în medii deschise și mai greu de cuantificat: atunci când se încearcă măsurarea fenomenelor sociale „este mult mai dificil să se găsească o relație stabilă” decât în științele naturii<sup>5</sup>, căci aceste sisteme de calcul complexe nu sunt replici ale creierului uman (care include procese perceptivă, învățare, autoorganizare, adaptare)<sup>6</sup>.

Specificul activității judecătorului uman rezidă în metoda de raționament pe care acesta o pune în operă. Acest raționament, cunoscut sub numele de *silogism judiciar*, este piatra de temelie în funcționarea justiției umane. Într-un mod simplificat, implică trei pași de reflecție în fiecare situație prezentată judecătorului: examinarea faptelor cauzei (este „minorul”); căutarea regulii de drept aplicabile faptelor (este „majorul”); aplicarea regulii la fapte (aceasta este concluzia silogismului).

<sup>2</sup> Y. Meneceur, *Intelligence artificielle et mémoire de la justice: le grand malentendu*, Les cahiers de la justice 2019 p. 277.

<sup>3</sup> Y. Meneceur, *Quel avenir pour une justice prédictive*, JCP 2018, Doctr. 190.

<sup>4</sup> Commission européenne pour l'efficacité de la justice (CEPEJ), Conseil de l'Europe, *Charte éthique européenne d'utilisation de l'intelligence artificielle dans les systèmes judiciaires*: <https://www.coe.int/fr/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment>.

<sup>5</sup> P. Jensen, *Pourquoi la société ne se laisse pas mettre en équation*, Éditions du Seuil, 2018.

<sup>6</sup> Y. Meneceur, *Intelligence artificielle et mémoire de la justice: le grand malentendu*, Les cahiers de la justice 2019, p. 277.

Raționamentul silogic trebuie să apară în motivarea hotărârii judecătorești, fiind cea care permite destinatarului să fie sigur de existența și fundamentul raționamentului judecătorului, de capacitatea sa de a include, în sistemul normelor juridice identificabile, elemente de fapt complexe. Așadar, *motivarea asigură transparență*. Este garanția că decizia luată este direct legată de litigiu și calificarea juridică dată situației este corectă.

În aceste condiții, se observă limitele aplicării algoritmilor în justiție, deoarece fazele esențiale ale raționamentului nu mai sunt „vizibile” și inteligibile.

Apoi, silogismul judiciar nu reflectă întreaga motivație folosită judecător, care este punctată uneori de interpretări și alegeri subiective, ascunse sau chiar inconștiente și care nu poate fi formalizată *a priori*. Învățarea automată este ineficientă în realizarea acestei munci de interpretare, proiectanții sperând cel mult că detectarea regularității în contexte lexicale reușește să reproducă la ieșire aceleași efecte ca procesul de luare a deciziilor care a produs aceste date.

Este semnificativ experimentul făcut de o echipă de cercetători englezi și americani<sup>7</sup>, specialiști în informatică, psihologie pozitivă, știința informației și drept, care a dezvoltat un algoritm capabil „să prezică” deciziile Curții Europene a Drepturilor Omului prin corelarea faptelor, argumentelor părților și a legii pozitive relevante. Programul vizează încălcarea a trei articole din Convenția europeană pentru protecția drepturilor omului și a libertăților fundamentale: interdicția torturii și a tratamentelor degradante (art. 3), dreptul la un proces echitabil (art. 6) și dreptul la respectarea vieții private (art. 8). Nu a fost întâmplătoare alegerea jurisprudenței CEDO, căci deciziile Curții pot furniza un eșantion reprezentativ și explică în detaliu faptele, argumentele părților, precum și dispozițiile legale relevante. Prin urmare, pentru fiecare caz, algoritmul a avut la dispoziție suficiente „intrări” pentru a-și îndeplini sarcina.

Rezultatele cercetărilor au fost publicate pe 24 octombrie 2016, într-un articol intitulat „Prezicerea deciziilor judiciare ale Curții Europene a Drepturilor Omului: o perspectivă de procesare a limbajului natural”<sup>8</sup>, anunțând că algoritmul creat a reușit să clasifice corect 79% din deciziile Curții Europene a Drepturilor Omului (încălcare sau nu). Analizând aproape 600 de cazuri judecate de Curtea Europeană a Drepturilor Omului, algoritmul a dat aceeași soluție ca magistrații în 8 din 10 cazuri.

Acest experiment de justiție algoritmică este bogat în lecții, generând o serie de întrebări și dezbateri<sup>9</sup>: faptul că un judecător uman și un judecător-robot, în

<sup>7</sup> Cercetători provenind de la universități din Londra, Sheffield, Pennsylvania.

<sup>8</sup> N. Aletras, V. Lampos, D. Tsarapatsanis, D. Preotiuc-Pietro, *Predicting Judicial Decisions of the European Court of Human Rights: a Natural Language Processing Perspective*, Peer Journal of Computer Science, 24 oct 2016 (disponibil pe <https://peerj.com/articles/cs-93/>).

<sup>9</sup> Pentru un studiu amănunțit, a se vedea: B. Barraud, *La croisée des savoirs - Un algorithme capable de prédire les décisions des juges: vers une robotisation de la justice?*, Les cahiers de la justice 2017, p. 121.

prezența aceluiași fapte, conduc adesea la aceleași soluții, nu este dovada că avem deja o justiție cvasi-mecanică? și acest fapt nu ar trebui salutat din punct de vedere al securității juridice și al previzibilității justiției? și dacă, în ceea ce privește cele două cazuri în care deciziile sunt divergente, sentințele judecătorului robotizat sunt cele care decurg cel mai logic din aplicarea strictă a dreptului pozitiv asupra faptelor, în timp ce ale judecătorului uman ar fi dictate de considerații externe, implicite? Apoi, care va fi evoluția profesiilor juridice: înlocuirea magistraților, avocaților, cu mașini „inteligente” care spun legea în mod automat, polițiști-roboti care arestează oamenii (în imaginația tehnopesimiștilor)? Sau va fi o asociere între magistrați/avocați/notari și aceste mașini „inteligente” – din perspectiva celor mai moderați?

Cele 21% de discrepante dintre hotărârile efective și hotărârile furnizate de algoritm arată că și alte elemente, în afară de faptele și dispozițiile legale relevante, joacă uneori un rol decisiv în alegerile magistraților. Dacă luarea deciziilor jurisdicționale ar respecta o logică formală strictă, cercetătorii ar fi reușit să construiască un judecător algoritmic mai performant, în acord cvasi-total cu judecătorul uman.

Originile diferențelor dintre deciziile judecătorului robot și cele ale judecătorului uman sunt multiple. Comportamentul judiciar este un „puzzle”, ale cărui componente nu au doar forme regulate. Judecătorul american Oliver Wendell Holmes a scris, la sfârșitul sec. XIX, că „viața Dreptului nu este logică; este experiență”<sup>10</sup>, iar Jean Carbonnier, unul dintre cei mai mari juriști ai secolului XX, sublinia că „judecătorul este un om și nu o mașină de silogism: la fel de mult cu cunoștințele sale despre reguli și cu logica sa, el judecă cu intuiția și sensibilitatea sa”<sup>11</sup>. Cunoașterea regulilor și a logicii sunt primare, dar intuiția și sensibilitatea joacă un rol complementar, influențând orientarea deciziilor judecătorești mai ales în cele mai dificile cazuri.

În lucrarea menționată<sup>12</sup>, cercetătorii indică faptul că „algoritmii au avut tendința de a oferi judecăți eronate atunci când se confruntă cu două cazuri similare care au condus la condamnare și necondamnare”, respectiv atunci când faptele identice sau cel puțin similare sunt tratate diferit în funcție de judecătorii cărora le sunt supuse. Cu alte cuvinte, judecătorul-robot și judecătorul uman au fost de acord mai ales în cazurile centrale și clare, în care regula este aplicabilă cu certitudine, și mai puțin în cazurile pentru care există motive atât pentru a afirma, cât și pentru a nega că se aplică o anumită regulă.

Marja de eroare a judecătorului robot nu arată cine greșește, ci că judecătorul-robot spune dreptul altfel decât judecătorul uman. Atunci, ar trebui să avem

<sup>10</sup> O.W. Holmes, *The Common Law*, MacMillan, 1881, p. 1, citat de B. Barraud, *op. cit.*, Les cahiers de la justice 2017 pp. 121-139.

<sup>11</sup> J. Carbonnier, *Droit civil vol. I: Introduction*, Puf, coll. Quadrige, 2004, p. 23.

<sup>12</sup> N. Aletras, V. Lampos, D. Tsarapatsanis, D. Preotiuc-Pietro, *op. cit.*

încredere în primul sau în al doilea? Este, credem, dificil să preferi justiția unei mașini decât justiția umană.

S-a susținut că anumite aplicații ale IA pot identifica prejudecăți datorate persoanei magistraților, *susceptibile* să hrănească suspiciuni de parțialitate. Dar, putem reuși să facem lumină asupra comportamentului judecătorilor pe baza unui tratament algoritmic al apariției numelor lor în anumite hotărâri judecătorești? IA construiește modele încercând să dezvăluie corelații ascunse într-o cantitate mare de date. Dar simpla corelație statistică dintre două evenimente nu este suficientă pentru a explica factorii cu adevărat cauzali. Măsurăm ceea ce poate fi eronat – științific și ideologic, din ambiția de a face deciziile judiciare obiective. Furnizarea unei explicații reale a unei hotărâri ar necesita o analiză mult mai detaliată a datelor fiecărui caz și nu ar putea apărea spontan dintr-o masă de legături lexicale în care apare numele unui magistrat. De exemplu, faptul că un judecător în materie de familie stabilește, mai des statistic, domiciliul unui copil la mamă nu reflectă neapărat o părtinire a acestui magistrat în favoarea femeilor, ci se poate explica și prin existența unor factori sociali, economici și culturali specifici populației din jurisdicția sa.

Riscurile explicațiilor false ale hotărârilor judecătorești sunt, prin urmare, extrem de ridicate. Justiția algoritmică va constitui un fel de *memorie mioapă* a justiției<sup>13</sup>, fiind lipsită de o analiză profundă a elementelor cauzale ale hotărârilor date de judecătorul uman. Valoarea statistică a rezultatelor obținute va rămâne discutabilă în absența unui control real al datelor prospectate și al efectului „cutiei negre” al anumitor algoritmi.

Desigur, există situații pentru care expunerea unor date obiective (timpul de lucru în întreprindere) poate corespunde unui răspuns obiectiv precis (durata concediului plătit). Acestea sunt cazuri în care două sau trei întrebări simple și recurente pot duce la o soluție, fără risc de eroare. Astfel de cazuri, bazate pe un silogism simplu, se pot împrumuta modelării mecaniciste, dar care nu este exact cea adoptată în general de IA. *Codul digital al muncii* din Franța oferă o astfel de ilustrare<sup>14</sup>.

Acest *Cod digital al muncii* este o platformă online, gratuită, care a început să funcționeze în ianuarie 2020. Pentru a realiza această platformă, o echipă de juriști și specialiști IT au lucrat împreună, timp de aproape doi ani. Introdus prin Ordonanța nr. 2017-1387 privind previzibilitatea și securitatea relațiilor de muncă, scopul codului digital este de a face dreptul muncii – prin utilizarea IA – „accesibil și eficient pentru toți”, deoarece reglementările în domeniu constituie uneori un

---

<sup>13</sup> Y. Meneceur, *Intelligence artificielle et mémoire de la justice: le grand malentendu*, Les cahiers de la justice 2019, p. 277.

<sup>14</sup> <https://www.fabrique.social.gouv.fr/startups/code-du-travail-numerique>.

labirint complex pentru angajați și angajatori. Concret, codul digital al muncii oferă următoarele servicii:

- **o bază de date** care integrează textele codului muncii și ale principalelor convenții colective, precum și fișe explicative, permițând utilizarea unor cuvinte-cheie în franceza fluentă și nu neapărat în limbaj juridic, pentru a obține trimiterea la documentele care par cele mai potrivite;

- **o cutie de instrumente**, care include simulatoare pentru calcularea despăgubirilor frecvente care pot fi ușor stabilite din răspunsul la unele date factuale (indemnizații de concediere, preaviz de demisie, salariu brut/net) și modele de acte.

Acest instrument este o bază de date inteligentă, dar nu este destinat efectuării unei analize juridice<sup>15</sup>. Poate da o potrivire textuală unui cuvânt cheie, dar nu poate ajuta la calificarea juridică a faptelor care i-au fost supuse. Apoi, utilizatorul este invitat să răspundă la o serie de întrebări închise care orientează iterativ sistemul. Acest proces imită raționamentul juridic pentru a ajunge la determinarea regulii de drept aplicabile și stabilirea unei (game de) durate sau compensații prevăzute de lege. Dacă situația este atipică, este posibil ca sistemul să nu ofere un răspuns și să sugereze contactarea unui departament juridic.

Pentru a determina textele și reglementările cele mai relevante, sistemul folosește un algoritm analog cu cel al motorului de căutare Google, care se bazează în principal pe asocieri statistice între cuvinte observate într-o bază de date mare. Sistemul s-a bazat inițial pe un *algoritm de învățare nesupravegheat*, în sensul în care oferă texte „apropiate” de cererea formulată; apoi, un algoritm *supravegheat* face posibilă o mai bună distincție între răspunsurile utile și cele mai puțin utile.

Acest instrument AI nu este imun la prejudecăți inerente oricărui sistem de învățare (dependent de baza de date de învățare, efect al alegerii unuia sau altuia criteriu de ajustare etc.), dar natura nedecizională a răspunsului pe care îl produce reduce evident consecințele acestor prejudecăți.

Legat de îmbunătățirea acestei platforme, s-a pus întrebarea despre capacitatea de a răspunde, prin integrarea jurisprudenței relevante, la întrebări mai specifice de la utilizatori („Pot fi demis pentru o întârziere la munca?”). Se poate oferi codului digital al muncii capacitatea de a furniza analize și un răspuns juridic la o situație de fapt ridicată de un individ? Opiniile specialiștilor și ale reprezentanților statului sunt în sensul că programului trebuie să-i fie limitate voluntar ambițiile „predictive” la un domeniu clar demarcat și cuantificabil al dreptului muncii; lărgirea ambiției predictive l-ar denatura și ar oferi garanții mai puține cetățenilor.

---

<sup>15</sup> Pentru detalii, a se vedea: L. Pécaut-Rivolier, S. Robin, *Justice et intelligence artificielle, préparer demain* – ep. III, publicat pe Dalloz Actualité, la 17.04.2020.

#### **d) Incapacitatea adaptării deciziei judiciare la circumstanțele concrete ale unei spețe**

Există *limitări ale dezvoltării IA* legate de natura specifică a activității judiciare, de necesitatea adaptării hotărârii la circumstanțele concrete ale unei spețe. Vrem să renunțăm, în favoarea predictibilității, la adaptarea deciziei judiciare raportat la caracteristicile specifice ale fiecărei spețe? Ne dorim un sistem penal care să nu lase putere de decizie judecătorului uman cu privire la pedeapsa aplicată sau un sistem informatic care codifică într-un mod automat condițiile de stabilire a domiciliului ale unui copil minor? Sancțiunea penală nu va mai fi legată strict de faptă și de autorul ei, ci se va adapta la previzibilitatea unei recidive atestate matematic.

*Este o alegere mai mult politică decât juridică, de a decide că, pentru anumite dispute, este mai valoroasă o soluție previzibilă și sigură decât o soluție adaptabilă și mai puțin previzibilă.*

Cert este că, indiferent de denumirea lor comercială, toate aplicațiile algoritmilor în justiție stabilesc tipuri de probabilități asupra rezultatului unui conflict interuman, tratând *nu o poveste brută, ci fapte deja calificate legal*. Construcția modelelor matematice se va baza pe o materie primă (pe hotărârile judecătorești deja pronunțate în aceste tipuri de litigii) și pe învățarea automată pentru a descoperi *corelații de grupuri lexicale*. Concret, aparatul va analiza, pe baza unor parametri identificați de proiectanți (pentru prestațiile compensatorii, de exemplu: durata căsătoriei, disproporția situației patrimoniale, vârsta și starea de sănătate a părților) posibilele legături cu soluții deja pronunțate de instanță (quantumul prestațiilor stabilite prin hotărâri anterioare). Cu toate acestea, aplicarea acestei abordări într-un proces judiciar necesită prudență maximă, întrucât corelațiile lingvistice descoperite în decizii sunt departe de a putea fi interpretate ca relații cauză-efect, în toate situațiile<sup>16</sup>.

#### **e) Incapacitatea algoritmilor de a substitui legea însăși**

Calculule probabilistice nu pot epuiza rolul legii. Se poate discuta aici despre problema utilizării baremelor judiciare: dacă aceste bareme au întotdeauna relevanță statistică, ele nu pot, prin simplul fapt al existenței lor, să substituie legea însăși, în special prin reducerea marjei de interpretare pe care norma de drept o oferă judecătorului.

Legea, deși imperfectă, este *expresia unui proiect social*, prin care un popor își stabilește valorile și prioritățile, pe când algoritmi ne impun, printr-un formalism științific aparent neutru, o nouă rațiune care este, în realitate, doar a proiectanților lor. În aceste condiții, democrația pare a fi înlocuită cu o nouă formă de aristocrație – *o aristocrație digitală*, cum o denumesc magistratul francez Y. Meneceur<sup>17</sup>.

<sup>16</sup> D. Cardon, *À quoi servent les algorithmes. Nos vies à l'heure des big data*, Seuil, La République des idées, 2015.

<sup>17</sup> Yannick Meneceur este magistrat francez, detașat la Consiliul Europei, cercetător asociat la IHEP (l'Institut des Hautes Études sur la Justice).



### f) Riscul decalajului digital

Vorbind despre aristocrația digitală, trebuie să adăugăm *riscul decalajului digital*<sup>18</sup>, care există nu numai pentru persoanele în vârstă, ci și pentru tinerii care, deși știu să găsească filme și muzică pe Internet, nu sunt la fel de pricepuți în completarea formularelor computerizate.

Într-un Raport din 2019 al Avocatului Poporului (Apărătorul Drepturilor) din Franța se arată că 20% dintre francezi nu au computer, 27% nu au un smartphone și peste 25% declară că au dificultăți în a găsi informații administrative pe internet<sup>19</sup>. În preambulul acestui Raport, se subliniază: „Dacă o singură persoană ar fi lipsită de drepturile sale din cauza digitalizării unui serviciu public, ar fi un eșec pentru democrația noastră și pentru Statul de drept”.

### g) Riscul reproducerii discriminării

Și mai grav, modelele de învățare pot reproduce și agrava discriminarea. Diferitele tehnici de învățare automată par neutre în ceea ce privește valorile sociale, căci științele exacte care le conduc sunt un formalism. Pe de altă parte, utilizarea acestui formalism cu o metodă și date părtinitoare va duce în mod sistematic la rezultate părtinitoare.

Este deja celebru exemplul algoritmului COMPAS<sup>20</sup>, utilizat în mod eficient în anumite state americane pentru a evalua pericolul indivizilor în vederea posibilei lor plasări în detenție provizorie sau pentru pronunțarea unei condamnări penale. Acest algoritm se bazează pe studii academice în criminologie și sociologie, pe diferite modele statistice și prelucrarea unui chestionar cu peste 130 de intrări, referitoare la persoana în cauză și la trecutul său judiciar fără nici o referire la originea sa etnică. Sistemul oferă apoi judecătorului „scoruri” diferite pe un orizont de doi ani: risc de recidivă, risc de comportament violent și risc de neprezentare pentru situații de detenție preventivă.

Însă, în mai 2016, reprezentanții ONG-ul ProPublica au analizat eficacitatea „predicțiilor” COMPAS asupra unei populații de aproape 10.000 de persoane arestate în ținutul Broward (Florida) între 2013 și 2014<sup>21</sup>. Acest studiu a găsit o rată relativ scăzută a „predicțiilor” corecte (61%) și analizând „falsurile pozitive”, a constatat că populațiile afro-americane erau ponderate cu risc mai mare de recidivă decât populațiile albe. În schimb, populațiile albe care au recidivat efectiv fuseseră de două ori mai clasificate ca fiind cu risc scăzut decât populațiile

<sup>18</sup> E. Jeuland, *Justice numérique, justice inique?*, Dalloz, Les cahiers de la justice 2019, p. 193.

<sup>19</sup> A se vedea: *Dématérialisation et inégalités d'accès au service public*, 2019, pe [www.defenseurdesdroits.fr](http://www.defenseurdesdroits.fr).

<sup>20</sup> Correctional Offender Management Profiling for Alternative Sanctions - este un algoritm dezvoltat de societatea privată Equivant (ex-Northpointe): <http://www.equivant.com/solutions/inmate-classification>.

<sup>21</sup> Studiu disponibil pe <https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>.

afro-americane. Așadar, fără a include etnia indivizilor, doar încrucișarea datelor (spre exemplu, locul de reședință) a supraponderat indirect acest aspect în detrimentul altor factori sociali individuali (educație, ocuparea forței de muncă, parcurs familial) și a condus la influențarea judecătorilor cu indicatori discriminatori. A reduce soarta unui individ la cea a grupului statistic căruia îi aparține nu înseamnă o reparație a unei doctrine penale deterministe...?

Amintim aici și cazul „*Loomis vs Wisconsin*”<sup>22</sup>: un tânăr condamnat la șase ani de închisoare pentru fapte pe care le-a negat, pe baza unei evaluări făcute de software-ul COMPAS. El a contestat această decizie în fața Curții Supreme a statului Wisconsin, argumentând că a fost condamnat fără să știe cu exactitate ce informații au fost utilizate de instrument, că metodologia utilizată de acest software (care este secret comercial) a fost necunoscută judecătorilor care l-au folosit pentru a-l evalua și că această decizie nu ia în considerare dreptul său la o sentință individualizată. Curtea Supremă a respins recursul pe motiv că programul COMPAS este doar un instrument aflat în serviciul judecătorilor.

Problema este aceeași pentru cazurile civile, administrative sau comerciale: ar fi interesant, de exemplu, să examinăm dacă, pentru o serie de beneficii compensatorii în care situația conjugală și economică este echivalentă obiectiv, sumele „prezise” de algoritmi vor fi ponderate diferit în funcție de domiciliul părților. Dacă da, ce criterii, evidente sau nu, ar fi putut avea o influență? Fără transparență cu privire la aceste mecanisme, riscăm ca o „cutie neagră” să influențeze rezultatul unei dispute într-o manieră discreționară și să reproducă posibile inegalități.

#### **h) Riscul înlocuirii statului de drept cu o societate guvernată de date**

Există riscul înlocuirii actualului cadru de referință, respectiv a statului de drept, cu o societate guvernată de date<sup>23</sup>, drum pe care ne îndreaptă aplicațiile formalismului matematic. Aceasta este o provocare majoră a timpului nostru: o societate guvernată de date (*data driven society*) ar fi mai eficient funcționabilă?<sup>24</sup>

Dreptul a devenit o țintă pentru antreprenorii care sunt în căutarea de noi piețe și care profită de aplicațiile IA pentru a dezvolta soluții comerciale într-un număr mare de sectoare (comerț, transporturi, urbanism, medicină etc.). În același timp, nevoia de a face răspunsul judiciar mai previzibil, combinată cu silogismul deciziilor judiciare, i-a făcut pe matematicieni să creadă că există material pentru

<sup>22</sup> Pentru detalii, a se vedea C.M. Drăgușin, *State of Wisconsin v. Loomis* sau intersecția dintre inteligența artificială și dreptul fundamental la un proces echitabil, în Revista „Dreptul” nr. 4/2020, p. 108; L.M. Stănilă, *Inteligența Artificială. Dreptul penal și sistemul de justiție penală, Amintiri despre viitor*, Editura Universul Juridic, București, 2020, pp. 175-178.

<sup>23</sup> Y. Meneceur, *Intelligence artificielle et mémoire de la justice: le grand malentendu*, Les cahiers de la justice 2019, p. 277.

<sup>24</sup> Despre *dataism*: a se vedea M. Hotca, *Impactul inteligenței artificiale (IA) asupra vieții sociale și afacerilor*, disponibil pe <https://www.universuljuridic.ro/impactul-inteligenței-artificiale-ia-asupra-vieții-sociale-si-afacerilor/2/>.

exploatare, astfel că investitorii au sponsorizat dezvoltarea costisitoare a soluțiilor bazate pe modele matematice de învățare automată în domeniul judiciar.

Magistratul și cercetătorul francez Y. Meneceur atrage atenția asupra construirii unor „catedrale algoritmice complexe” – cu mari investiții financiare și umane – unde „tehn-evangheliștii” ne promet că dezvăluie un adevăr immanent, surprins din corelații de învățare automată<sup>25</sup>. Cu toate acestea, ne avertizează magistratul, acești „coloși” au fundații fragile, pentru că sistemele juridice rămân foarte deschise și o schimbare poate invalida decenii de construcție a jurisprudenței, în timp ce algoritmi sunt construiți, mai presus de toate, pe statistici ale evenimentelor din trecut.

### **i) Riscul unei „dejudiciarizări” forțate**

Într-o primă etapă, se produce *riscul unei accelerări intempestive și fragmentării actului de justiție*: timpul prin termene, faptele prin procesarea în masă a documentelor, pledoarii tot mai absente sau limitate în timp și în puncte precise, judecăți pe cadre sau tratamentul separat al soluțiilor prin instrumentele așa-numitei justiții predictive.

IA este implementată în justiție pentru a răspunde la criticile cu privire la costul, încetinirea și imprevizibilitatea sa. Renunțăm la varii interpretări și dezbatere, dar câștigăm timp și previzibilitate.

Evident, reducerea limbajului oral și a tensiunii dezbaterii în favoarea ecranului afectează dimensiunea personală și pasională a oricărui proces. Nu mai este o privire pe care o avem în fața noastră, ci un ecran fără prezență, o voce fără origine. Conexiunea înlocuiește relația, corelația înlocuiește dovada. Această *discontinuitate între a vedea, a auzi și a simți*, între text și context, *sărăcește experiența dreptății*<sup>26</sup>.

*Abundența de date, omniprezența ecranelor, strategiile predictive* sunt caracteristici ale revoluției digitale care afectează componentele funcției de a judeca și poate duce la o „dejudiciarizare forțată”.

Această intervenție „perturbatoare” are un *impact direct asupra practicienilor Dreptului* care sunt tot mai provocați de noua „justiție digitală”. Dar voi reveni asupra acestei problematici spinoase într-un studiu viitor.

## **3. Condiții generale pentru a avea un instrument de IA eficient în justiție**

Pentru a avea un instrument eficient de IA care oferă informații diferiților actori ai justiției, ajutându-i și dându-le siguranță, este esențială întrunirea unor condiții prealabile:

<sup>25</sup> Y. Meneceur, *idem*. Despre „religia datelor”, a se vedea și Y.N. Harari, *Homo deus*, Ed. Polirom, București, 2018, p. 321.

<sup>26</sup> D. Salas, *Data, écrans, prévisions*, Les cahiers de la justice 2019, p. 201.

### a) Transparența algoritmilor

Orice algoritm se bazează pe o serie de alegeri în materie de reprezentare a observațiilor, măsurare a similarității sau criterii de optimitate. Dacă aceste opțiuni nu sunt controlate de utilizator, ci sunt automate, aparatul va avea propriile sale criterii de determinare.

Din punctul de vedere al proiectantului algoritmului, este important ca problema să fie gestionabilă matematic; pentru programator, confortul matematic constituie o prima regulă de alegere. Astfel, dacă este vorba de *compararea a două texte pornind de la frecvența unei serii de cuvinte în interiorul lor*, programul se va baza pe distanțele (pozitive sau negative) între aceste frecvențe. Pentru a neutraliza semnul pozitiv sau negativ, care contează prea puțin pentru a verifica abaterile, se poate folosi pătratul abaterii sau valoarea sa absolută, iar rezultatele furnizate de algoritm vor fi diferite, în funcție de utilizarea valorii pătrate sau absolute.

Criteriul de măsurare a potrivirii dintre o *predicție și răspuns observat în jurisprudența existentă* reglează importanța acordată posibilelor erori. Astfel, să luăm cazul unei *predicții binare* („responsabil” / „non-responsabil”): eroare înseamnă a prezice „responsabil” într-o situație juridică în care justițiabilul nu ar trebui să răspundă sau a prezice „neresponsabil” într-un caz legal de răspundere. Criteriul de măsurare a potrivirii poate presupune că ambele erori sunt la fel de dăunătoare sau, dimpotrivă, să acorde o greutate mai mare primei; spre exemplu, în cazul predicției cuantumului unei despăgubiri, în funcție de criteriului ales, o predicție mai mare cu 10% pentru o indemnizație va fi sau nu considerată echivalentă cu o sub-predicție de 10%.

Astfel, *cunoașterea precisă a datelor alese pentru alimentarea bazei de date și alegerea criteriul similarității sau al optimității constituie chei esențiale ale sistemului*. Aceste opțiuni, aparent benigne și fără interes, vor decide, de fapt, orientările răspunsurilor și predicțiilor, fără ca nimeni să fi fost conștient sau să le fi gândit. Prin urmare, este responsabilitatea juriștilor să se asigure că fiecare acțiune operațională este verificată sub aspectul consecințelor sale.

### b) Asigurarea imparțialității (fairness) și loialității (loyalty)

Abordarea mimetică adoptată de majoritatea algoritmilor de învățare înseamnă că formulele de predicție pe care le produc *tind să întrețină prejudecățile prezente în bazele de date care le-au alimentat inițial*. Din punct de vedere statistic, prejudecata este o eroare sistematică cu privire la un obiectiv de predicție.

Astfel, s-au observat prejudecăți sistematice în deciziile judecătorești bazate pe originea socială sau etnică a justițiabililor, pe care algoritmi, care „s-au instruit” pe baza acestor decizii, le-au reprodus fidel; există cercetări actuale care se concentrează pe posibilitatea de a detecta astfel de prejudecăți și a le corecta automat<sup>27</sup>.

<sup>27</sup> S. Barocas, M. Hardt, A. Narayanan, *Fairness in machine learning*, NIPS Tutorial, 2017.

Aceasta implică, de exemplu, să ne asigurăm că, dacă spețele a doi litiganți diferă doar prin originea lor etnică, soluția furnizată de algoritm va fi aceeași (sanctiunea, valoarea compensației etc.).

Noțiunea de *fairness* utilizată în IA poate completa astfel noțiunea de *loyalty* (cu sensul de *bună-credință*), prima urmărind să corecteze astfel de prejudecăți în cadrul algoritmului însuși (rămâne de decis cine ar avea responsabilitatea pentru această intervenție), în timp ce a doua impune o corecție externă făcută de judecătorul uman. Doar în aceste condiții, contribuția sa poate fi considerabilă.

Cuvântul *loialitate* este frecvent folosit atât în discursurile juriștilor, cât și în cele ale specialiștilor în IA, dându-se uneori sensuri foarte diferite. Pentru juriști, Loialitatea, adesea legată de buna-credință, constituie un principiu transversal oricărei acțiuni în exercitarea unui drept, fiind menționată frecvent în relațiile contractuale în dreptul civil, în dreptul muncii, în desfășurarea proceselor etc.: „Atitudinea loială se referă la respectarea regulilor probității și onoarei”<sup>28</sup>.

*Loialitatea* găsește un ecou special în materie de IA. Pe de o parte, *în amonte de implementarea instrumentului*, ne obligă să ne întrebăm despre sensul regulii înainte de a-i decide consecințele. Pe de altă parte, *în aval*, ne autorizează să ieșim din instrument ori de câte ori soluția furnizată contrazice noțiunea de loialitate, adică are ca rezultat o aplicare denaturată a regulilor. Așadar, raportat la IA, loialitatea presupune un control extern. În acest context, judecătorul uman poate corecta prejudecăți de fapt, legate de o situație de constrângere, sau prejudecăți sociale, legate de o situație de discriminare, de exemplu.

Am mai afirmat că legat de folosirea IA în justiție nu există neutralitate, există doar opțiuni care nasc întrebări asupra *liniei de echilibru* pe care dorim să o stabilim între *răspunsul securizat, complet digital și răspunsul adaptat*, pentru care omul rămâne cel mai bun garant<sup>29</sup>. Iar conceptele de *fairness* și *loyalty*, chiar înțelese ușor diferit de matematicieni și juriști, pot ajuta la trasarea acestei frontiere.

### c) Nevoia unei abordări multidisciplinare

Pentru a face să se întâlnească dreptatea și IA, trebuie se întâlnească oamenii care le acționează. Pentru proiectarea instrumentelor de IA, să existe o *echipă multidisciplinară* formată din juriști, informaticieni, statisticieni, matematicieni, pentru o perioadă suficient de lungă, care să permită tuturor să se identifice și să înțeleagă ce face celălalt: juriștii să știe exact instrucțiunile care au fost date și specialiștii IT să poată împărtăși, în detaliu, alegerile cu care se confruntă și la care pot avea tendința de a răspunde implicit. Dificultatea este ca o astfel de echipă să se mențină

<sup>28</sup> W. Cherbonnier et al., *La loyauté: de la règle morale au principe juridique*, Rev. jur. Ouest, 2012, p. 327; disponibil și pe [https://www.persee.fr/doc/juro\\_0990-1027\\_2012\\_num\\_25\\_3\\_4730](https://www.persee.fr/doc/juro_0990-1027_2012_num_25_3_4730).

<sup>29</sup> L. Pécaut-Rivolier, S. Robin, *Justice et intelligence artificielle, préparer demain* – ep. III, publicat pe Dalloz Actualité, la 17.04.2020.

în timp, datorită actualizărilor multiple pe care instrumentul le necesită. În caz contrar, instrumentul va scăpa treptat sub controlul designerilor.

În Franța, pe fondul intrării în vigoare a *Legii nr. 222/2019 privind programarea 2018-2022 și reforma justiției*<sup>30</sup>, lege care vizează implementarea unor transformări digitale în sistemul judiciar francez, profesioniștii din domeniul juridic și-au manifestat îngrijorarea (demonstrații, articole științifice, interviuri etc.) față de riscurile unei digitalizări prost gândite, nu în sensul de a critica, în general, utilizarea digitalului în justiție, ci lipsa unui filtru, ca și cum totul ar fi bun în transformarea digitală<sup>31</sup>.

O abordare doar administrativă, managerială și tehnică a problemelor reflectă o necunoaștere a funcțiilor judecătorului, care țin de *raționament și motivare, adaptabilitate și loialitate*. Este necesară **o abordare umanistă și interumanistă a digitalizării justiției**, o abordare relațională care ia în considerare atât interesele individuale ale părților, cât și relația lor în complexitatea ei rațională, emoțională și uneori inconștientă.

Desigur, faptul că existau analfabeți în trecut, nu a împiedicat trecerea la procedura scrisă. Tot astfel, și un analfabet digital poate cere ajutor unui avocat pentru a efectua pașii procedurali online.

Însă, adeseori, sentimentul de dreptate se naște din faptul sau posibilitatea ca justițiabilul să fie ascultat nemijlocit de un judecător. Este percepută ca fiind gravă *lipsa cadrului judiciar responsabil de conținerea emoțiilor* (principiul contradictorialității și al oralității, prezența fizică și judecătorul uman), corelată cu excesele rațiunii algoritmice și riscurile asociate capturii realului de către obiecte și persoane conectate<sup>32</sup>.

Pentru a exista dreptate și nu aplicarea mecanică a unor reguli, cu consecințe care sunt uneori inadecvate sau chiar nedrepte, este necesar ca un judecător uman să intervină, cu puterea sa de a modula decizia și de a elimina automatismul pe care algoritmul îl creează.

Așadar, instrumentele IA trebuie să asigure respectarea principiilor fundamentale ale procedurii, cărora li se adaugă principiile specifice justiției digitale: *un principiu de prezență fizică* (care să ofere fiecărei părți dreptul de a iniția o audiere în persoană), *principiul judecătorului natural* și *al cooperării cu noile tehnologii pentru cercetarea faptelor și legii*<sup>33</sup>.

---

<sup>30</sup> Disponibilă pe <https://www.legifrance.gouv.fr>. Consiliul Constituțional sesizat cu *Legea privind reforma în domeniul programării și justiției 2018-2022* a confirmat (prin decizia CC nr. 2019-778), în general, constituționalitatea aspectelor digitale ale legii.

<sup>31</sup> E. Jeuland, *Justice numérique, justice inique?*, Dalloz, Les cahiers de la justice 2019, p. 193.

<sup>32</sup> E. Jeuland, *Le juge et l'émotion*, disponibil pe <https://hal.archives-ouvertes.fr/hal-01790855/>.

<sup>33</sup> E. Jeuland, *Justice numérique, justice inique?*, Les cahiers de la justice 2019, p. 193.

#### 4. Scurte concluzii

Sintetizând, IA necesită un control total al concepției sale. Nu putem lăsa spațiu pentru alegerile implicite de proiectare – ceea ce este ușor, de regulă, pentru informaticieni. Înțelegerea aprofundată a modului în care funcționează IA este o responsabilitate a juriștilor și a factorilor de decizie din lumea juridică. Juriștii trebuie să fie conștienți de acest lucru și să stăpânească instrumentul, astfel încât IA să îi asculte, și nu invers.

Doar în aceste condiții, contribuția IA în justiție poate însemna un adevărat pas înainte, contribuind la o transformare sănătoasă a justiției.